Opetus- ja kulttuuriministeriö

Ministry of Education and Culture

Education

Assessment of undergraduate students' generic skills in Finland

Findings of the Kappas! project

Publications of the Ministry of Education and Culture, Finland 2021:31

Jani Ursin, Heidi Hyytinen & Kaisa Silvennoinen (eds.)

Publications of the Ministry of Education and Culture 2021:31

Assessment of undergraduate students' generic skills in Finland

Findings of the Kappas! project

Jani Ursin, Heidi Hyytinen & Kaisa Silvennoinen (eds.)

Ministry of Education and Culture Helsinki 2021

Publication sale

Online bookstore of the Finnish Government

vnjulkaisumyynti.fi

Publication distribution

Institutional Repository for the Government of Finland Valto

julkaisut.valtioneuvosto.fi

Ministry of Education and Culture

ISBN pdf: 978-952-263-901-1
ISSN pdf: 1799-0351
ISBN painettu: 978-952-xxx-123-xxx
ISSN painettu: xxxx-xxxx
Layout: Government Administration Department, Publications
Helsinki 2021 Finland

Description sheet
18 August 2021

Assessment of undergraduate students' generic skills in Finland
Findings of the Kappas! project

Publications of the Ministry of Education and Culture, Finland 2021:31		**Subject**	Education
Publisher	Ministry of Education and Culture		
Editors	Jani Ursin, Heidi Hyytinen & Kaisa Silvennoinen		
Language	English	**Pages**	166

Abstract

The study investigated the level of Bachelor-level students' generic skills, what factors are connected with the level of generic skills, and to what extent these skills develop during higher education studies. The assessed generic skills included analytic reasoning and evaluation, problem solving, writing effectiveness, and writing mechanics. The participants (n = 2402) were students at initial and final stages of their Bachelor degree programmes from seven universities of applied sciences (UASs) and eleven universities. According to the results, for nearly 60 percent of the higher education students, the generic skills were on a basic or lower level while for the rest, about 40 percent, these were on a proficient or higher level. The variation in students' generic skills was explained mainly by factors pertaining to student's educational and socioeconomic background. In addition, a cross-sectional analysis indicated that generic skills develop to some extent during the studies. Based on the research findings, attention should be paid to the learning of generic skills already at the lower educational levels and also in learning environments outside school. Moreover, the role of generic skills in student admissions should be investigated, and in efforts to develop generic skills in higher education, the different goals of UAS and university education should be considered and learning of generic skills supported in a goal-oriented fashion.

Keywords argumentation, assessment, generic skills, higher education teaching, know-how, polytechnics, problem solving, skills, students, universities, education

ISBN PDF	978-952-263-901-1	**ISSN PDF**	1799-0351
URN address	http://urn.fi/URN:ISBN:978-952-263-901-1		

Kuvailulehti
18.8.2021

Korkeakouluopiskelijoiden geneeristen taitojen arviointi
Kappas!-hankkeen tuloksia

Opetus- ja kulttuuriministeriön julkaisuja 2021:31		**Teema**	Koulutus
Julkaisija	Opetus- ja kulttuuriministeriö		
Toimittaja/t	Jani Ursin, Heidi Hyytinen & Kaisa Silvennoinen		
Kieli	englanti	**Sivumäärä**	166

Tiivistelmä

Tutkimuksen tarkoituksena oli selvittää, millä tasolla korkeakoulututkintoa suorittavien opiskelijoiden geneeriset taidot ovat, mitkä tekijät ovat yhteydessä geneeristen taitojen tasoon ja missä määrin geneeriset taidot kehittyvät korkeakouluopintojen aikana. Tutkimuksessa arvioitavia geneerisiä taitoja olivat analyyttinen päättely ja arviointi, ongelmanratkaisu, argumentatiivinen kirjoittaminen sekä kielen hallinta. Tutkimukseen osallistui yliopistojen alemman korkeakoulututkinnon ja ammattikorkeakoulututkinnon alku- ja loppuvaiheen opiskelijoita (n = 2402) ammattikorkeakouluista (n = 7) ja yliopistoista (n = 11). Tulosten mukaan lähes 60 prosentilla korkeakouluopiskelijoista geneeriset taidot olivat korkeintaan tyydyttävällä tasolla ja noin 40 prosentilla vähintään hyvällä tasolla. Geneeristen taitojen osaamiseroja keskeisimmin selittivät korkeakouluopiskelijan koulutus- ja sosioekonomiseen taustaan liittyvät tekijät. Poikkileikkausasetelma antoi myös viitteitä siitä, että geneeriset taidot kehittyvät jossain määrin opintojen aikana. Tutkimuksen tulosten perusteella geneeristen taitojen oppimiseen on kiinnitettävä huomiota jo alemmilla koulutusasteilla ja koulun ulkopuolisissa oppimisympäristöissä, geneeristen taitojen merkitystä opiskelijavalinnoissa on selvitettävä, geneerisiä taitoja kehitettäessä on otettava huomioon ammattikorkea- ja yliopistokoulutuksen erilaiset tavoitteet sekä geneeristen taitojen oppimista on tavoitteellisesti tuettava korkeakouluopinnoissa.

Asiasanat ammattikorkeakoulut, argumentointi, arviointi, geneeriset taidot, korkeakouluopetus, ongelmanratkaisu, opiskelijat, osaaminen, taidot, yliopistot, koulutus

ISBN PDF	978-952-263-901-1	**ISSN PDF**	1799-0351
Julkaisun osoite	http://urn.fi/URN:ISBN:978-952-263-901-1		

Presentationsblad
18.8.2021

Utvärdering av högskolestuderandes generiska färdigheter
Resultat från Kappas! -studien

Undervisnings- och kulturministeriets publikationer 2021:31		**Tema**	Utbildning
Utgivare	Undervisnings- och kulturministeriet		
Redigerare	Jani Ursin, Heidi Hyytinen & Kaisa Silvennoinen		
Språk	engelska	**Sidantal**	166

Referat

Syftet med studien var att ta reda på vilken nivå de studerandes generiska färdigheter ligger för studerande med kandidatexamen, vilka faktorer som är relaterade till generiska färdigheternas nivå och i vilken utsträckning generiska färdigheter utvecklas under högskolestudierna. De generiska färdigheter som utvärderades i studien var analytiskt resonemang och utvärdering, problemlösning, argumenterande skrivande och språkhantering. I studien deltog studerande som var i början och i slutet av sina kandidatstudier (n = 2402) från yrkeshögskolor (n = 7) och från universitet (n = 11). Enligt resultaten låg nivån för de generiska färdigheterna på högts en tillfredsställande nivå för närmare 60 procent av de högskolestuderande och för cirka 40 procent minst på en bra nivå. Skillnaderna i generiska färdigheter förklarades främst av faktorer som var relaterade till de högskolestuderandes utbildnings- och socioekonomiska bakgrund. Tvärsnittsforskning gav också indikationer på att generiska färdigheter utvecklas till en viss del under studierna. Baserat på resultaten av studien måste uppmärksamhet ägnas åt de generiska färdigheternas inlärning redan på de lägre utbildningsnivåerna och i inlärningsmiljöer utanför skolan. Vikten av generiska färdigheter vid intagningen av de studerande måste utredas och uppmärksamhet ska också läggas vid utvecklingen av de generiska färdigheterna med hänsyn till de olika målen som yrkeshögskolans utbildning och universitetsutbildningen har. Inlärningen av de generiska färdigheterna ska också målmedvetet stödjas vid högskolestudier.

Nyckelord	argumentation, färdigheter, generiska färdigheter, högskoleundervisning, kunnande, problemlösning, studerande, universitet, utvärdering, yrkeshögskolor, utbildning

ISBN PDF	978-952-263-901-1	**ISSN PDF**	1799-0351
URN-adress	http://urn.fi/URN:ISBN:978-952-263-901-1		

Contents

FOREWORD

The purpose of the Assessment of Undergraduate Students' Learning Outcomes in Finland (Finnish acronym Kappas!) project was to produce information on students' skills in order to develop higher education and improve its quality. The project has produced important information on students' mastery of generic skills for higher education institutions. It has also put the generic skillsets considered essential for working life and continuous learning at the centre. Among other things, these skills are expressed and they can be developed in higher education studies in the form of critical thinking, problem solving skills and logical reasoning. A key question thus is to what extent generic skills should be consciously addressed in teaching and highlighted as a clear part of higher education institutions' curricula.

The Kappas! project is particularly significant from the perspective of teaching development. Learning outcomes are influenced by many factors, starting from the student's capability and possibilities for studying. A link has been found between teaching methods, arrangements and approaches on the one hand and students' study motivation and learning outcomes on the other. The project has striven to produce for higher education institutions information that would provide concrete benefits for efforts to develop education and teaching.

Higher education institutions' quality management forums and various initiatives have drawn attention to the quality of teaching. Student feedback on teaching and guidance and feedback obtained from employers as part of the quality assurance system are important tools for evaluating education and outcomes. Information on learning outcomes adds an important dimension to this big picture. The Kappas! project offers interesting observations on the experiences of higher education institutions that have selected different pedagogical solutions and teaching practices, among other things.

The global nature of the research instrument used in the Kappas! project makes it possible to draw on international benchmarking data and learn from other countries. Finland has previously participated in an OECD project which examined the possibilities of mapping higher education students' learning outcomes at the international level. In this project, too, developing teaching and learning was an important perspective. Finland is an interesting comparison country in terms of education policy, and Finland's initiative in assessing higher education students' learning outcomes as a national project has attracted

interest around the world. The OECD's aim is to build international understanding and trust in a world where different types of rankings are used to compare higher education institutions.

The objective of the Kappas! project was to illustrate the importance of generic skills and to offer higher education institutions opportunities, information and tools for developing teaching and supporting students' preconditions for learning. The participating students gained valuable information about their generic skills and possibilities of improving them.

The project involved 18 higher education institutions, more than 2,400 students took the test, and a large number of higher education institution staff participated in the practical arrangements of the research project. I would like to thank all participating higher education institutions, staff members and students for your major contribution to the Kappas! project!

Maarit Palonen, Counsellor of Education
Chair of the Kappas! project's support group

1 Introduction

Jani Ursin & Maarit Palonen

1.1 Background

The capabilities produced by higher education have been discussed actively over the last few decades. In European higher education policy, this debate has been closely associated with improving the quality of higher education and the aim of creating a European Qualifications Framework (EQF) to make national qualifications easier to understand and compare (e.g. Ursin 2014a). Consequently, the European Union has for its part emphasised the wide range of skills higher education graduates need to meet the new requirements of working life: not only professional competence but also problem solving, team work and interaction skills (e.g. European Commission 2013). Additionally, generic skills are believed to be one of the key competences required for continuous learning and digitalisation (European Commission 2019). The European Economic and Social Committee states in its opinion that in the era of digitalisation and artificial intelligence

> *[i]t is relevant to ask which skills bring added value over machines and robots and which skills we want to keep anyway. This highlights the need for a strong foundation in cross-cutting skills such as logical reasoning, critical thinking, creativity and interpersonal and interaction skills. (Opinion of the European Economic and Social Committee…2020.)*

The skills imparted by higher education are also regarded as important in the national context. Finland's goal is to be the world's most competent country in which the higher education community produces the globally best learning and learning environments. The competence created by higher education plays a crucial role in responding to the transition of the labour market and the needs of continuous learning. High-quality education is thus also considered to provide tools for global influence. (Ministry of Education and Culture 2017.)

Finnish higher education policy has to a great extent been based on striving for consistent quality. We have learned from international comparisons of skills that evaluating the quality of teaching and learning is not simple. Consequently, international comparisons of higher education have mainly focused on research, while teaching has received less attention. One exception to this was the International Assessment of Higher Education Learning Outcomes (AHELO) carried out by the OECD, which investigated the possibilities

of measuring higher education students' learning with a comparative approach (Tremblay, Lalancette & Roseveare 2012; Ursin 2014b).

What knowledge and skills is higher education then expected to produce? A key goal of higher education studies is to develop the student's expertise in their selected field. This expertise is based on knowledge and skills specific to each field of study but also on different generic, or universal, skills. While there are many types of generic skills (see Chapter 2 in this report), higher-order cognitive skills are usually put centre stage in higher education, including an ability for critical thinking and argumentation, analytical reasoning and making justified decisions (e.g. Zoller & Tsaparlis 1997; Arum & Roksa 2011; Lemons & Lemons 2017). In the much-used Bloom taxonomy, which describes levels of learning, these generic skills are placed in the top categories that contain skills in analysing, creating and evaluating information (Bloom, Englehart, Furst, Hill & Krathwohl 1956). Rather than only being essential in working life, generic skills are thus already needed for studies and continuous learning. Previous studies have found, however, that higher education students experience challenges in such areas as argumentation, evaluation of information and making conclusions (e.g. Badcock, Pattison & Harris 2010; Arum & Roksa 2011; Evens, Verburgh & Elen 2013; Hyytinen, Nissinen, Ursin, Toom & Lindblom-Ylänne 2015).

1.2 Objectives of the Kappas! project

The generic skills assessed in the Kappas! project are: (1) analytical reasoning and evaluation (recognising weaknesses and strengths in different justifications), (2) problem solving (identifying a problem situation and resolving it with justifications), (3) writing effectiveness (producing text with convincing arguments), and (4) writing mechanics (mastering the conventions of written language). These skills have been found essential in such studies as the universities' latest career monitoring survey (Suorsa & Sainio 2020), and they are important in all fields of study (e.g. Shavelson 2010; Tuononen et al. 2017). The goal of the Kappas! project was to find out about

1. the level of Finnish higher education students' generic skills
2. the factors associated with the students' mastery of generic skills, and
3. the extent to which these skills develop during higher education studies.

The idea was that the information produced by the project could be used following the principles of enhancement-led evaluation (Finnish Education Evaluation Centre 2020). In practice, this meant that each student who participated in the Kappas! study received a report on their test score as well as support material which allows them to improve their generic skills, should they wish to do so. Additionally,

each participating higher education institution received a report on their students' performance in the test. A webinar was organised for interested higher education institutions to review each institution's Kappas! results. In particular, the webinar focused on how generic skills could be integrated into degree programme curricula and teaching. At the national level, the project produces information on the generic skills of Bachelor's degree students' at universities and universities of applied sciences. The national results were compared to reference data from the United States to obtain a more comprehensive picture of Finnish higher education students' mastery of generic skills. This helps identify weaknesses and strengths related to mastering generic skills at the level of the higher education system.

1.3 Kappas! research organisation and report structure

In compliance with the Act on Public Procurement and Concession Contracts, the Ministry of Education and Culture organised a call for tenders to select a national coordinator for the assessment of higher education students' learning outcomes. The deadline for submitting tenders was 31 August 2018. The Ministry of Education and Culture concluded a contract on the coordination tasks with the University of Jyväskylä's Finnish Institute for Educational Research. The project was carried out between 1 October 2018 and 31 December 2020 by the University of Jyväskylä's Institute for Educational Research (KTL) together with the University of Helsinki's Centre for University Teaching and Learning (HYPE). The project was coordinated by a 23-strong support group appointed by the Minister of Education and Culture, which included a representative from each higher education institution participating in the project and from the Ministry of Education and Culture. The support group was chaired by Counsellor of Education Maarit Palonen from the Ministry of Education and Culture. In addition to coordinating the project, the support group selected the tests used in the project and commented on the project's draft final report. The Council for Aid to Education (CAE), a US non-profit organisation, participated in the project as an international partner. Its tasks included developing the test instrument used in the project, implementing an electronic testing environment and producing reports for higher education institutions and students.

This final report of the Kappas! project has six chapters. The Introduction describes the background, objectives and research organisation of the project. Chapter 2 illustrates the meaning and assessment methods of generic skills in higher education. Chapter 3 describes project implementation: the research instrument and skill categories, translation and pre-testing of the instrument, student selection, scoring, collection and analysis of data, and feedback reports received by higher education institutions and students. The chapter also discusses the reliability of the Kappas! study. Chapter 4 presents the key results of the project by research question, and Chapter 5 showcases higher education

institutions' and students' experiences of the assessment project. The final chapter presents the key conclusions and development proposals of the project and points the way to the next steps. To make the report easier to read, the central observations of each chapter are briefly presented at its beginning.

2 Generic skills and their assessment in higher education

Heidi Hyytinen, Katri Kleemola & Auli Toom

- Generic skills are the key to higher education students' learning and the building of expertise specific to the field of study.
- Generic skills consist of skills, knowledge and willingness.
- Self-assessments, performance tasks and observation are used to assess generic skills.
- Different methods produce different types of knowledge of generic skills.

2.1 What are generic skills in higher education?

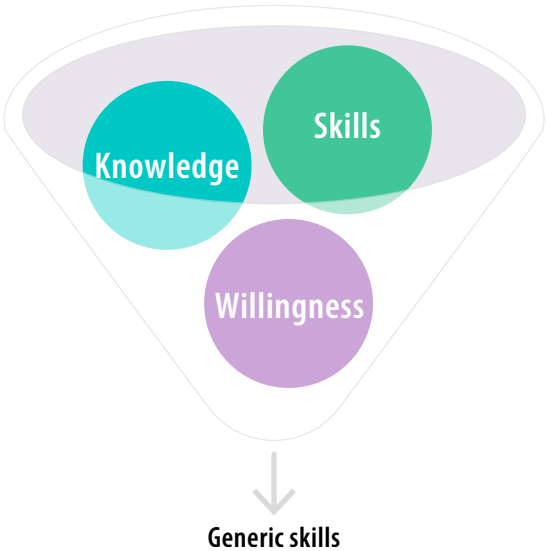
Generic skills have in recent years made it to the top of the agenda both in higher education institutions' curriculum work and in discussions on higher education policy (e.g. Miettinen 2019; Chapter 1). It has become apparent that working life is changing rapidly and that it is not solely sufficient for higher education graduates to master field-specific skills. Generic skills, including problem solving and interaction skills, have become key skills that higher education aims for. This is why learning generic skills, rather than only building expertise and learning knowledge specific to the field, is seen as the purpose of higher education studies. In a changing working life it is essential that graduates are capable of continuous learning, and the key to continuous learning is mastering generic skills (cf. Muukkonen, Lakkala, Lahti-Nuuttila, Ilomäki, Karlgren & Toom 2020).

Generic skills are universal expert skills needed in higher education studies and working life and equally important in all fields. While different skills lists are often used to define generic skills, none of them are exhaustive (see e.g. Virtanen & Tynjälä 2013; Hyytinen, Toom & Shavelson 2019; Tuononen, Kangas, Carver & Parpala 2019a; Muukkonen et al. 2020). The generic skills that play a key role in higher education studies include critical thinking, problem solving, evaluation of information sources, justification, interaction skills, self-regulation and writing (e.g. Virtanen & Tynjälä 2013; Halpern 2014; Hyytinen et al. 2019;

Tuononen et al. 2019a; Kleemola, Hyytinen & Toom 2021). Generic skills enable students to draw on their field-specific expertise and competence both during their studies and later in working life (Arum & Roksa 2011; Virtanen & Tynjälä 2013; Tuononen, Parpala & Lindblom-Ylänne 2019b). Writing skills, for example, enable you to make your ideas and conclusions visible to others. The skills of justification and argumentation, on the other hand, help you defend your views. If students have challenges related to their generic skills, they may also not be able to fully apply field-specific information, whereas those who master generic skills can apply their field-specific knowledge to a variety of situations and draw on their expertise more diversely when working together with experts of different fields. Links have been found between generic skills and study success, progress in studies as well as a deep approach to learning (Badcock et al. 2010; Arum & Roksa 2011; Tuononen et al. 2019b).

Generic skills are tightly intertwined with field-specific knowledge and skills. To enable students to make full use of their knowledge and generic skills, willingness is also needed (Halpern 2014; Heijltjes, van Gog, Leppink & Paas 2014; Hyytinen et al. 2019; Figure 1). Willingness comprises self-efficacy, perseverance, curiosity and an understanding of when the skills are needed. Willingness helps the student cope also with challenging tasks.

Figure 1. Rather than just skills, generic skills are also about knowledge and willingness



Major differences have been found in the generic skills of initial stage higher education students (e.g. Utriainen, Marttunen, Kallio & Tynjälä 2017; Hyytinen, Toom & Postareff 2018). The skills may also develop unevenly: the students' strengths and challenges vary

(Hyytinen 2015; Kleemola et al. 2021). In addition, students have different perceptions of the usefulness of various generic skills, and not all students recognise their skills (Tuononen, Parpala & Lindblom-Ylänne 2017; Tuononen 2019; Tuononen et al. 2019b).

2.2 How can generic skills be assessed?

Generic skills can be assessed using a number of different methods, including self-assessments, performance-based assessments and observation, which are discussed in detail in this section. In the Kappas! project, a performance-based method was used.

The most traditional assessment method is self-assessment. Self-assessment data are typically collected by means of interviews, learning journals or surveys, for example by asking students how well they think they master different generic skills or what knowledge they have learned during specific courses or studies (e.g. Virtanen & Tynjälä 2013, 2018; Tuononen et al. 2017, 2019a, 2019b). The advantage of self-assessments is that they are relatively easy and cost-effective to conduct. Surveys, in particular, can at best reach a large group of respondents. Self-assessment is an effective method for examining students' personal viewpoints, including their experiences, beliefs and perceptions of their mastery of generic skills. However, self-assessments tell us nothing about students' actual levels of mastery and the way they apply their skills in authentic situations. Studies have shown that students find it difficult to assess their competence and skills levels realistically (Tuononen et al. 2017; Tuononen 2019). Students may either overestimate or underestimate their skills (e.g. Hyytinen, Holma, Shavelson & Lindblom-Ylänne 2014; Hyytinen, Postareff & Lindblom-Ylänne 2020a). Estimates of generic skills based on self-assessments are thus only indicative and cannot replace information on students' skills produced by performance-based assessments.

Performance-based assessments are one way of drilling down into students' generic skills (McClelland 1973; Shavelson 2010; Hyytinen 2015; Hyytinen et al. 2015; Shavelson, Zlatkin-Troitschanskaia & Marino 2018; Hyytinen & Toom 2019). In these assessments, students are typically given a task they could encounter in real life, for example in their studies or at the workplace, and they must draw on different generic skills in order to solve the task. For example, students could be asked to analyse information obtained from different sources, apply the results of their analysis to problem solving and reasoning, and communicate their response, proposal for a solution or recommendation for action either by writing or by selecting the correct answers to selected-response questions (Shavelson 2010; Hyytinen et al. 2015; Shavelson et al. 2018).

Various types of assessments emphasise different skills and thus highlight different aspects of generic skills (Hyytinen et al. 2015; Hyytinen, Ursin, Silvennoinen, Kleemola & Toom 2020b; Kleemola et al. 2021). Selected-response questions can be used to assess logical reasoning, literacy and skills in analysing the credibility and consistency of arguments, among other things. It is important to note, however, that selected-response questions do not provide information on students' ability to produce answers independently, including their ability to build arguments. For this, a performance task in the form of an essay is needed, in which the students give their response by writing. Answering selected-response questions does also not provide information on the thinking processes through which the students reach their answers (Messick 1994; Hyytinen et al. 2015), as these tasks primarily guide the students to select the best option among the responses offered to them. Written answers to performance tasks give indications of the processes that underlie the answer, including of how well students understand the materials given to them and draw on the materials in their answers. Selected-response questions usually focus on individual and separate skills, whereas it has been found that performance tasks require a combination of several different skills (Shavelson 2010) and versatile self-regulation (Hyytinen & Toom 2019; Hyytinen et al. 2020b). In addition, guessing and excluding unlikely options are commonly used strategies in selected-response questions (Hyytinen et al. 2015, 2020b). As different types of assessments focus on different skills, they produce different types of information and perspectives on students' skills (Hyytinen et al. 2015). The advantage of performance-based assessments is that students receive feedback on their performance and level of mastery. The assessments themselves are also learning situations: they can stimulate students to reflect on their skills. Reflection together with feedback enable students to recognise their skills and development needs (Tuononen et al. 2017; Tuononen 2019).

The challenge of using performance-based assessments lies in that they require a great deal of resources and expertise of those setting and organising the assessment compared to those based on self-assessment (Zlatkin-Troitschanskaia, Shavelson, Schmidt & Beck 2019). To develop the assessment, you need an understanding of both the generic skills that you aim to assess and the assessment methods that capture the desired skills. To assess open-ended written responses (and to interpret and score the answers), accurate assessment criteria based on the objective of the task are required and, to ensure consistent scoring, at least two trained scorers (Borowiec & Castle 2019). Test situations may also put participants under stress, as responding carefully to a performance-based assessment takes time and effort. Only a limited number of skills can be assessed in a single test situation: due to the length of the time required, the test situation may become too stressful for students, or an attempt to assess many different skills at the same time may undermine the reliability of the assessment (see Zlatkin-Troitschanskaia et al. 2019). When designing performance-based assessments, the aim is to make them interesting

for students and to inspire students to put in an effort in the assessment task (Ercikan & Pellegrino 2017).

Observation is one way of getting a handle on the processes of generic skills. It can provide information on how students operate in situations requiring them to use their generic skills. It also allows the observer to assess some generic skills, including interaction skills, which are difficult if not impossible to pin down by other methods. Observation is used either independently or in addition to and in support of interviews or surveys. In observation related to research in generic skills, the thinking aloud method is often used: when solving the task, the student speaks their thoughts aloud (Leighton 2019). The observed situations are video recorded to avoid focusing solely on oral expression when analysing the material. Gestures, expressions, postures and other behaviour can also be analysed. It goes without saying that observation is extremely laborious, and using it to assess a large group of students is challenging and requires a great deal of resources.

In order to draw conclusions on a student's generic skills, we need to somehow get an idea of what they know, what kinds of thinking processes they have, and how they act in situations calling for generic skills (McClelland 1973; Shavelson 2010; Hyytinen & Toom 2019; Zlatkin-Troitschanskaia et al. 2019). No individual assessment method responds to these needs on its own. When we understand their strengths and weaknesses, we can choose an appropriate method or combination of methods for each need. An individual method or task may not be sufficient to describe the multiple facets of generic skills (Hyytinen et al. 2020a; Kleemola et al. 2021). The more in-depth perception of the student's skills we wish to obtain, the more diversely their skills must be assessed using different tasks and methods. When interpreting the results of the assessment, the student's willingness should also be taken into account. The results may be influenced by the student's motivation to participate, the effort they put in when performing the test, and their metacognitive skills, or ability to regulate their thinking, emotions, behaviour and performance in the test situation (Arum & Roksa 2011; Hyytinen & Toom 2019; Hyytinen et al. 2020b).

3 Implementation of the study

Jani Ursin, Heidi Hyytinen, Kari Nissinen & Kaisa Silvennoinen

- The generic skills assessed in this study were analysis and problem solving, writing effectiveness, writing mechanics, critical reading and evaluation, scientific and quantitative reasoning and critiquing an argument.
- The duration of the test was around two hours, and computers were used to take it. The test contained a performance task in essay format, 25 selected-response questions and a background information survey.
- Pre-testing was carried out to ensure that the test versions translated into Finnish and Swedish worked as intended.
- Students who represented different fields of study as comprehensively as possible at the national level were selected to participate in the study.
- The participants comprised 2,402 students from 7 universities of applied sciences and 11 universities.
- The data were analysed using design-based methods.
- The students and higher education institutions received reports detailing their test results.

3.1 Assessment of generic skills in the Kappas! project

3.1.1 CLA+ International test as a research instrument

The performance-based Collegiate Learning Assessment (CLA+) International test developed by the Council for Aid to Education (CAE), a US non-profit organisation, was used in the Kappas! project. The CLA+ test has been used to assess generic skills in a number of higher education institutions in the United States, Italy and the United Kingdom (Zahner & Ciolfi 2018; Zlatkin-Troitschanskaia et al. 2019). The CLA+ test consisted of two main sections: an essay-form performance task and selected-response questions (Table 1).

In the performance task, a problem they could encounter in real life was presented to the students, who were asked to produce a written solution to it. In the Kappas! project, the students had to compare life expectancies in two cities and consider if some measures were needed to increase the life expectancy in one of them. In their responses, the students had to present a solution to the problem and give a recommendation for possible measures. They were asked to use documents relevant to the task found in a document library to produce their responses. The library contained five different source documents: a blog text, a transcribed podcast, a memorandum, a newspaper article and infographics. In order to respond to the performance task, the students had to draw on several generic skills simultaneously, such as analysing the information in the source documents, evaluating the reliability of the information, distinguishing between essential and non-essential information, using problem solving skills, making conclusions and giving justifications. The students were given 60 minutes to respond to this task in writing.

In the second part of the test, the students answered 25 selected-response questions. Similarly to the performance task, the students had to draw on the information found in the documents of the document library to answer the selected-response questions. Ten of the questions were relevant to the background documents, which dealt with secretion of proteins in the brain. These questions measured scientific and quantitative reasoning. A second set of ten questions, which was based on a reader's letter about nanotechnology sent to an imaginary journal, measured critical reading and evaluation. The last five questions, which related to an opinion piece on women in combat, assessed the students' ability to analyse arguments, including possible logical fallacies. In the selected-response questions, the students were asked to select the correct answer among four options. The students were given 30 minutes to respond to this section of the test.

At the end of the test, the students answered background information questions (Appendix 1), the first 17 of which are a permanent part of the CLA+ test. The following 20 questions were jointly tailored by the project team and the support group. The background information questionnaire was used to collect information on the participating students' backgrounds (including socioeconomic and educational background as well as mother tongue and mathematics grades in the matriculation examination) and their current education and studies. Their socioeconomic background was measured by parental education on the one hand and the number of books in their childhood homes on the other. The number of books in the childhood home is regarded as an indication of a culture of learning and reading (e.g. De Graaf 1986; Sikora, Evans & Kelley 2019), and many international large-scale assessments have found it one of the strongest predictors of learning outcomes for both basic education pupils (Sulkunen & Nissinen 2012; Arffman & Nissinen 2015; Vettenranta, Hiltunen, Nissinen, Puhakka & Rautopuro 2016; Leino, Nissinen, Puhakka & Rautopuro 2017; Hiltunen & Nissinen 2018; OECD 2019) and adults (e.g. Sikora et al. 2019). The background questionnaire also asked how engaging the students found the test and how much effort they put in to complete it.

Table 1. Sections and skill categories in CLA+ International test

Test sections	Skill categories
1) Performance task (60 minutes)	• Analysis and problem solving • Writing effectiveness • Writing mechanics
2) 25 selected-response questions (30 minutes)	• Critical reading and evaluation • Scientific and quantitative reasoning • Critiquing an argument
3) Background questionnaire	• Demographic data, including age and gender • Questions about socioeconomic status • Questions about studies and field of study • Questions about educational background • Questions related to the test situation

The students took the CLA+ test in a standardised electronic test interface. The tasks in the different sections were designed to eliminate the need for any subject-specific knowledge or other special expertise. In the test situation, the test interface provided the students with instructions for completing each section and other information needed to do the tasks. The test situations were supervised. Due to copyright issues and because the test tasks are still used internationally, the performance task and selected-response questions used in the Kappas! project cannot be published. However, an example of the CLA+ test sections and test interface can be found at http://www.starttest.com/7.0.0.1/programs/clacross/CLA%20Practice%20Test%20 Page.htm.

3.1.2 Skill categories

The generic skills assessed in the Kappas! project focused especially on critical thinking and writing skills. Regardless of the field of study, these skills are important in studies, working life and civic activity (e.g. Shavelson 2010; Arum & Roksa 2011; Hyytinen et al. 2019). The assessment project also focused on the most central skills needed in different fields, one of the reasons for this being that performance-based assessments, including CLA+, can only be used to assess a limited number of skills at once (Chapter 2.2). Additionally, the skills assessed in the project required an active approach of the students: in order to complete the tasks, they had to be willing to carefully peruse the documents in the document library, have the skills to evaluate information related to the situation

and put it into a context, look at matters from multiple perspectives, make informed choices and decisions, and weigh conclusions from different perspectives and justify them convincingly (Shavelson 2010; Arum & Roksa 2011).

The students' responses in the performance task were assessed in three areas: analysis and problem solving; writing effectiveness; and writing mechanics (Zahner & Ciolfi 2018). *Analysis and problem solving* refer to the student's ability to identify a problem situation, analyse information related to it and, based on their analysis, present a conclusion and solution to the situation. The assessment also focuses on the student's ability to analyse the reliability of information and examine conflicting information or alternative justifications. *Writing effectiveness* refers to the student's ability to form consistent and logically structured arguments. *Writing mechanics*, on the other hand, measures the extent to which the student masters and follows the conventions of written language in their response.

Critical reading and evaluation in the selected-response section of the CLA+ test refers to the student's ability to evaluate the reliability of information, to identify conflicting information in the documents and to make interpretations. *Scientific and quantitative reasoning* refers to the student's ability to interpret quantitative data, to produce conclusions and hypotheses based on the information provided, to identify questionable assumptions and to assess alternative conclusions. *Critiquing an argument* is understood as the student's ability to assess the logic of arguments, including their strengths and weaknesses, and to recognise fallacies.

3.2 Selection, translation and pre-testing of CLA+ test tasks

The support group of the Kappas! project selected the tasks used in the study. In June 2018, the support group members familiarised themselves with three performance tasks and two series of selected-response questions developed by CAE. They selected for the project one performance task and one set of 25 selected-response questions, which were considered the best suited for students in both higher education contexts. The support group also found it important that no specialist competence specific to a field of study was required to understand the source documents provided for the tasks. In addition, tasks considered too difficult by the support group were eliminated.

The CLA+ test was translated into Finnish and Swedish. The translation and adaptation of the test progressed through four main steps in each language version (cf. Hambleton & Patsula 1998). In the first phase cApStAn, a subcontractor of CAE, translated the English

test into Finnish and Swedish. Two translators in Finland then reviewed, confirmed and, if necessary, proposed changes or corrections to the translations independently of each other. In the third phase, the Kappas! project team decided on the final versions of the translations based on the translators' proposals. The translated test was then pre-tested in Finnish and Swedish in so-called cognitive laboratories, ensuring that the translation and adaptation phase had not changed the meaning, level of difficulty or internal structure of the test.

The cognitive laboratories were carried out with 20 target group students from two different higher education institutions (Table 2). Cognitive laboratories are based on the think-aloud method: the students strive to verbalise their thought process (Leighton 2017, 2019). At the beginning of the cognitive laboratories, the students were informed of the purpose of the pre-testing and asked to complete a research permit, and they gave their consent to having the situation filmed. The students then practised thinking aloud with easy sample tasks. Next, they moved on to the actual thinking aloud phase. They took a computer test that included the same elements as the actual test: instructions, a performance task, a selected-response section and a background questionnaire. As thinking aloud is a slow method, however, the students were allowed more time than in the actual test situation. Each cognitive laboratory was attended by one or two researchers in addition to the student. The researchers did not intervene in the situation, except to remind the student to think aloud if the student did not do so. After the think-aloud phase, the students were also interviewed briefly. All students answered questions about the adequacy of the instructions, comprehensibility of the test, their response strategy and their engagement with the test in the interview. Additionally, the researchers made notes of potentially important observations or ones requiring further clarification, and if such observations came up, they were discussed with the student at the concluding interview. All cognitive laboratories were recorded, filmed and transcribed to document the relevant information. Each student participating in the cognitive laboratory was given a cinema ticket.

Table 2. Students who participated in the cognitive laboratories (n = 20)

Gender	Year of birth	Field of study	Higher education sector	Language
male	1993	Humanities or Arts	university	Finnish
female	1996	Humanities or Arts	university	Finnish
male	1990	Science	university	Finnish
female	1991	Science	university	Finnish
male	1993	Humanities or Arts	university	Finnish
female	1997	Social Sciences	university	Swedish
male	1989	Social Sciences	university	Swedish
female	1997	Social Sciences	university	Swedish
female	1997	Science	university	Swedish
female	1997	Social Sciences	university	Swedish
male	1996	Science	university	Swedish
female	1964	Health or Welfare	university of applied sciences	Finnish
female	1982	Health or Welfare	university of applied sciences	Finnish
female	1971	Health or Welfare	university of applied sciences	Finnish
female	1997	Services	university of applied sciences	Finnish
male	1983	Engineering, Manufacturing, Architecture or Construction	university of applied sciences	Finnish
male	1985	Arts	university of applied sciences	Finnish
female	1993	Arts	university of applied sciences	Finnish
female	1989	Arts	university of applied sciences	Finnish
male	1997	Science	university	Swedish

The cognitive laboratories confirmed that the tests translated into Finnish and Swedish worked as intended and that the translation and adaptation phases had not changed the meaning or level of difficulty of the test (Ercikan & Pellegrino 2017; Leighton 2017, 2019). However, minor changes were made to the translated CLA+ test based on the cognitive laboratories: for example, the instructions for the performance task were shortened to fit on a single page, in addition to which minor linguistic clarifications were made, and spelling errors were corrected.

3.3 Student selection

The Ministry of Education and Culture invited all Finnish higher education institutions to participate in the Kappas! project. Seven universities of applied sciences and eleven universities accepted the invitation (Table 3). One third of the universities of applied sciences and most universities in Finland took part in the study. In addition, the participating higher education institutions represented Finnish higher education institutions of different sizes and in different regions rather comprehensively. Two of the participating higher education institutions were Swedish-speaking, three were bilingual and the rest were Finnish-speaking.

Table 3. Higher education institutions participating in the Kappas! project

University of applied sciences	University
1. Häme University of Applied Sciences	1. Aalto University
2. South-Eastern Finland University of Applied Sciences	2. University of Helsinki
3. Lapland University of Applied Sciences	3. University of Eastern Finland
4. Laurea University of Applied Sciences	4. University of Jyväskylä
5. Metropolia University of Applied Sciences	5. Lappeenranta-Lahti University of Technology LUT
6. Savonia University of Applied Sciences	6. National Defence University
7. Tampere University of Applied Sciences	7. University of Oulu
	8. Hanken School of Economics
	9. University of the Arts Helsinki
	10. University of Turku
	11. Åbo Akademi University

The objective of the Kappas! study was to collect nationally representative data concerning the students of the participating higher education institutions. Foreign exchange students and other students who did not speak the languages used in the CLA+ test (Finnish or Swedish) as their mother tongue were excluded from the target population. The main criterion for representativeness was obtaining a sufficient number of responses from all the different fields of study offered in universities of applied sciences and universities in Finland. In addition, an effort was made to ensure an equally balanced representation of initial stage students and those (third-year) students who were in the final stage of the Bachelor's degree.

Representativeness within individual higher education institutions was not sought in the data collection. The main reason for this was the practical impossibility of recruiting enough students from multidisciplinary higher education institutions to ensure that the data would cover all fields of study at the level of each institution with sufficient accuracy. In multidisciplinary higher education institutions, data were only collected on a few selected fields of study, ensuring data representativeness at the national level. This meant that students from different fields of study were selected from different higher education institutions, and when the data obtained from different institutions were aggregated, sufficient national representation of all fields of study was achieved. At the same time, however, it was ensured that students from more than one higher education institution were tested in each field of study. An (imaginary) example illustrating such differentiated data collection would be selecting students of technology, business and culture from one higher education institution, students of business, social and health care and tourism from a second institution, students of social and health care, natural resources and tourism from a third institution, and students of technology, culture and natural resources from a fourth one. In this example, in each institution data would be collected from three fields of study, and at the same time each field of study receives data from two different institutions. The institutions' preferences were taken into account in the data collection as far as possible; however, the final selection of the education or degree programmes and their students was carried out by random sampling conducted independently of the higher education institution and programme.

As the participation of higher education institutions in the study was voluntary, the selection of students from the national student population could not be completely random, students being drawn from the seven universities of applied sciences and eleven universities that wished to participate in the study (Table 3). The practical constraints of the student selection also included the workload that organising Kappas! test events imposed on individual higher education institutions. The number of students selected for the study from each higher education institution was kept limited to avoid imposing an unreasonable workload on the institution (fitting the test events in the students' schedules, reservations of rooms or computer classes, personnel resources needed for

supervising the tests). In this sense it was helpful to the institutions that students only needed to be tested in a few fields of study (following the differentiated data collection design illustrated above).

The basis of the sampling plan was to select 200 initial stage and 200 final stage students from each participating higher education institution. It was expected that roughly one half of the students invited to take the test would not turn up, which would result in a data set of about 1,800 initial stage and 1,800 final stage students. However, a higher number of students from the largest multidisciplinary institutions was invited to take the test to ensure that sufficient representation of several fields of study would be secured in the data set.

Prior experience, including data collection for the AHELO feasibility study in 2012 (Tremblay et al. 2012; Ursin 2014b), has shown that it is difficult to get Finnish higher education students to participate in these types of studies on a voluntary basis and in their own time. Selecting individuals by random sampling has produced very poor results (in the AHELO study, the participation rate of Finnish students was as low as 14%). This is why in the Kappas! project a decision was made to use cluster sampling based on tutor groups (initial stage students) or seminar groups or other courses included in the curriculum (final stage students). At the first stage, the education or degree programmes whose students (initial stage or final stage) would be invited to the tests were randomly selected within the higher education institution. If the programme selected in this way had several tutor groups or the students were otherwise divided into several parallel groups (courses or seminar groups), one or more of the groups were then randomly picked, ensuring that the targeted number of students was reached. If the total number of students in the programme selected for the study was not very large in total (in other words, the resources of the higher education institution made it possible to test everyone), all students of this programme were invited to take the test with no division into smaller groups. The students of the selected groups were expected to attend a test session at a given time under the direction of their tutor or teacher. This method resulted in a clearly higher participation rate than the method based on individual sampling in the AHELO study, for example, but anyway the participation rate reached was nowhere near the 50% target: in the Kappas! study, the overall student participation rate was 25%. There was major variation between higher education institutions and fields of study, however. A great deal depended on how active the tutor or teacher of the group included in the sample was. Initial stage university of applied sciences students had the highest participation rate (39%), while this rate for initial stage university students was 25%. The participation rate of final stage students was lower than that of initial stage students, being 26% for universities of applied sciences and as low as 15% for universities.

3.4 Test implementation

The CLA+ tests were organised in higher education institutions between August 2019 and March 2020. While the plan was to continue the tests until early April 2020, the coronavirus situation made it necessary to finish the testing earlier than planned in six higher education institutions. However, most higher education institutions were able to complete their testing before the exceptional situation began.

All tests were performed using computers in a secure network environment. Consequently, the student was unable to browse the Internet or use other software installed on the computer during the test. The students could take the test in either Finnish or Swedish. The test interfaces intended for the students and test proctors as well as all the instructions for test arrangements thus had to be translated into Finnish and Swedish. While hardware belonging to the higher education institutions was mainly used for the tests, students' personal devices could also be used if necessary. The test situations were supervised, and the planned duration of the test, including reading the instructions, was 2 hours and 15 minutes in total. The students had to complete the test sections in a specific order. They could only move on to the selected-response questions once they had completed the performance task. The three sets of questions in the selected-response section were presented in the same order to all students (1. 'brain protein', 2. 'nanotechnology', and 3. 'women in combat'), but the students were able to move around freely between the three sets if they so wished (section 3.8). Finally, the students filled in a background questionnaire. The students were able to finish the different sections of the test at different times, and they were allowed to leave the test as soon as they had finished. The time allowed for completing the test could also be modified for those students who needed individual test arrangements due to such reasons as a disability, illness or restriction of their functional capacity. In the Kappas! project, students could have their time extended by 50%.

The tests were carried out following detailed instructions issued by CAE. Before the data collection started, two similar training sessions on test implementation were organised for the contact persons in higher education institutions. Additionally, a written guide for test proctors and a letter template used to invite students to the test were sent to the institutions. The higher education institutions completed the practical arrangements and advance preparations for their test sessions independently, including booking the necessary rooms, coordinating and supervising the test sessions, and inviting students to take the test. The practices used by different higher education institution varied. Some institutions also provided different external incentives, including lunch vouchers, cinema tickets, gift cards or lotteries for smartphones to reward the students for their participation.

In technical terms, the tests were implemented quite successfully. The national project team and CAE as the international coordinator provided support for higher education institutions where necessary during the testing. While the electronic test system worked faultlessly most of the time, some technical challenges also emerged during the testing. For example, in a few cases the connection to the test was cut off suddenly for one reason or another, or the test froze while the student was working on it. However, in these cases the students were mostly able to return to the test, either immediately by logging in again or at the latest within a week, and continue from the point where they were when the incident occurred.

Some problems related to the test and test system also had to be tackled together with the institutions' IT support and/or the test developer. Individual problems were related to such issues as the functionality of the software used for the test in the higher education institutions' computers. One more extensive problem that affected several higher education institutions also occurred in the test, as a result of which students were unable to respond to the background questionnaire at the end of the test. The problem did not affect the actual test, and it was fixed relatively soon. While the higher education institutions also provided students with opportunities to come back and complete the test, the responses to the background information questions were ultimately not received from many students who had encountered this problem.

In a few cases, an individual student's need for additional time for taking the test emerged too close to the test session or only after the test had been completed. It was not possible to take this into account later, however, for example in the scoring of the student's responses in the performance task. Many issues associated with the test arrangements were easy to solve as they were of a highly practical nature and related to the administration and supervision of the tests. All in all, few students in different higher education institutions were completely unable to take the test due to various technical problems.

3.5 Scoring of performance task responses

In late 2019 and spring 2020, students' responses in the performance task were scored as set out in the scoring rubric (Appendix 2) on a scale of 1 to 6 in three areas: analysis and problem solving, writing effectiveness and writing mechanics. Each performance task was scored by two trained scorers. The students consequently scored at least three and at most 18 points. If there was a difference of more than two points between the scores given by the two scorers, the response was assessed by a third one to ensure consistency. If a student's response could not be scored, for example because it was very unclear or irrelevant to the topic, the score was 0. The scoring was carried out in a secure network

interface, and it took place in a random order. The electronic system allotted the responses to be assessed by each scorer.

In September 2019, CAE trained seven people to score the performance task responses. During their training, they assessed authentic student responses in English, Finnish and Swedish under the direction of a CAE trainer. The training familiarised the scorers with the elements they were expected to assess in the students' responses and how to do it. The objective of the training was to ensure that the scorers understood the contents of the performance task and scoring rubric, learned to use the rubric consistently when scoring students' responses, and were able to independently score responses in each skill category.

To ensure consistent scoring, 50 Finnish and 30 Swedish responses were scored in advance (as verification responses). After the training, each scorer was asked to assess ten responses in the scoring system, ensuring that the scores were within two points of the pre-defined scores. Only then was the scorer allowed to assess students' responses independently. During the scoring process, the verification responses also ensured that the scoring was consistent. The system randomly assigned these responses to the scorers to assess. If the scores given for two consecutive verification responses deviated from the pre-defined scores by more than two points, the system was locked and the scorer had to complete the assessment of ten verification responses again. The scoring was supervised by the lead scorer (Kari Nissinen). For more information on the scoring, see the attached scoring rubric (Appendix 2).

During the Kappas! project, the consistency of the performance tasks scores between the United States and Finland was further ensured by translating 20 Finnish responses into English. These responses were assessed by two US scorers. Correspondingly, 20 responses from the United States were assessed by two Finnish scorers. When the scores were compared, no significant difference was found between them.

3.6 Data sets and analyses

3.6.1 Data sets

National data

In total, 2,402 students from 7 universities of applied sciences and 11 universities participated in the study. Of these students, 1,538 (64%) were initial stage (first year) students and 864 (36%) were third year students (in the final stage of Bachelor's degree studies). The data contained 1,273 university students (53%) and 1,129 (47%) university

of applied sciences students. Compared to the student population in the whole country, university students are overrepresented in the data, as according to the statistical service Vipunen (2019), approx. 38% of Finnish higher education students study at universities and 62% at universities of applied sciences, taking into account students enrolled as present in a Bachelor's or a Master's degree programme. The over-representation of universities in the data set is due to the fact that a larger share of the country's universities than universities of applied sciences participated in the study.

Of the participating students, 156 (6.5%) took the CLA+ test in Swedish, while the remainder took it in Finnish. It should be noted that 42% (66) of the students who took the test in Swedish said that their mother tongue is Finnish (the mother tongue of 97% of the students who took the test in Finnish was Finnish). As the number of Swedish-speaking students was too small to make reliable conclusions, the scores are not compared between the language groups.

The data included 1,178 male and 1,158 female students in total, as well as 58 who did not wish to state their gender. In other words, 49% of the data concerns men and 48% women. Men accounted for 51% of university students and 47% of UAS students in the data. The data are not fully representative of the gender distribution, as according to Vipunen database, 55% of Finnish university students were women in 2019. The proportion of female students in universities of applied sciences was 52%.

Women were particularly underrepresented among final stage students: as few as 40% of the final stage university students in the Kappas! study were women, while this proportion for UAS students was 47%. The gender distribution of initial stage students in the data set was more similar to the national population: 53% of initial stage UAS students and 52% of initial stage university students were women.

The median age of the students in the data set was 23 years. The median age was 21 for initial stage and 23 for final stage students. On average, university students were younger than UAS students by one year. The median age was 21 years for initial stage and 23 years for final stage university students, while the median ages for UAS students were 22 and 24 years respectively. The fact that almost one out of four UAS students was aged 28 years or over is significant. While the proportion of this age group in universities of applied sciences was 27% for students in the final stage, 21% of initial stage UAS students were also aged 28 or over. Among university students, the proportion of this age group was as low as 9% (8% for initial stage and 10% for final stage students).

In the entire data set, 80% of the students had completed the matriculation examination. This proportion was 92% for university students and 66% for UAS students. On the other hand, 42% of UAS students had completed an upper secondary vocational qualification,

whereas this figure for university students was as low as 8%. However, one out of three of the UAS students who had a vocational qualification had also completed the matriculation examination. Eight per cent of the students in the data set had already completed some other higher education degree. The proportion of such students was similar in universities (9%) and universities of applied sciences (7%). The data set also contained a small number of students who had a prior higher education degree but who had not completed the matriculation examination.

For several reasons, the representativeness of the data varied not only by gender but also by field of study. Firstly, higher education institutions participated in the study on a voluntary basis. This is why the fields of study represented by the participating higher education institutions dominate the data set at the expense of those that did not participate. Secondly, due to constraints related to testing resources and other practical data collection arrangements, students from all fields of study at a participating (multidisciplinary) higher education institution could not be picked. Because of these constraints, the data collection aimed for the best possible representativeness at the national rather than at the institution level. This was achieved by only selecting students from a few different fields of study in multidisciplinary higher education institutions (section 3.3). The goal of aggregating the data from different higher education institutions was to obtain a national data set in which students from all fields of study would be represented in proportion to the national student population. In practice, however, the success of different higher education institutions in recruiting students for the study varied. Consequently, the fields of study available at the institutions that were the most successful in recruiting students came to be overrepresented in the data. The third reason, which is partly related to the above, was the varying participation rates of students in different fields of study in the higher education institutions. Aspects that had a major influence on this included fitting the test sessions in with the students' schedules and the contribution of teachers or persons responsible for the assessment to motivating the students. At universities, the most overrepresented fields were Services, Health or Welfare, and Engineering, Manufacturing, Architecture or Construction, whereas Agriculture as well as Social Sciences and Education were underrepresented. In universities of applied sciences, the most overrepresented fields were Science, Services and Agriculture, while Health or Welfare were the most underrepresented fields.

Reference data from the United States

Reference data collected in the United States were made available for the Kappas! project. This data comprised a total of 71,403 initial stage students and 54,340 final stage students who were in their fourth year in the US system. The reference data were collected using the US version of the CLA+ test, which contained the same elements as the CLA+ International test. However, a higher number of performance tasks and sets of selected-response questions have been used in the United States than in Finland.

3.6.2 Analyses

For each student who participated in the study, a so-called sampling weight was calculated, which was used to correct the distortions in the data at the analysis stage. This meant that students from the underrepresented groups were given a higher weighting in the analyses than students from overrepresented groups. When determining the sampling weights, account was taken of the student's gender and field of study, the higher education sector (university/university of applied sciences) and the year in which the student started their studies (2017/2019). The data concerning the student population were retrieved from the Vipunen database, and the weightings were determined to ensure that the composition of the data would correspond to the national distributions of gender, field of study, higher education sector and starting year. As the statistical service Vipunen only has two categories for gender, the sampling weights for those 58 students who did not wish to state their gender in the background information survey was determined on the basis of the field of study, higher education sector and starting year alone. All analyses discussed below, including such basic statistics as means, standard deviations and percentages with their standard errors and regression analyses, were calculated using these sampling weights.

When analysing the Kappas! data, the potential clustering of students by field of study should be taken into account. This clustering follows from the method of selecting students for the sample by education or degree programme. Clustering means that individuals studying in the same programme at the same higher education institutions are not necessarily independent of each other. Typically, a group of students studying in the same field is more homogeneous than a random group of students; students in the same programme are likely to be more similar to each other than students arbitrarily selected from different programmes. This phenomenon can be seen in the data as so-called positive intra-cluster correlation. If this is ignored, the standard errors and confidence intervals of the statistics calculated based on the data are underestimated. In practice, this means overestimating the accuracy of the statistics and, consequently, analysis results whose statistical significance is overstated. To address the intra-cluster correlation, appropriate statistical methods are needed, including hierarchical two-level models (in which the student is level 1 and the cluster is level 2) or survey analysis methods adapted to the properties of the cluster setting. Usual statistical analyses rely on an assumption of independent observations, which is why applying them to the Kappas! data is not advisable. The data analyses presented in this report (means, standard deviations, percentages, regression analyses) were produced using so-called design-based survey methodology. What is typical of them is that the standard errors of statistics, including means, percentages or regression coefficients, are calculated in ways derived from the characteristics of the sampling design. In the analyses of this report, all standard errors were computed using Taylor series approximation. This is a commonly used approach in analyses of clustered and sampling weighted data. The main computation tools were the

Surveymeans and Surveyreg procedures of the SAS statistics software. In the analyses, students who had selected the same field of study at the same educational institution were regarded as belonging to the same cluster.

Of the 2,402 students who participated in the Kappas! study, a total of 2,300 completed the CLA+ test in a manner that made it possible to determine their total scores for the performance task, which measured generic skills, and for the selected-response questions. A performance task score could be calculated for 2,379 students and a selected-response task score for 2,315 students. The scores were converted to standard scales derived for the CLA+ test based on reference data concerning US students who completed the test in 2013–2018. The same standard scales were applied equally to all students, regardless of which tasks they completed. This was made possible by the equating procedure. Regardless of which PT or SR sets that were administered, equating places student scores on the same scale. The midpoint of the performance task scale is approximately 990 points, while its range is approximately 510–1,470 points. These are rounded figures. The midpoint of the scale for the selected-response questions, on the other hand, is approx. 1,090 points, and the range is approx. 550–1,630 points. The student's total score is the mean of the scores for the performance task and the selected-response questions. The midpoint of this scale is 1,040 points and the range 530–1,550 points.

The sub scores for the selected-response questions (critical reading and evaluation, scientific and quantitative reasoning and critiquing an argument) were calculated by applying the difficulty level of the questions to the number of correct answers, after which the scores were converted to a standardised scale whose average is 500 and standard deviation 100. On this scale, the scores typically vary between 200 and 800 points. In order to get a score for this section, the student had to answer at least one half of the questions. The score for the selected-response section was obtained by combining and rescaling these partial scores. In this process, the subscores are weighted. The SQR and CRE sets, both have 10 items (1 point per item, 10 possible points per set), have a higher weighting than the CA set, which has 5 items (1 point per item, 5 possible points).

Based on the total score in the CLA+ test, the students' level of mastery in generic skills was classified into five levels: advanced, accomplished, proficient, basic and below basic. The cut-off points of mastery levels were defined based on the US reference data, and the score limits are used to report CLA+ results internationally. For the distribution of mastery levels in the reference data, see Table 5 in Chapter 4; for a description of the contents of the mastery levels, see Appendix 3.

3.7 Test reports for students and higher education institutions

3.7.1 Reports for students

The students received a personal report on their test scores by e-mail (Appendix 4). The results were reported to the students both as mastery levels (Appendix 3) and as so-called percentile scores. The percentile score describes the student's ranking and indicates how well the student performed compared to other Finnish students and to students in their own field of study who took the test. The students received feedback not only on their overall performance but also separately on each area of assessment: analysis and problem solving, writing effectiveness, writing mechanics, scientific and quantitative reasoning, critical reading and evaluation, and critiquing an argument. The report included instructions for interpreting and using the test results and definitions for the levels of mastery. In addition, the covering letter reminded the students that many factors may have affected the test results besides their skills, including the situation in which the test was taken and the amount of effort the student put in. It was also explained to the students that any challenges related to learning, including dyslexia, were not taken into account in the assessment of the responses. In addition to the test results, the students were provided with additional information about generic skills and tips for developing them (Appendix 5). The highest-performing students (with at least proficient mastery level) also received a digital 'CLA+ badge' which, should they wish, they can add to their CVs or similar. The students received their reports in the language in which they took the test.

3.7.2 Reports and webinars for higher education institutions

The higher education institutions also received reports on their test results in English (Appendix 6). In these reports, the students' test results were described as scores and mastery levels. The test results were reported separately for initial stage and final stage students. In the aggregate report received by the institutions, singling out the results of an individual student was not possible. The report compared the institution's results to the overall results in Finland and to the US reference data. The covering letter of the report reminded the higher education institutions that the participants were selected to obtain a sample that was as representative as possible at the national level in each field of study. This is why the results of an individual institution are not representative and should not be used for comparisons between higher education institutions. The appendices at the end of the report described in greater detail the test process and the different areas of the test, among other things. Each higher education institution received its own, anonymised data. A two-hour webinar was also organised for interested higher education institutions

(n = 12). At the webinar, a short introduction to the Kappas! project was provided, and the institution's Kappas! results were discussed. An introduction to developing generic skills as part of higher education studies was also given at the webinar. In the group work which concluded the webinar, the participants focused on considering how generic skills could be addressed better in the higher education institution's degree programmes, curricula and teaching.

3.8 Evaluation of the reliability of the study

In the different stages of the Kappas! project, an effort was made to maintain a high scientific standard and, where necessary, to improve the data collection and research quality. For example, international standards were complied with in the translation, adaptation and pre-testing of the CLA+ research instrument (Hambleton & Patsula 1998; Ercikan & Pellegrino 2017), different mechanisms were used in the scoring of the performance tasks to ensure consistency (section 3.5), and appropriate design-based methods were used to analyse the data (section 3.6.2). When interpreting the results, however, the following limitations should be taken into account:

Limitations related to skill categories. While the project focused on key generic skills from the perspective of higher education and continuous learning (analytical reasoning, problem solving, writing effectiveness and writing mechanics), it should be emphasised that the skill categories selected for the project only comprise one generic skillset. There are many types of generic skills (Chapter 2), and the ones studied in the Kappas! project do not represent the full range which, for instance, includes interaction and intercultural skills. Consequently, the results of the project cannot be generalised and applied to all generic skills.

Limitations related to the study design. A cross-sectional design was used in the Kappas! study: data were only collected once, and as they were gathered separately from initial and final stage students, these two groups represented different cohorts. This should be noted especially when looking at the development of generic skills as part of higher education studies (section 4.3). A longitudinal study design would naturally provide a more reliable answer to this question, but the duration of the Kappas! project did not allow longitudinal research.

Limitations related to data and data collection. As the higher education institutions were involved in the project on a voluntary basis and 18 of them participated in it (there was a total of 36 higher education institutions in Finland at the time the data collection took place), a fully randomised sample of students among the entire national population could not be selected. The data collection was additionally labour-intensive, and student

input of around two hours was required to take the test. Participation in the study was also voluntary for the students: this was important, among other reasons because participation on a voluntary basis increases the likelihood of the students taking the test seriously. All these factors affected the participation rate, which was as low as 25%. While the participation rate can be considered reasonable for a study of this type, it also means that the data have certain limitations from the perspective of non-response. An effort was made to take these challenges related to data representativeness into account in the analyses as far as possible, however, for example by calculating sampling weights. Additionally, including a higher number of tasks selected at random could have improved the reliability of the research instrument, especially from the perspective of generalising the measured levels of mastery and using the reference data. Due to time constraints, careful translation, adaptation and pre-testing of more tasks were not possible. The background information was additionally obtained from the participating students themselves. For example, the students' matriculation examination grades were based on the students' reports rather than retrieved from the matriculation examination register. This approach naturally somewhat increases the possibility of errors and incorrect information.

Limitations related to the reference data. While this report compares the results of Finnish higher education students to their American peers, the reference data available for the Kappas! project were not intended for comparisons of generic skills levels between Finnish and American students as such. Above all, the purpose of the reference data was to reflect the way in which information on generic skills has been obtained in a different higher education institution context using the same method. In other words, this was about comparing the outcomes produced by one higher education system to those produced by another in the spirit of benchmarking (see Karjalainen 2002). The data sets are also different, and they were collected at different times. Comparability between the Finnish data and the reference data is undermined by the fact that the Finnish data are only based on a single performance task and a single set of selected-response questions, which were completed by all the students included in the sample, whereas the reference data are based on an extensive collection of performance tasks and selected-response questions with varying levels of difficulty. While the scores given for different tasks can be converted to a common scale by means of equating, differences in the tasks' level of difficulty affect the scaling. The selected-response questions used in Finland, on the other hand, were of medium difficulty and allow better comparability between the Finnish and US data.

Limitations related to the test system. The online test system maintained by the CAE through its subcontractor placed restrictions on test implementation. The system only allowed a certain number of background questions in a specific format (with a limited number of answers, for example). Due to the constraints of the system, responding to the first 17 background questions was mandatory for the students (questions set by CAE) and

the remaining 20 questions were voluntary (nationally tailored questions). In addition, random alteration of the two different types of tasks (performance task versus selected-response section) and the three sets of questions in the selected-response section was not possible in the system. This may have affected the scores for the last section, in particular (critiquing an argument), for example as the student ran out of time or became fatigued – albeit that some students on their personal initiative answered the sets of questions in the selected-response section in a different order than that which the system offered.

4 Higher education students' generic skills

Kari Nissinen, Jani Ursin, Heidi Hyytinen & Katri Kleemola

- The generic skills of 59% of university students were basic or below basic, whereas few students had advanced generic skills.
- University students had a higher level of mastery in generic skills than UAS students.
- Of initial stage UAS students, 62% had at most basic level generic skills. Of final stage university students, on the other hand, 28% had at least an accomplished level of generic skills.
- Women did well in the performance task and men in the selected-response questions.
- There were no systematic differences between the fields of study.
- The student's previous educational background was associated with the mastery level of generic skills: success in the mother tongue test in the matriculation examination, in particular, explained strongly a high level of generic skills.
- Higher education students' socioeconomic background was also associated with the mastery level of generic skills. In particular, this was true for the number of books in the childhood home, which reflects a culture of reading and learning and which strongly explained the student's levels of mastery.
- The students' attitudes to the test were associated with their levels of mastery. Especially the effort the student put in when taking the test affected their success.
- When comparing Finnish students' levels of mastery to the US reference data, we notice that a large proportion (40%) of Finnish students had basic level skills, whereas the students in the reference data were more evenly distributed across different levels of mastery.

4.1 Mastery level in generic skills

4.1.1 Higher education students' mastery of generic skills

This section examines higher education students' mastery of generic skills by score and level of mastery (Appendix 3) in the entire data set and by stage of studies and higher education sector. Finnish students' scores are also compared to reference data from the United States.

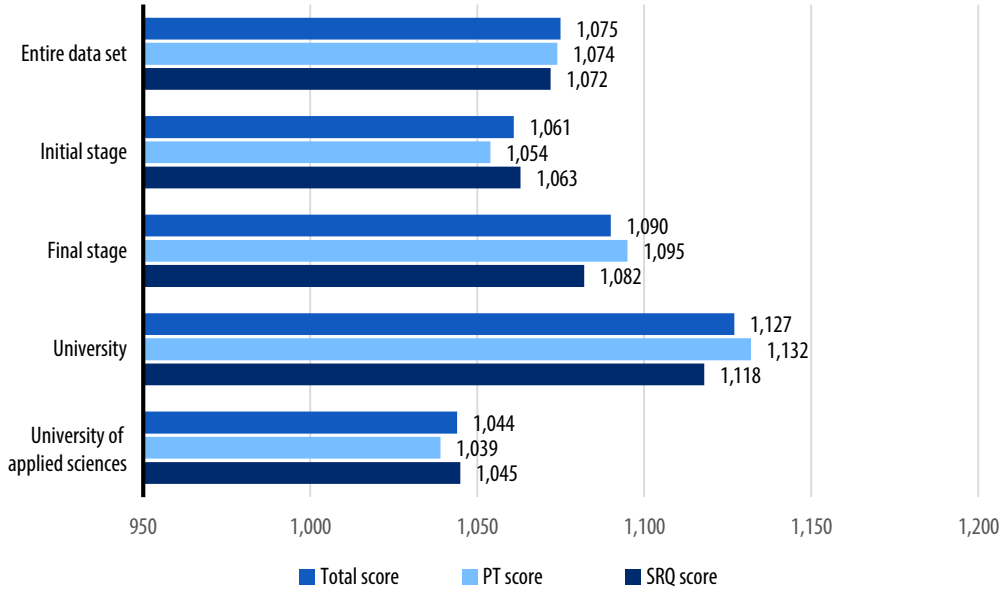
In the entire data set, the total CLA+ mean score was 1,075 points and the standard deviation 118 points (Figure 2). The mean scores for the performance task and selected-response questions were similar: on average, students scored 1,074 points in the former and 1,072 points in the latter. The standard deviations were 133 points and 159 points respectively (Table 1, Appendix 7). The mean score for initial stage students was 1,061 points, while the standard deviation was 117 points. In the performance task, their mean score was 1,054 points and average deviation 129 points, whereas in the selected-response section, these figures were 1,063 and 158 points. The corresponding means for final stage students were 1,090, 1,095 and 1,082 points, and the standard deviations were 116, 135 and 158 points respectively. A difference of 29 points in the total score, 41 in the performance task and 19 in the selected-response questions was observed between initial stage and final stage students. The differences in the total score and the performance task were statistically significant. There was no significant difference between the standard deviations for initial stage and final stage students; in other words, the variations in the scores obtained by initial stage and final stage students were of the same magnitude. As a whole, the students' scores were at a similar level in the performance task and the selected-response section: there were no statistically significant differences in the mean scores. This was equally true for the entire data set, the two higher education sectors as well as initial stage and final stage students.

The mean scores in the US reference data were 1,056 points for initial stage and 1,110 for final stage students. In the reference data, the mean score for initial stage US students was 1,049 points in the performance task and 1,059 points in selected-response section, while the average total score was 1,056 points. The corresponding standard deviations were 170, 186 and 149 points. The mean score of final stage US students was 1,102 points in the performance task and 1,113 points in the selected-response questions, while their average total score was 1,110 points. The standard deviations were 171, 183 and 148 points. The average difference between initial stage and final stage students in the reference data was 53 points in the performance task, 54 in the selected-response section, and likewise 54 in the total score. Consequently, a difference of slightly over 50 points can be considered a reference value which describes the expected improvement in higher education students' generic skills during approximately three years of study as measured with the CLA+ test.

While there was no statistically significant difference between the average total scores of initial stage Finnish students and those in the reference data, the average total score for final stage students was statistically significantly lower than the corresponding average in the reference data. The same observation can be made regarding the scores for the selected-response questions. On the other hand, there was no statistically significant difference in the mean scores between final stage Finnish students and the reference data in the performance task. It should be noted, however, that the performance task used in Finland was relatively easy; had the task been more demanding, Finnish students' scores might have been better than the ones now recorded.

The standard deviations in Finnish students' total scores and performance task scores were smaller than those in the reference data, whereas in the SRQ scores, the deviation was larger for Finnish students than those in the reference data. In other words, Finnish students' total scores and the points scored for the performance task varied less, and the points scored for the selected-response section varied more, than among American students.

Figure 2. CLA+ mean scores of the participants by stage of studies, type of HEI and in total



Note: The mean Total score is higher than the mean subscores because of different number of observations in each category (see page 34).

There was a clear difference between the mean scores of university and UAS students (Figure 2). The mean score was 1,127 points for university students and 1,044 for UAS students. This difference of 83 points was statistically highly significant. The difference

was 93 points in the performance task and 73 in the selected-response questions, both in favour of university students. These differences were also statistically highly significant. Their magnitude suggests that there is a major difference between the student populations of universities and universities of applied sciences. This will be discussed later in the context of examining the background variables (section 4.2.4). However, there was no notable difference between the standard deviations of university and UAS students: for example, the standard deviation of the total score was 111 points for university students and 110 points for UAS students. This means that the variation in the total scores of university and UAS students was similar.

Table 4. Distribution of mastery levels of participating higher education students in the entire data set and by stage of studies and higher education sector (%)

	Advanced	Accomplished	Proficient	Basic	Below basic
Entire data set	0.2	10	31	40	19
Initial stage students	0.0	7	29	41	23
Final stage students	0.4	13	32	40	15
University students	0.6	22	39	32	7
UAS students	0	5	27	44	24

When the scores are examined by levels of mastery, we notice that few students reached the highest level (Table 4). With the exception of a single initial stage university student, all those who reached the advanced level were final stage university students about to complete a Bachelor's degree (Table 6). Over 40% of students reached at least a proficient level. Of UAS students, 24% only achieved the below basic level, whereas this proportion among university students was as low as 7%.

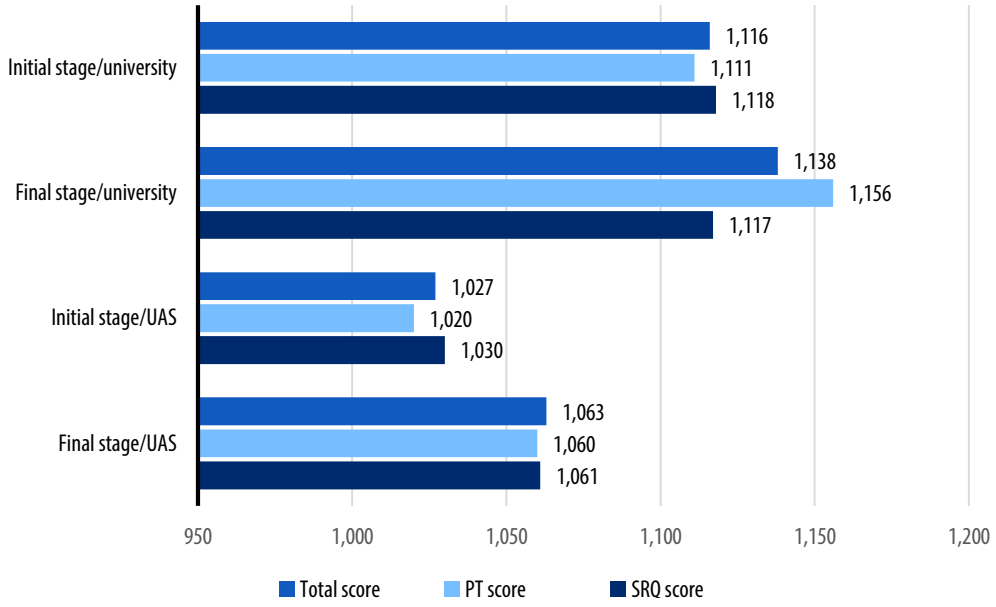
Table 5. Mastery level distribution in the CLA+ test in US reference data (%)

	Advanced	Accomplished	Proficient	Basic	Below basic
Initial stage students	2	13	26	32	28
Final stage students	3	20	31	28	17

When we compare the Finnish data to the US reference data (Table 5), we notice that there is less dispersion in the former. In the Finnish data, the majority (40%) of students reached a basic level of mastery, whereas in the reference data, the students were more evenly distributed between the different levels (the most common levels were proficient and basic, and also below basic for initial stage students). In the reference data,

2% of initial stage students reached an advanced level, whereas among the almost 1,500 students in the Finnish data, this level of mastery was only achieved by a single student in the initial stage. In the reference data, 15% of initial stage students reached at least an accomplished level, while this proportion in the Finnish data was as low as 7%. On the other hand, 23% of initial stage Finnish students were at the below basic level, whereas this proportion was slightly larger in the reference data, or 28%. Consequently, the distribution of Finnish students concentrates on the middle levels of mastery more strongly than in the reference data. For students in the final stage of Bachelor's degree studies, the differences between the Finnish data and the reference data were similar. In the reference data, 3% of students reached the advanced level and 23% reached at least the accomplished level, whereas the corresponding proportions among Finnish students were 0.4 per cent and 13 per cent. The proportion of final stage students who performed poorly was also slightly lower in Finland (15%) than in the reference data (17%).

Figure 3. CLA+ mean scores of initial and final stage students by the type of HEI



When we look at the mean scores of university and UAS students by the stage of studies, we notice that final stage university students achieved the highest scores as a whole (Figure 3). The difference between them and the other student groups examined in the study was particularly large in the performance task, in which they scored on average 45 points more than initial stage university students and 96 points more than final stage UAS students. These differences were statistically highly significant. In addition, initial stage university students' scores were statistically significantly higher (a difference of 91 points) than those of initial stage UAS students. The average difference between initial stage and final stage UAS students in the performance task was 40 points, but this difference was not statistically significant.

When the association of the stage of studies and the higher education sector with CLA+ scores was examined was examined using regression analysis, the higher education sector explained 11% of the variation in the total score and the performance task score, and 5% of the variation in points scored for the selected-response questions. The coefficient of determination of the stage of studies was essentially smaller: it only explained approx. 2% of the variation in the total scores and performance task scores, and less than 1% per cent of the variation in the SRQ scores. Combined, the stage of studies and higher education sector explained 13% of the variation in the total scores, 14% of the variation in the performance task scores, and approx. 6% of the variation in the SRQ scores.

Table 6. Mastery level distribution of initial stage and final stage students in the CLA+ test by higher education sector (%)

	Advanced	Accomplished	Proficient	Basic	Below basic
Initial stage university students	0	17	42	33	8
Final stage university students	1	27	35	30	7
Initial stage UAS students	0	3	24	45	29
Final stage UAS students	0	7	31	44	18

When we look at the mastery levels of university and UAS students, attention is drawn to the large proportion of initial stage UAS students with a below basic level of mastery (29%) and, on the other hand, to the fact that the proportion of final stage university students who have reached at least the accomplished level (28%) clearly stands out from

the other groups (Table 6). The proportion of both initial stage and final stage university students with at least an accomplished level of mastery corresponds fairly well to the reference data from the United States (Table 5). On the one hand, the share of those with below basic mastery is significantly smaller among Finnish university students than in the reference data. On the other hand, the proportion of below basic scores among Finnish UAS students is close to the corresponding shares in the reference data, whereas the proportion of accomplished and advanced scores is significantly lower than in the reference data.

4.1.2 Higher education students' generic skills by sub-area

In this section, we look at students' mastery of generic skills by skill category (Chapter 3, Table 1). The responses students gave to the performance task were assessed in three different areas: analysis and problem solving, writing effectiveness and writing mechanics. Each sub-area was scored on a scale of 1 to 6, in which level 1 represented a very poor response and level 6 an excellent one. The total score for the performance task consisted of the combined scores for the three areas, which were converted to the standard CLA+ test scales. The tasks of the selected-response section also measured three different skill categories: critical reading and evaluation, scientific and quantitative reasoning, and critiquing an argument. The results of these skill categories were converted to a scale where the average is 500 points and standard deviation 100 points. The tasks' levels of difficulty were taken into account in the scaling. The total score for the selected-response questions was obtained by combining the scores for these three skill categories and converting them to a standard scale.

See Tables 7 to 9 for the distribution of Finnish students at the different levels of the skill categories in the performance task. To facilitate comparisons, the Tables also show the corresponding distributions in the US reference data. Of the three skill categories measured by the performance task, Finnish students did the best in writing mechanics. The distribution of their success in analysis and problem solving was more or less similar than in writing effectiveness. The most typical level for Finnish students was 4 (excluding writing mechanics among university students, for whom the most common level was 5), whereas the most common level in the US reference data was level 3 in all skill categories of the performance task. However, this difference is at least partly explained by the fact that the performance task completed by Finnish students was somewhat easier than the

tasks completed by American students on average. In other words, the distribution of Finnish students is not entirely comparable with the reference data in this case.

Table 7. Distribution of initial and final stage higher education students' analysis and problem solving skills by higher education sector (%)

	Level 6	Level 5	Level 4	Level 3	Level 2	Level 1
Initial stage university students	3	29	56	10	1	0.2
Final stage university students	8	37	50	4	1	0
Initial stage UAS students	0.04	11	61	24	3	1
Final stage UAS students	2	18	56	21	1	1
Entire Kappas! data set	3	21	57	17	2	1
Reference data	1	3	23	47	23	3

Table 8. Distribution of initial stage and final stage higher education students' writing effectiveness skills by higher education sector (%)

	Level 6	Level 5	Level 4	Level 3	Level 2	Level 1
Initial stage university students	4	34	50	11	2	0.2
Final stage university students	7	36	53	4	0.2	0.2
Initial stage UAS students	0.1	12	57	26	3	1
Final stage UAS students	2	20	51	23	2	1
Entire Kappas! data set	3	23	53	18	2	1
Reference data	1	4	27	47	20	2

Table 9. Distribution of initial and final stage higher education students' writing mechanics skills by higher education sector (%)

	Level 6	Level 5	Level 4	Level 3	Level 2	Level 1
Initial stage university students	8	45	40	6	0.3	0
Final stage university students	16	49	32	2	0.2	0
Initial stage UAS students	2	23	55	18	1	1
Final stage UAS students	5	31	53	10	1	1
Entire Kappas! data set	7	34	47	10	1	0.4
Reference data	1	4	44	44	7	1

In all three skill categories measured by the performance task, final stage university students achieved the best total scores. Their scores were statistically significantly better in all skill categories than the scores of UAS students. The scores of initial stage university students were also statistically significant higher in all skill categories than the scores of UAS students. In the areas of analysis and problem solving as well as writing mechanics, the scores of final stage university students were statistically significantly better than the scores of initial stage university students, whereas the difference in writing effectiveness was not statistically significant.

Figure 4. Higher education students' mean scores in the skills measured by the selected-response section of the CLA+ test by higher education sector

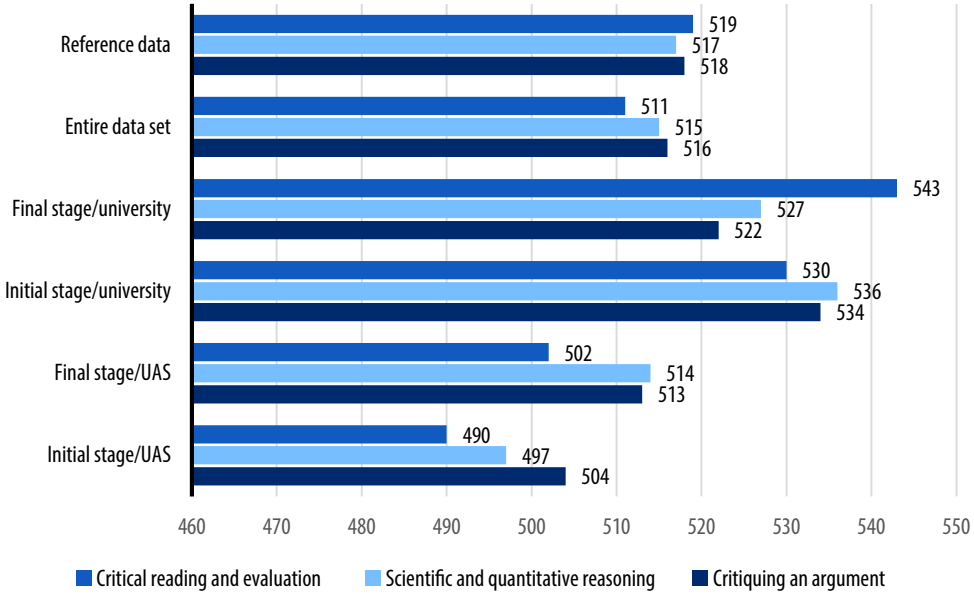


Figure 4 shows the mean scores for the three skill categories in the CLA+ test's selected-response section for Finnish students and in the reference data. As a whole, Finnish students' scores in these skill categories were at the same level as those in the reference data; the differences were not statistically significant. In critical reading and evaluation, Finnish university students' mean scores were statistically significantly better than those of UAS students; however, there was no significant difference between initial stage and final stage students in either group. In scientific and quantitative reasoning and critiquing an argument, the best mean scores were achieved by the group of initial stage university students. However, the differences between them and final stage university students were not statistically significant. The lowest mean scores were recorded for initial stage UAS students in all three skill categories of the selected-response section. The difference between them and initial stage university students was statistically significant in all skill categories. On the other hand, the difference between final stage university students and final stage UAS students was only statistically significant in the area of critical reading and evaluation.

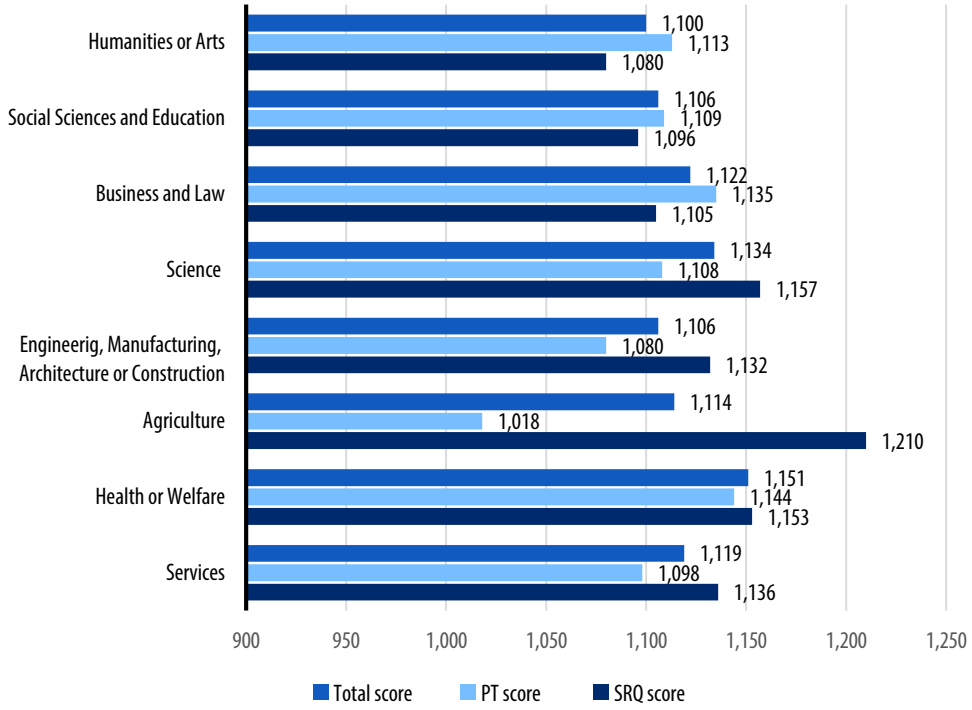
4.2 Associations between key factors and generic skills

4.2.1 Field of study and generic skills

The students were divided into eight fields of study in the data set. Due to the technical constraints associated with the questionnaire (section 3.8), the field of study information had to be collected using a US classification that does not fully correspond to the classification of education used in Finnish statistics. It groups education and social science fields together in the same category, and ICT is included in science. The same classification was applied to students from universities and universities of applied sciences. In this section, we look at differences in generic skills between students in different fields of study by higher education sector and stage of studies.

Figure 5 shows the CLA+ mean scores of initial stage university students in different fields of study, while Table 10 shows the distributions of mastery levels as percentages based on the total scores. The number of Agriculture students was very small in the data set, which is why no reliable conclusions can be made regarding this group. In the case of initial stage students, it should be noted that they mainly took the test at the beginning of their first academic year, which means that they had had little chance of completing any actual studies. This is why the differences between the fields of study describe, above all, the average differences in the initial level between students of different fields rather than the impact of studies in their fields on their generic skills.

Figure 5. CLA+ mean scores of initial stage university students who participated in the study by field of study



In some respects, the comparison of university students' fields of study led to different results depending on whether the total score, the performance task score or the score for the selected-response section was examined. Students of Science, Engineering, Manufacturing, Architecture or Construction and Services performed significantly better in the selected-response questions than in the performance tasks, whereas students of Humanities or Arts as well as Business and Law scored higher in the performance task (however, the difference to the scores for selected-response questions was not statistically significant in these fields). Students of Health or Welfare did well in the test, both in the performance task and the selected-response section. These differences at least partly reflect the gender distribution in the fields of study: the data indicate that men performed significantly better in the selected-response questions than in the performance task, whereas the situation was opposite for women. Science as well as Engineering, Manufacturing, Architecture or Construction were the most male-dominated fields of study in the data set. Another possible explanation is that some of the documents related to the selected-response questions could slightly favour students who studied Science or Medicine.

Some statistically significant differences were observed between the fields of study, although they depended on the score that was examined – the fields of study whose students did the best in the performance task and in the selected-response section were not the same in all respects. In the performance task, Health or Welfare students achieved the best mean scores. This result differed significantly from the scores achieved by students of Engineering, Manufacturing, Architecture or Construction and Services, who were the least successful of all fields of study in the performance task (the scores achieved by students of Agriculture in the performance task were also very low, but this finding was only based on the scores of seven students).

In the selected-response section, the best scores were achieved by students of Agriculture (although there only were seven of them in the data set), Science and Health or Welfare. The students of Humanities or Arts had the lowest mean scores, and the difference between them and students of not only Agriculture and Science but also Engineering, Manufacturing, Architecture or Construction and Services was statistically significant. There were no other statistically significant differences in the mean scores for the selected-response section.

As for the total score, students of Health or Welfare and Science had the highest means, while the lowest average levels were recorded for students of Humanities or Arts, Social Sciences and Education as well as Engineering, Manufacturing, Architecture or Construction. However, the differences in the total scores between the fields were fairly small and not statistically significant. The reason for this is that many fields of study that did well in the performance task did less well in the selected-response section and vice versa, and when the scores for the performance task and selected-response section are combined, the differences largely cancel each other out. As we have seen, initial stage students of Health or Welfare did well in both types of tasks. The total scores of students in the fields of Humanities or Arts as well as Social Sciences and Education were reduced by a poor performance in the selected-response questions, whereas the total scores of students of Engineering, Manufacturing, Architecture or Construction were affected by doing less well in the performance task.

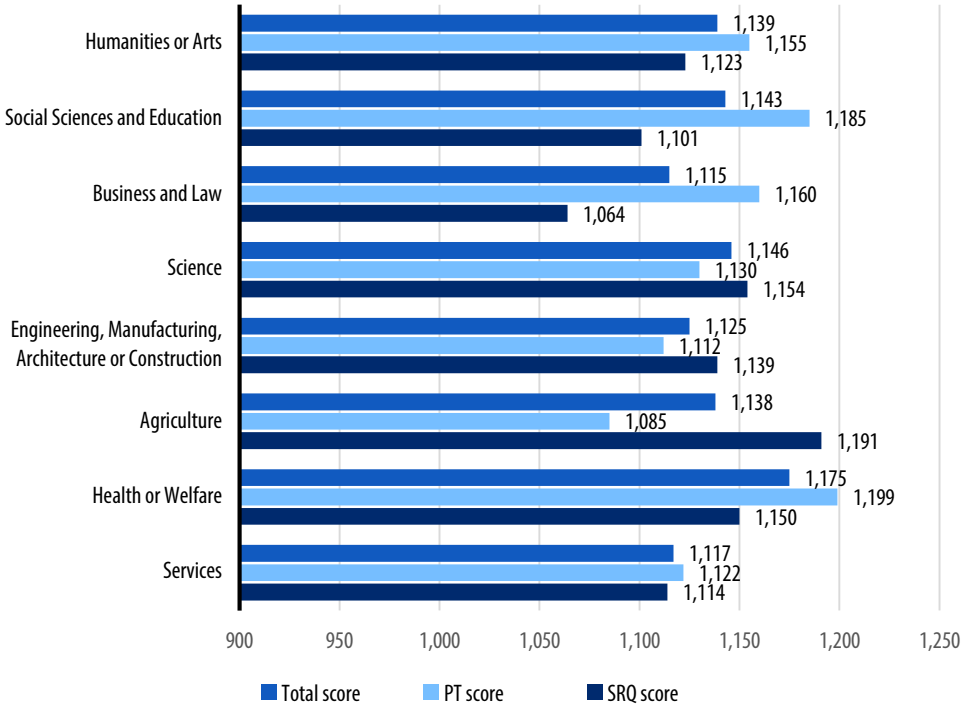
For initial stage university students, the field of study explained 2% of the variation in the total score, 3% in the performance task score, and 4% of the SRQ score. The good overall performance of initial stage university students of Health or Welfare is explained above all by the fact that almost one in four students showed accomplished skills and the number of low scores was small (Table 10). One in five students in the fields of Services and Social Sciences and Education showed accomplished skills but, on the other hand, the proportion of students below basic level was also high in these fields of study.

Table 10. Level of generic skills of participating initial stage university students by field of study (%). Number of observations = n

Field of study	n	Advanced	Accomplished	Proficient	Basic	Below basic
Humanities or Arts	100	0	10	40	43	7
Social Sciences and Education	148	0	19	33	37	11
Business and Law	103	0	15	50	31	4
Science (including ICT)	141	0	19	51	24	6
Engineering, Manufacturing, Architecture or Construction	151	0	14	44	31	11
Agriculture	7	0	12	35	53	0
Health and Welfare	49	0	23	43	29	4
Services	117	1	20	27	39	13
Initial stage university students, total	818	0.01	16	43	33	8

For the results of final stage university students, see Figure 6 and Table 11. The usability of the findings is in this case undermined by the fact that the numbers of observations are generally rather small. Only two students of Agriculture took the test, and they were excluded from the comparisons between the fields of study. Statistically significant differences in the averages between fields of study were only found in the performance task and selected-response questions, and similarly to initial stage university students, final stage students also performed differently in the two task types. When the scores for the performance task and selected-response questions were combined to obtain the total score, the observed differences between the fields of study again cancelled each other out, and there were no statistically significant differences in the total scores between the fields.

Figure 6. CLA+ mean scores of final stage university students who participated in the study by field of study



The best scores for the performance task were achieved by students in the fields of Health or Welfare as well as Social Sciences and Education (Figure 6). Their scores were statistically significantly better than the scores achieved by students of Science and Services, which were among the lowest in the comparison. In the selected-response section, on the other hand, the mean score of Science students was the highest of all fields of study, and statistically significantly higher than the scores achieved by students in the fields of Business and Law. The same phenomenon was already observed among initial stage students, and in this context, too, it can be explained by variation in the shares of men and women in different fields of study, and possibly also the topic of the selected-response section. For final stage university students, the field of study explained approx. 2% of the variation in the total score, 6% in the performance task score, and 3% in the SRQ score.

The groups in which at least accomplished level was the most common were students in the fields of Social Sciences and Education as well as Health or Welfare (Table 11). More than 30% of the students reached at least an accomplished level of mastery in these two fields of study, whereas one half of the students in the fields of Business and Law only reached a basic or below basic level of mastery.

Table 11. Mastery level distribution of final stage university students who participated in the study by field of study (%). Number of observations = n

Field of study	n	Advanced	Accomplished	Proficient	Basic	Below basic
Humanities or Arts	57	3	25	28	38	6
Social Sciences and Education	35	0	32	36	21	11
Business and Law	36	0	13	37	46	4
Science (including ICT)	78	1	22	48	23	6
Engineering, Manufacturing, Architecture or Construction	69	1	16	50	24	9
Agriculture	2	0	50	0	50	0
Health or Welfare	54	5	27	38	27	3
Services	123	0	7	55	33	6
Final stage university students, total	455	1	23	39	29	7

The differences between fields of study among initial stage UAS students were only statistically significant in the performance task, in which the mean scores of students in the fields of Health or Welfare as well as Humanities or Arts were better than the scores achieved by students of Engineering, Manufacturing, Architecture or Construction (Figure 7 & Table 12). In the selected-response questions, the highest mean scores were obtained by students of Science as well as Social Sciences and Education; however, the differences between them and other fields of study were not statistically significant due to the low number of observations. In this case, too, the differences between fields of study were not the same in the performance task and selected-response questions. While students of Science and Engineering, Manufacturing, Architecture or Construction did poorly in the performance task, they were among the top three fields of study in the selected-response section. In these two fields, the scores for the selected-response section were significantly better than the scores obtained for the performance task. For initial stage UAS students, the field of study only explained one per cent of the variation in the total score. The field of study explained 5% of the variation in the score for the performance task but only 0.3% of the variation in the selected-response questions.

Figure 7. CLA+ mean scores of initial stage UAS students who participated in the study by field of study

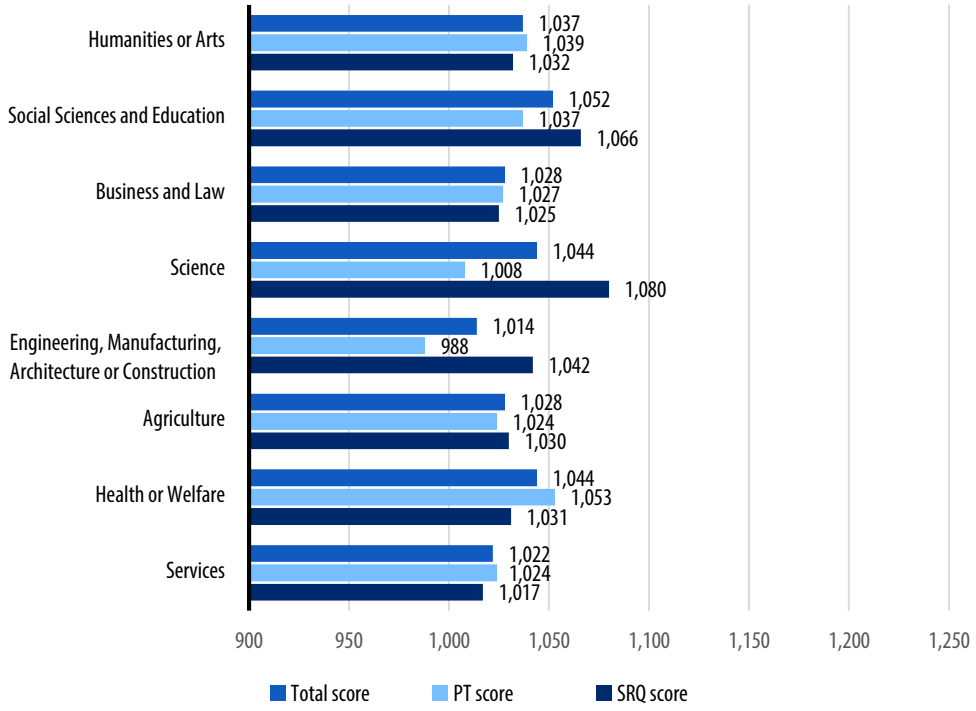


Table 12. Mastery level distribution for initial stage UAS students participating in the study by field of study (%). Number of observations = n

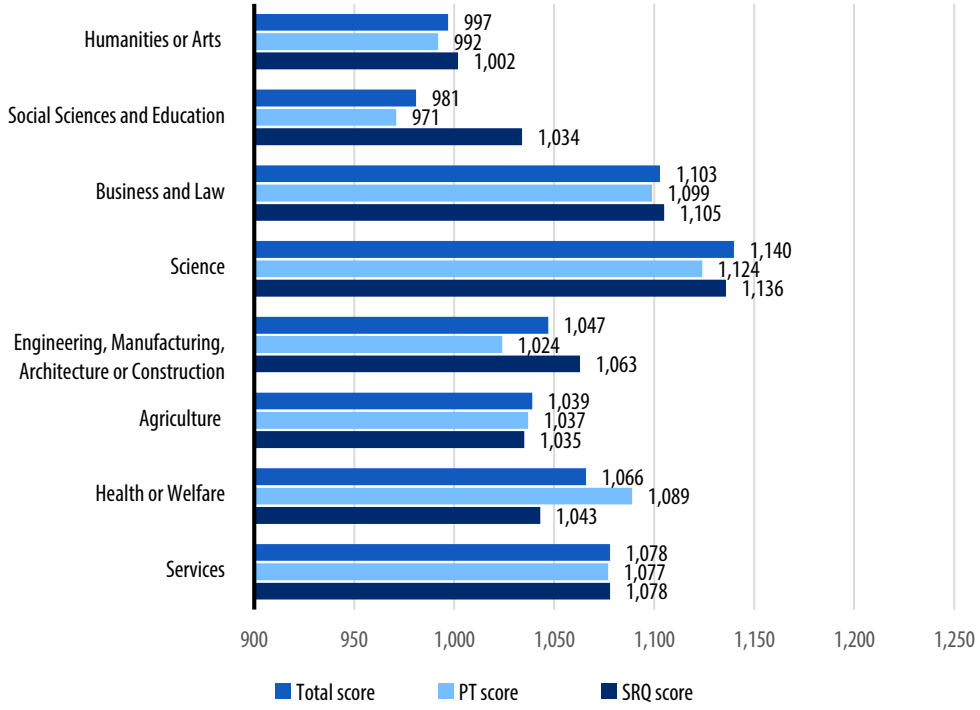
Field of study	n	Advanced	Accomplished	Proficient	Basic	Below basic
Humanities or Arts	41	0	8	16	54	22
Social Sciences and Education	16	0	0	41	35	24
Business and Law	82	0	2	24	42	32
Science (including ICT)	62	0	0	30	47	24
Engineering, Manufacturing, Architecture or Construction	204	0	1	20	51	28
Agriculture	35	0	0	37	33	30
Health or Welfare	137	0	4	28	41	27
Services	117	0	2	27	45	27
Initial stage UAS students, total	720	0	3	24	45	29

When we look at the scores of final stage UAS students by field of study, the small numbers of observations in several fields should be accounted for (Figure 8). In the fields of Social Sciences and Education, only four final stage UAS students took the test, and their scores were very low. The number of Humanities or Arts students who took the test was also small. These fields have been excluded from the comparisons between fields of study. Once again, students in the fields of Engineering, Manufacturing, Architecture or Construction had significantly higher scores in the selected-response section than in the performance task, whereas the opposite was true for students of Health or Welfare.

Science students scored the highest points in the data set among final stage UAS students (Figure 8). The number of observations concerning them is small, however, and as this was combined with relatively large variation in the scores, differences between Science and other fields of study could not be considered statistically significant regarding students' success in the performance task and selected-response questions.

If we exclude Science, the scores of students in the fields of Business and Law were the highest of the fields of study included in the comparison in the performance task, selected-response questions and total score alike. At the other extreme were found students of Agriculture and Engineering, Manufacturing, Architecture or Construction (when Humanities or Arts and Social Sciences and Education are excluded). Students in the fields of Engineering, Manufacturing, Architecture or Construction performed less well to a statistically significant degree than students of Business and Law in all scores of the CLA+ test. This time Health or Welfare students, who performed well in other student groups, did not stand out among the other fields of study by their good scores. In the performance task, the mean score of Health and Welfare students was roughly the same as the mean score of Business and Law students, but in the selected-response questions, the former did clearly less well than the latter. Consequently, a statistically significant difference in the total scores was also observed between Health or Welfare students on the one hand, and Business and Law students on the other, in favour of the latter group. For final stage UAS students, the field of study explained 4% of the variation in the total score, 6% in the performance task score, and 3% in the score for the selected-response questions.

Figure 8. Mean scores of final stage UAS students who participated in the CLA+ test by field of study



None of the UAS students achieved the advanced level in the CLA+ test (Tables 12 & 13). Among final stage students, the highest number of accomplished results was recorded in the field of Science (33%). On the other hand, almost one half of the students in this field did not perform beyond the basic or below basic level. Of Business and Law students, 12% reached the accomplished level. In the fields of study that had enough participants to enable a statistical analysis, poor performance was the most common among students of Engineering, Manufacturing, Architecture or Construction (22%).

Table 13. Mastery level distribution for final stage UAS students who participated in the study by field of study (%). Number of observations = n

Field of study	n	Advanced	Accomplished	Proficient	Basic	Below basic
Humanities or Arts	10	0	8	16	16	60
Social Sciences and Education	4	0	0	0	77	23
Business and Law	93	0	12	40	38	10
Science (including ICT)	21	0	33	19	44	4
Engineering, Manufacturing, Architecture or Construction	128	0	6	28	44	22
Agriculture	30	0	4	21	57	18
Health or Welfare	93	0	5	31	52	12
Services	22	0	5	35	42	18
Final stage UAS students, total	409	0	7	31	44	18

4.2.2 Student's age and generic skills

In this section, we examine the association between age and generic skills by stage of studies, separately for both higher education sectors. In this examination, initial stage students are grouped into five categories by their age: those aged under 20, 20 to 21, 22 to 23, 24 to 27 and over 27 years (Table 1, Annex 8). As the age distribution of final stage students is understandably slightly different, they are grouped into four categories: under 22, 22 to 23, 24 to 27, and over 27 years (Table 2, of Annex 8).

The most common age group for initial stage students was those aged from 20 to 21 years, who accounted for 44% of university students and 34% of UAS students. Of initial stage university students, 22% were under the age of 20, whereas only 12% of initial stage UAS students were in this age category. Similarly, 21% of initial stage UAS students were aged over 27; for university students, this proportion was as low as 8%.

The most common age group for final stage Bachelor's degree students was 22 to 23 years. This group accounted for 47% of university students and 36% per cent of UAS students. Of final stage students, 17% of university students but only 8% of UAS students were aged under 22.

More than one out of four final stage UAS students, or 27%, were aged over 27, whereas this age category accounted for 10% of university students.

Figure 9. CLA+ mean scores of initial stage university students in different age groups

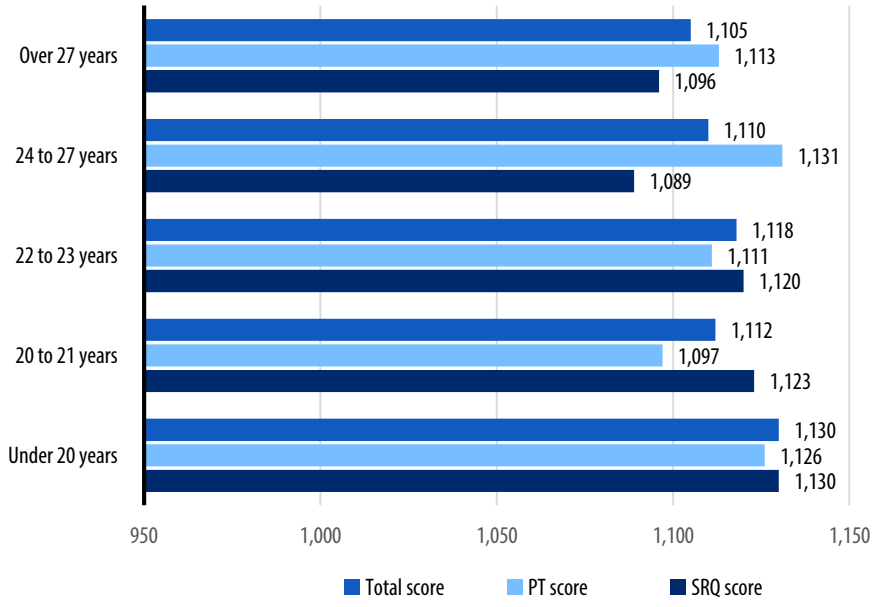


Figure 10. CLA+ mean scores of final stage university students in different age groups

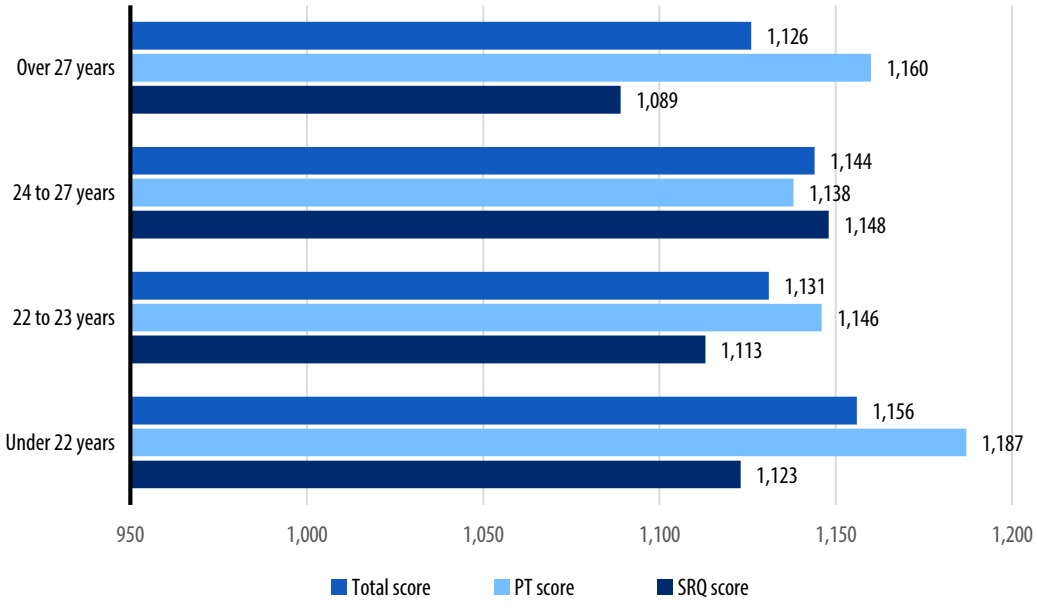


Figure 11. CLA+ mean scores of initial stage UAS students in different age groups

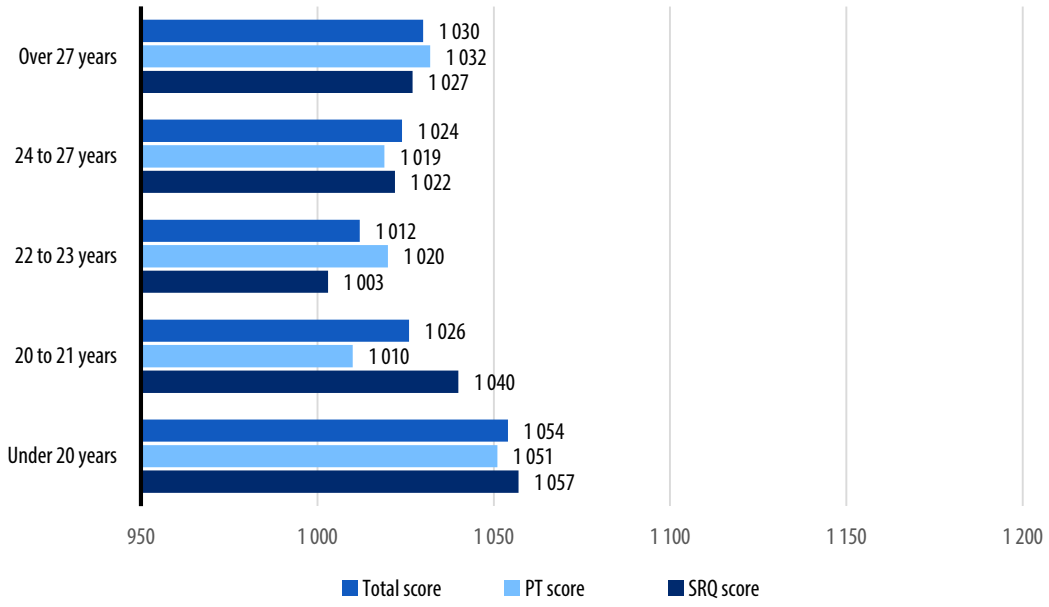
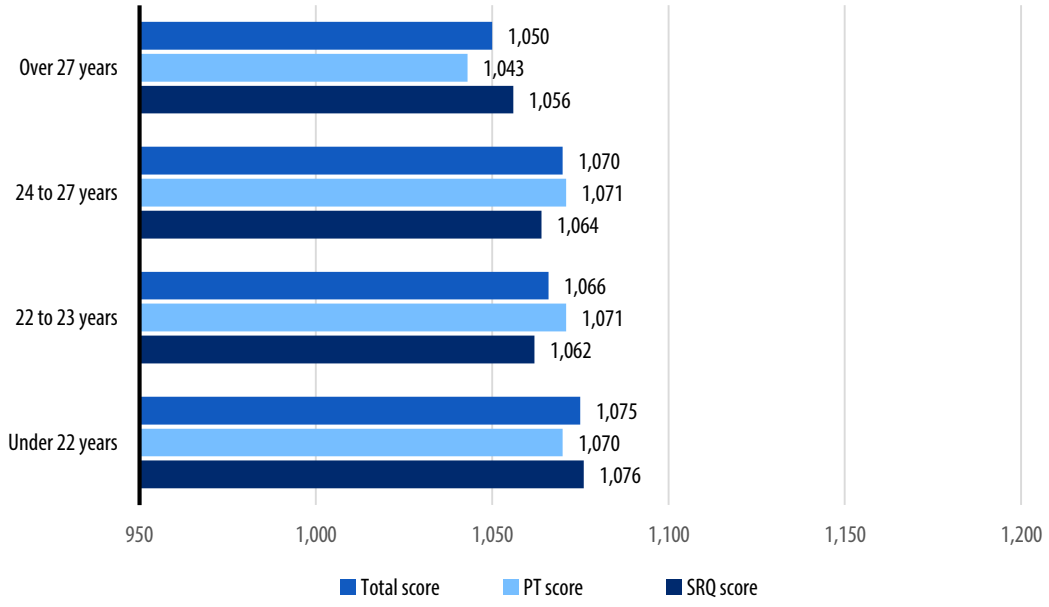


Figure 12. CLA+ mean scores of final stage UAS students in different age groups



Figures 9 to 12 show the mean CLA+ scores by age group for initial stage and final stage university and UAS students. As a rule, there were no statistically significant differences between the age groups, and the coefficients of determination were low throughout (0 to 3%). The only exception was initial stage UAS students (Figure 11). In this group, the

scores of those aged under 20 were statistically significantly higher than the scores of the next two age groups, or those aged 20 to 21 and 22 to 23 years. A watertight explanation for this difference is difficult to find. A possible explanation is, however, that a significantly higher number of students who had been awarded good grades in the mother tongue or mathematics test in the matriculation examination (M or higher) was found in the group of initial stage UAS students aged under 20 than in the other age groups. In addition, 74% of initial stage UAS students aged under 20 were women, whereas in the other age groups, the proportion of female students varied between 51% and 63%. This fact may be related to better results scored by students aged under 20 in the performance task.

4.2.3 Gender and generic skills

Regardless of the higher education sector and the stage of studies, gender seemed to be systematically associated with CLA+ scores. This association depended on the score that was examined. In the performance task, women achieved a mean score that was significantly higher than men, while this finding was reversed in the selected-response questions, however with a smaller difference in favour of men (Figure 13). When the scores for the performance task and selected-response questions were combined to obtain the total score, the gender differences almost cancelled each other out and a significant difference between men and women was no longer observed, excepting among final stage UAS students. A similar phenomenon was observed in both the university and UAS sector, and among initial stage and final stage students alike. The differences between fields of study in the performance task and SRQ scores discussed above are easier to understand when they are examined from the perspective of gender differences: the best scores in the performance task were obtained by students in the most female-dominated fields, whereas the best scores in the selected response questions were obtained by students in the most male-dominated fields.

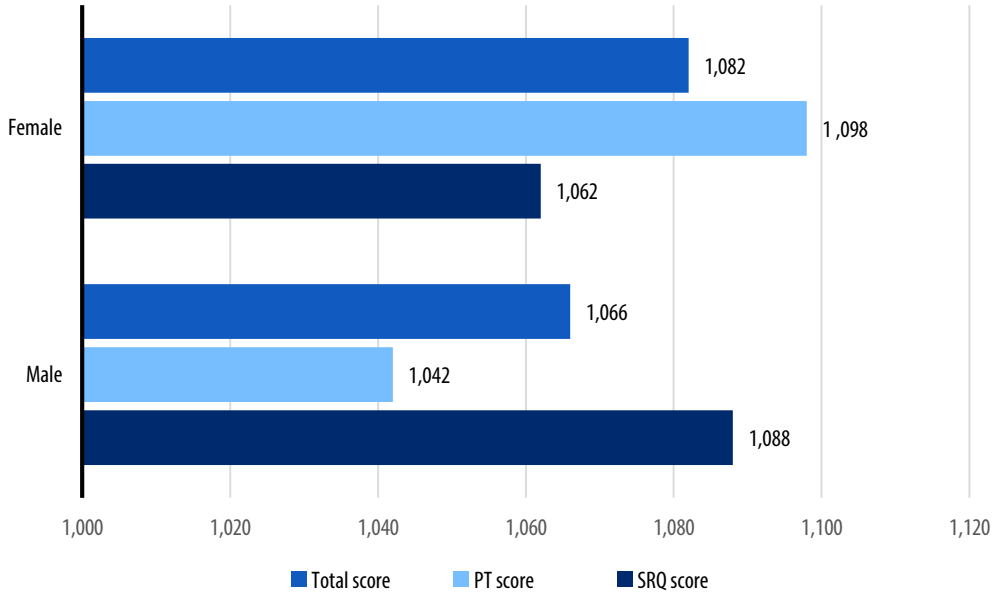
The magnitude and significance of the gender difference varied somewhat between the subgroups. In the performance task, the average gender difference was statistically significant in favour of women in all the subgroups examined, although the magnitude of this difference varied somewhat. The greatest difference of 66 points was observed among final stage UAS students, whereas the smallest difference of 36 points was recorded for final stage university students. On the other hand, the differences in the scores for the selected-response questions were the largest among final stage university students (38 points in favour of men, although the difference was not statistically significant) and the smallest among final stage UAS students (17 points in favour of men). Gender differences in the averages among initial stage students fell between these two extremes.

However, initial stage university students were the only subgroup in which a statistically significant difference between men and women was found in the selected-response questions (33 points in favour of men).

In the CLA+ mean scores, a statistically significant gender difference was only registered among final stage UAS students (23 points in favour of women). Among initial stage UAS students, this difference was 16 points in favour of women, whereas it was 7 points among initial stage university students. Neither of these differences were statistically significant. There was no difference between the total mean scores of women and men among final stage university students.

Figure 13 shows the CLA+ mean scores of men and women calculated for the entire data set. The graphs obtained for the sub-groups formed by stage of studies and higher education sector would be very similar, with the difference that the total score level is the highest for final stage university students and the lowest for initial stage UAS students. For the more detailed numerical results of the different subgroups, see Appendix 7, Table 2.

Figure 13. CLA+ mean scores of men and women in the entire data set



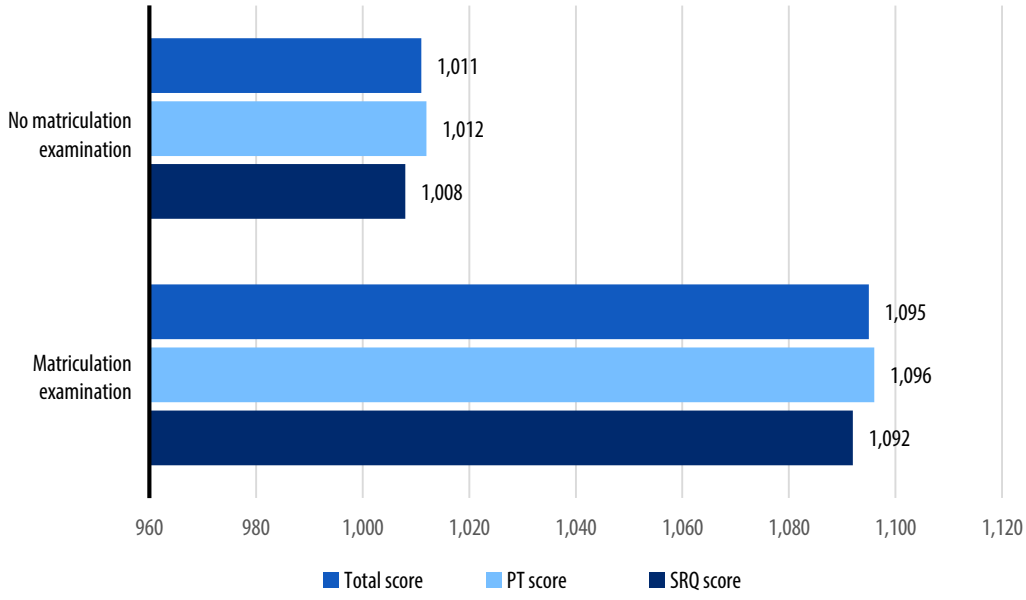
While a statistically significant association was found between gender and the student's success in the performance task and the selected-response questions, it did not explain a large proportion of the variation in scores. The coefficients of determination of gender were as low as in the range of one per cent, regardless of the score being examined. The only exception was the score for the performance task

among final stage Bachelor's degree students at universities, for whom this coefficient was slightly higher (3%) than for others. The low coefficients of determination indicate that the gender difference was very small compared to individual variation in the skills within the male and female student populations. This appeared to be equally valid for initial stage and final stage students as well as students in the different higher education sectors.

4.2.4 Educational background and generic skills

The information on educational background consisted of whether or not the respondent had taken the matriculation examination, how the respondent had succeeded in the mother tongue and mathematics tests in it, and whether the respondent had a previous degree or qualification. The proportion of students who had completed the matriculation examination in the data set was 80%. This figure was 92% for university students and 66% for UAS students. The CLA+ scores of students who had sat the matriculation examination were on average 84 points better than those of students who had not (Figure 14). This difference was statistically highly significant. Some of the average differences between university and UAS students can indeed be explained by a larger share of those who have completed the matriculation examination among university students.

Figure 14. Completion of the matriculation examination and CLA+ mean scores in the entire data set



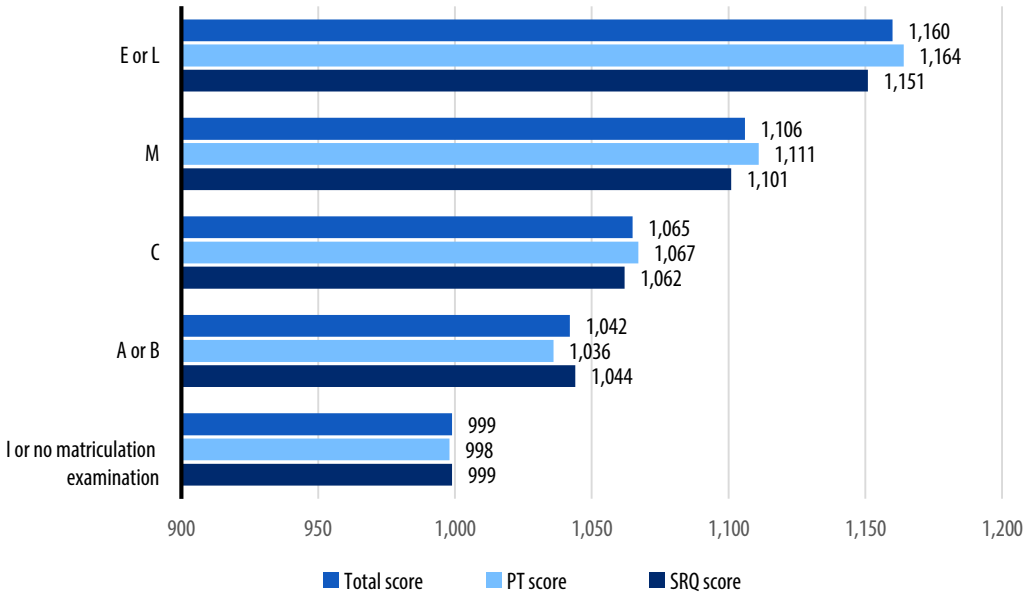
The results are very similar regardless of whether we look at initial stage or final stage students separately, or UAS students as a group. The differences in favour of those who had completed the matriculation examination were 61 to 88 points depending on the score examined, and in all cases statistically highly significant. The differences were smaller among university students, or around 40 points, and they were not statistically significant. However, this was mainly due to the small number of those who had not completed the matriculation examination among university students. In the entire data set, completing the matriculation examination explained 9% of the variation in the total score, 7% of the variation in scores for the performance task, and 5% of the variation in the scores for the selected-response section.

When we look at the association between the grade achieved in the mother tongue test in the matriculation examination and CLA+ scores, we observe that the CLA+ mean score increases in an almost linear fashion with the grade, and the differences between the groups were statistically significant (Figure 15). The students who had not taken the matriculation examination or whose mother tongue test had been failed had the lowest scores of all test participants. The mother tongue grade awarded in the matriculation examination was one of the strongest predictors of CLA+ scores among the background variables in the data set: it explained 22% of the variation in the total score, 20% in the performance task score, and even 11% in the score for the selected-response questions.

Of students who had completed the matriculation examination, 41% had taken the mathematics test in the advanced syllabus and 46% in the basic syllabus, whereas 14% had not taken the mathematics test at all. As many as 60% of university students had taken the advanced syllabus test in mathematics, whereas this rate was as low as 25% among UAS students. The proportion of those who had not taken the mathematics test in the matriculation examination was 11% for university and 16% for UAS students. Consequently, 59 % of UAS students who had completed the matriculation examination had taken the basic syllabus test in mathematics. At universities, their proportion was 29%.

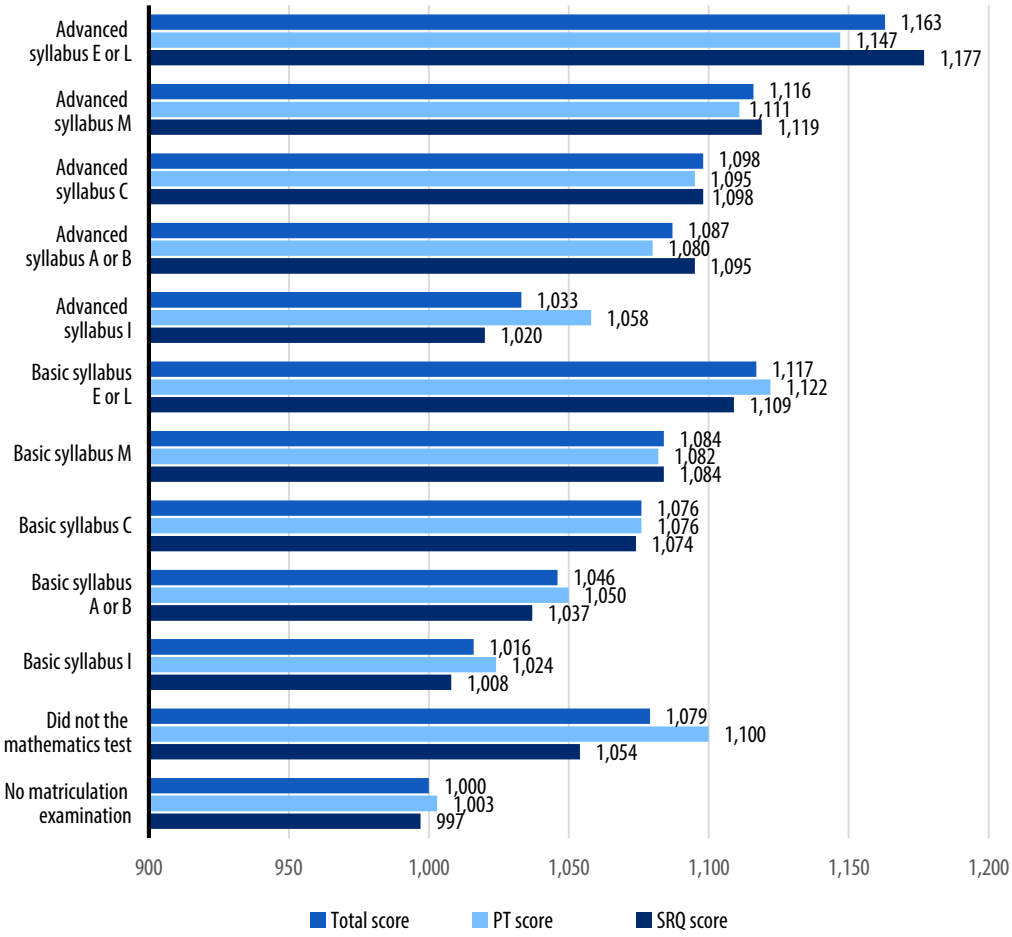
A significant proportion (32%) of those who took the advanced syllabus test in mathematics in this data set had received the grade eximia cum laude approbatur (E) or laudatur (L). Among those who took the basic syllabus test in mathematics, this proportion was 23%. On the other hand, 17% of those who took the advanced syllabus test in mathematics and 27% of those who took the basic syllabus test had received a grade below cum laude approbatur (C). There was a clear correlation between the mother tongue and mathematics grades in the data set: almost one half of those who had received at least the grade eximia cum laude approbatur (E) in mathematics (in either the advanced or basic syllabus) had also received the same grade or higher in the mother tongue test.

Figure 15. Grade in the mother tongue test in the matriculation examination and CLA+ mean score



When we look at the CLA+ mean scores itemised by the grade awarded to the student in the matriculation examination mathematics test, we observe that in this case, too, the CLA+ mean scores improve in an almost linear fashion with the matriculation examination grade (Figure 16). The best results were achieved by students who had received the grade eximia cum laude approbatur (E) or laudatur (L) in the advanced syllabus test of mathematics. The scores of those who had been awarded at least the grade eximia cum laude approbatur (E) for the basic syllabus in mathematics were in the same range with those who had received the grade magna cum laude approbatur (M) in the advanced syllabus test in mathematics. We can also note that the average success of students who had not taken the mathematics test at all in the matriculation examination was fairly good in the performance task. The mathematics grade received in the matriculation examination was also a strong predictor of the CLA+ scores. Taking into account not only the grade but also the syllabus in which it was obtained, 18% of the total score variation could be explained, or 12% of the variation in the performance task scores and 11% of the variation in the selected-response section scores.

Figure 16. Grade in the mathematics test in the matriculation examination and CLA+ mean score



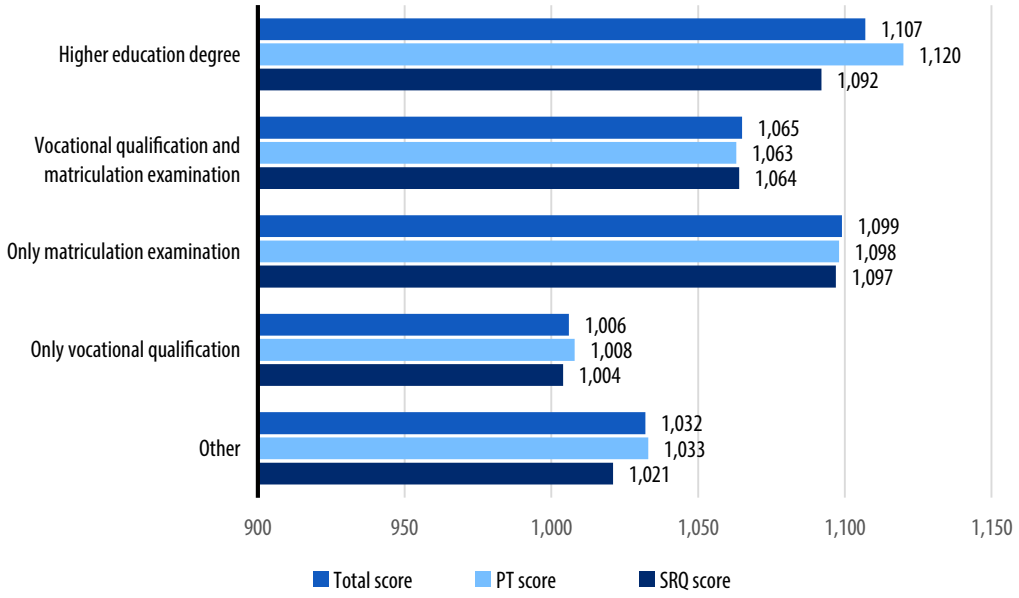
We noted above that 80% of the students in the data set had completed the matriculation examination. However, some of them also had some other degree or qualification, usually a vocational upper secondary qualification or a higher education degree. When we take a closer look at the students' qualifications, we notice that 56% of the students in the data set had only completed the matriculation examination. Approximately 18% had completed an upper secondary vocational qualification but not the matriculation examination, and 12% had completed both a vocational qualification and the matriculation examination. Of the participating students, 8% were completing their second higher education degree.

There was a clear difference between university and UAS students in the types of previous qualifications they had: 76% of university students and 45% of UAS students had only completed the matriculation examination, whereas 27% of UAS students but only 2% of university students had only completed a vocational qualification. Those who had

completed both a vocational qualification and the matriculation examination accounted for 16% of UAS students and 7% of university students. Of UAS students, 7% already had another higher education degree; this proportion of university students was 10%.

Students who had already completed a previous higher education degree performed the best in the CLA+ test, but the difference between them and those who had only completed the matriculation examination was not statistically significant (Figure 17). Those who had only completed a vocational qualification differed in their mean scores from the groups that performed better to a statistically significant degree, including those who had completed both a vocational qualification and the matriculation examination. Notably, the CLA+ mean scores of those students who had only completed the matriculation examination were better to a statistically significant degree than the scores achieved by the students who had completed both a vocational qualification and the matriculation examination. The difference may be explained by the different profiles of the two student groups. Those students in the data set who had only completed the matriculation examination were on average younger, they had on average received higher grades in the matriculation examination, and about one half of them studied at a university. Of those who had completed both a vocational qualification and the matriculation examination, 79% studied at a university of applied sciences.

Figure 17. Previous degree or qualification and CLA+ mean score



4.2.5 Socioeconomic background and generic skills

Two indicators describing the student's socioeconomic background were available in the data set: parental education and the estimated number of books found in the student's childhood home. Parental education was measured on a scale of six categories (Table 14) by asking the student to think of the parent or guardian the student felt had had the greatest influenced on him or her and to indicate the highest level of education this person had. When looking at the educational level distribution of higher education students' parents, we notice that university students' parents are more likely to have a high level of education than UAS students' parents. In this data set, 43% of university students' parents had at least a Master's degree, whereas this proportion for UAS students was 20%. On the other hand, 47% of UAS students' parents had completed at most a secondary level qualification, while this proportion among university students' parents was 27%.

Table 14. Distribution of parents' educational level for university and UAS students

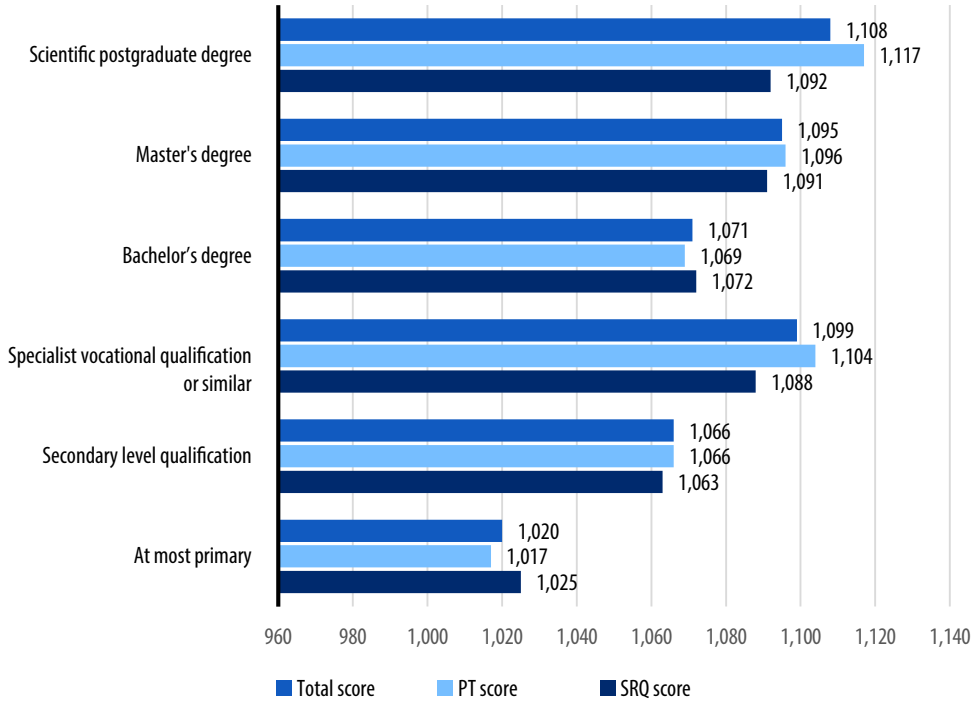
Level of education	Universities	Universities of applied sciences	All
	%	%	%
At most primary	5	12	9
Secondary	22	35	30
Specialist vocational qualification or similar	11	8	9
Bachelor's degree	19	25	23
Master's degree	37	18	25
Scientific postgraduate degree	6	2	4
Total	100	100	100

When we examine the association between parental education and the student's CLA+ scores, we observe that the students whose parents had completed at most primary level education stood out the most clearly from the other groups (Figure 18). Their mean scores were statistically significantly lower than those of all other groups. The differences between the other groups were relatively small, and only some of the differences in the mean scores for the performance task or total scores were statistically significant. The performance task scores received by students whose parents had a Master's degree or a specialist vocational qualification (or similar) were better to a statistically significant degree than the scores of students whose parents had a Bachelor's degree or a secondary level qualification. This also applies to the total score but not to the score for the

selected-response questions. The best total scores were achieved by the group with at least one parent who had a scientific postgraduate degree.

Due to the small size of this group, however, the differences between it and the other groups, apart from the group with the lowest level of parental education, were not statistically significant. Parental education did not explain the variation in generic skills to a particularly high degree in the data set: it explained 4% of the variation in the total score, 3% in the scores for the performance task, and only 1% in the scores for the selected-response questions.

Figure 18. Parental education and CLA+ mean scores



We noted above that there was a strong relation between the student's grade in the mother tongue test in the matriculation examination and the CLA+ test score. There is also a positive relation between parental education and the mother tongue grade in the data set. More than one out of three students whose parents had a scientific postgraduate degree had been awarded the grade eximia cum laude approbatur (E) or laudatur (L) in the mother tongue test. Of students in this group, 87% had completed the matriculation examination. At the other end of the scale are students whose parents had completed primary level education or at most a secondary level qualification, of whom 14% achieved the grade cum laude approbatur (C). The proportion of students who had completed the matriculation examination was also lower in this than the other groups, or 70%.

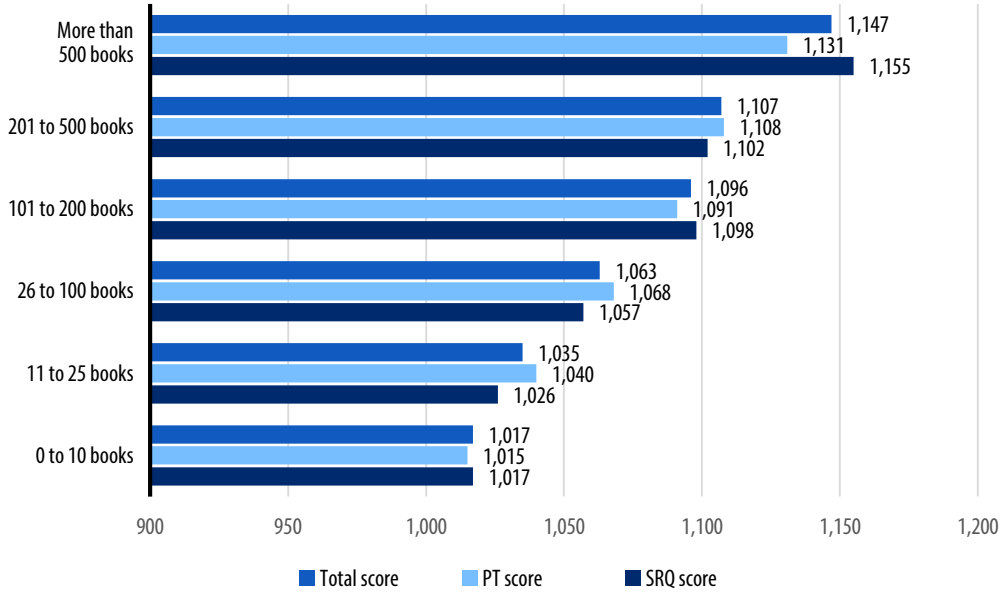
The number of books in the student's childhood home can be used as an indicator of a culture of reading and learning associated with the student's home background. Table 15 shows the distribution of the number of books in the childhood home in the data set. The data indicate that university students had on average a higher number of books in their childhood homes than UAS students.

Table 15. Distribution of the number of books in the childhood home among university and UAS students

	Universities	Universities of applied sciences	All
	%	%	%
0 to 10 books	7	11	9
11 to 25 books	10	15	13
26 to 100 books	23	35	31
101 to 200 books	27	22	24
201 to 500 books	25	13	18
More than 500 books	9	4	5
Total	100	100	100

The CLA+ mean scores improve in a fairly straightforward fashion as the number of books in the student's childhood home increases (Figure 19). This positive trend was statistically highly significant both regarding the total score, the score for the performance task and the score for the selected-response questions. The number of books explained 8% of the variation in the total scores and 5% in both the performance task and SRQ scores. Consequently, the association between the number of books and the CLA+ scores was slightly stronger than the association between parental educational and the scores. The number of books and parental education together explained 10% of the variation in the total scores, 7% in the performance task scores and 6% in the SRQ scores.

Figure 19. Number of books in the childhood home and CLA+ mean scores



The number of books in the childhood home also showed a clear positive correlation to the grade awarded to the student in the matriculation examination mother tongue test. Two thirds of the students who had had more than 500 books in their childhood homes received at least the grade magna cum laude approbatur (M) in the mother tongue test (35% received at least the grade eximia cum laude approbatur, or E). In this group, 94% of the students had completed the matriculation examination. Of those students who had no more than 25 books in their childhood homes, approximately 27% achieved at least the grade magna cum laude approbatur (M). Approx. 65% of them had completed the matriculation examination. We can consequently assume that the number of books in the childhood home has a significant indirect link with the student's generic skills; in particular, this link is underpinned by command of the mother tongue.

4.2.6 Students' attitudes towards the test and generic skills

The participating students were asked how engaging they found the tasks included in the test and how much effort they put in to complete the tasks. These questions were asked separately for the performance task and the selected-response questions, and the students answered them on a five-step scale. The higher the numerical values of the answers, the more engaging the student had found the tasks and the more effort they had put in to complete the tasks. The answers concerning student engagement in the performance task and selected-response questions were combined into a single score that indicates how engaging the student found the CLA+ test as a whole. The same procedure was followed for the variables measuring student effort.

The scores measuring student engagement in the CLA+ test were grouped into four categories (Table 16): (1) not at all or slightly engaging, (2) moderately engaging, (3) very engaging and (4) extremely engaging. The scores measuring student effort were also grouped into four categories (Table 17): (1) little or no effort, (2) a moderate amount of effort, (3) a lot of effort, (4) my best effort.

Table 16. Distribution of student engagement with the test among university and UAS students

	Universities	Universities of applied sciences	All
	%	%	%
Not at all or slightly engaging	7	16	13
Moderately engaging	30	47	40
Very engaging	51	33	40
Extremely engaging	12	5	7
Total	100	100	100

The distribution of student interest was rather symmetrical in the data; the majority of students found the test very engaging or moderately engaging. University students were more likely to find the test engaging than UAS students. There were no notable differences between initial stage and final stage university or UAS students.

Approximately four out of five students said they had made a lot effort or their best effort when completing the CLA+ tests (Table 17). One out of three university students reported they had made their best effort in the test. Making this type of major effort was more common among final stage than initial stage students, both in universities and universities of applied sciences. Final stage university students were the most likely to make their best effort (39%), whereas initial stage UAS students were the least likely to do so (11%). The proportion of students who said they made little or no effort in the test was as low as a few per cent in the data set.

Table 17. Distribution of student effort in the CLA+ test among university and UAS students

	Universities	Universities of applied sciences	All
	%	%	%
Little or no effort	2	3	2
Moderate effort	13	23	19
A lot of effort	52	59	56
My best effort	33	15	22
Total	100	100	100

Both finding the test engaging and making an effort in it had a straightforward and statistically highly significant association with the test scores (Figures 20 & 21). Experiencing the test as engaging explained 8% of the variation in the total scores, 7% in the performance task scores, and 4% in the scores for the selected-response questions. The coefficients of determination of student effort, on the other hand, were 9%, 9% and 4% respectively.

Figure 20. Student engagement in the CLA+ test and CLA+ mean score

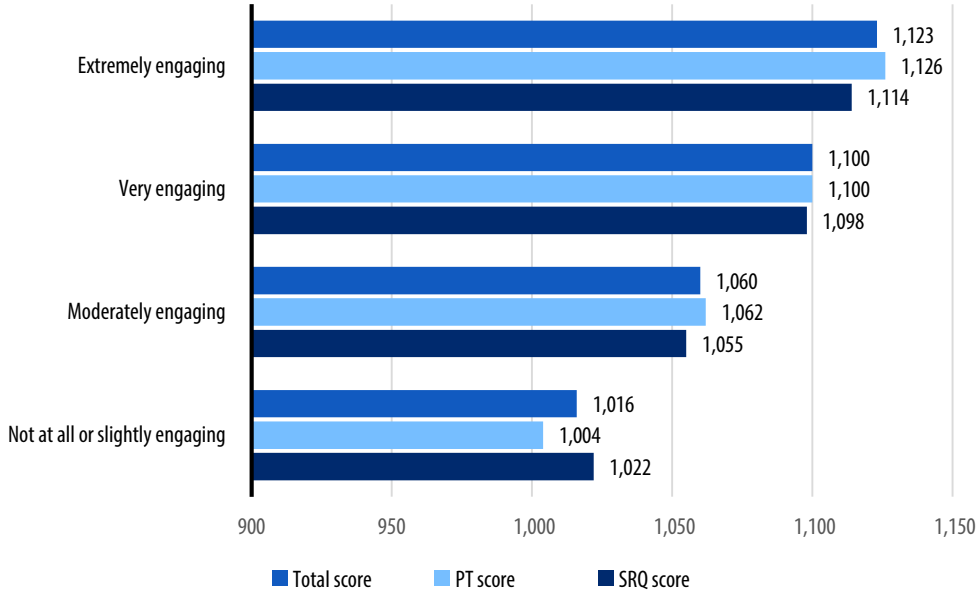
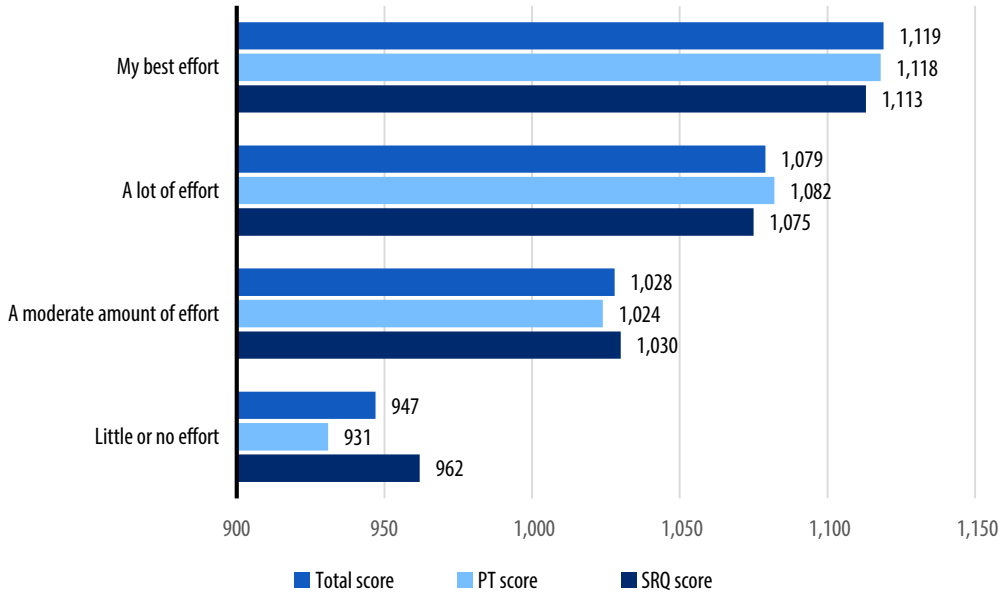


Figure 21. Student effort in the CLA+ test and CLA+ mean score



4.2.7 Strongest explanatory factors of higher education students' generic skills

Multivariate regression analyses of the CLA+ scores measuring students' generic skills (total score, performance task score and SRQ score) were conducted, in which all background variables examined in the previous sections were used as explanatory factors (stage of study, field of study, age and gender of student, family background, information on prior education, attitude to the CLA+ test). The multivariate regression analyses were used in examining the associations between the CLA+ scores and background variables when the background variables were controlled for each other (correlations between them were taken into account). In the previous sections, each background variable was examined separately and independently of any others. Controlling for other variables typically weakens the explanatory power of the variables, with the exception of the strongest explanatory factors. Of particular interest are the background variables which are statistically significant explanatory factors of generic skills after the statistical control. The analyses were carried out separately for university and UAS students, and all explanatory variables were treated as categorical, enabling the analysis to also bring up non-linear relations. Possible interactions between the explanatory variables were additionally examined during the modelling, but few of them were statistically significant. Significant interactions, on the other hand, were not meaningful in terms of interpretation or had no essential significance for the conclusions, which is why they were excluded from the final models.

Table 18 shows in a concise form the background variables which, based on the regression analyses, statistically significantly explained the variation in generic skills. The variables which were not statistically significant in any of the cases were left out. However, the stage of studies was retained in the model, whether it was statistically significant or not, as it serves as an indicator describing the development of generic skills during studies. For the more detailed numerical results of the regression analyses, see Table 3, Appendix 7 (university students) and Table 4, Appendix 7 (UAS students).

The CLA+ scores achieved by the students were the most systematically predicted by their mother tongue grade in the matriculation examination and the amount of effort they had put in when taking the test (Table 18). These were the statistically most significant explanatory variables of all CLA+ scores for both university and UAS students. The number of books in the childhood home explained the variation in CLA+ scores statistically significantly in other cases except in the performance task for university students. In this case, too, a variable measuring the student's socioeconomic background was related to the score achieved, albeit the number of books in the childhood home was replaced by parental education. The roles of the other tested background variables varied between the scores. Notably, differences between the fields of study and gender differences lose their statistical significance in many cases when the other background variables are controlled (mother tongue grade and student effort in the test are key factors here). The student's age, the mathematics grade in the matriculation examination, or whether the student took the test in advanced or basic syllabus in mathematics no longer were statistically significant explanatory variables in any of the models. The minor role of mathematics at least partly followed from its strong correlation with the mother tongue grade in the matriculation examination. Consequently, mathematics no longer had statistically significant explanatory power regarding generic skills once performance in the mother tongue test in the matriculation examination had already been introduced in the model. The regression models' coefficients of determination were higher for the performance task than for the selected-response questions, and higher for UAS students than university students. Especially variation in UAS students' performance task scores could be explained rather well by the background factors.

Table 18. Statistically significant factors explaining variation in CLA+ test scores in multivariate regression models

	University students			UAS students		
	PT score	SRQ score	Total score	PT score	SRQ score	Total score
Number of observations	n = 1,216	n = 1,175	n = 1,183	n = 1,049	n = 1,024	n = 1,042
Coefficient of determination	R^2= 27%	R^2= 15%	R^2= 20%	R^2= 32%	R^2= 16%	R^2= 26%
Stage of studies (initial stage/final stage)	***	ns	ns	***	ns	**
Field of study	ns	***	ns	**	ns	ns
Gender (male/female)	ns	***	ns	*	**	ns
Mother tongue grade in the matriculation examination	***	**	***	***	***	***
Student has a qualification/degree other than the matriculation examination (vocational qualification/higher education degree)	**	ns	**	ns	ns	ns
Number of books in the childhood home	ns	***	***	**	*	**
Parental education	*	ns	ns	ns	ns	ns
Student effort in the CLA+ test	***	***	***	***	**	***
Student engagement with the CLA+ test	***	ns	ns	*	*	ns

*** p<0.001; ** p<0.01; * p<0.05; ns = not statistically significant

When we examine university students' scores in more detail (Appendix 7, Table 3), we observe that the difference between initial stage and final stage students was only statistically significant in the performance task (33 points in favour of final stage students). Differences between fields of study were only found in the SRQ scores. The best scores were achieved by students of Agriculture, Science and Services. It should be noted that when determining differences between the fields of study, the effects of the student's gender and mother tongue grade in the matriculation examination had been controlled in the calculation. Consequently, the differences between fields of study initially observed in the performance task, for example (Figures 5 & 6) can be largely explained by the fact that students in various fields of study differed from each other in terms of their background variables, including their performance in the matriculation examination. No actual impact of the field of study on generic skills was observed in the performance task and the total score. Gender only had a statistically significant link to the scores obtained for the selected-response section: on average, men received 50 points more than women for this section.

Success in the mother tongue test in the matriculation examination was the strongest individual predictor of university students' generic skills, and it also had a strong and statistically significant correlation with the scores after the background variables were controlled. In particular, if the grade awarded to the student in the mother tongue test had been E or L, they could be expected to score highly in the CLA+ test (in the performance task, the difference between them and the group of students who had not taken the mother tongue test in the matriculation examination was as much as 100 points). Students who received the grade M also stood out positively from the reference group (66 points) in the performance task.

If a university student had already completed another higher education degree, his/her expected score in the performance task was more than 40 points higher than the scores of students with no higher education degree. However, the difference in the total score was only 21 points, which was not quite statistically significant at the 5% significance level. On the other hand, the CLA+ mean scores of the students who had a vocational upper secondary qualification were statistically significantly lower than the other students' scores.

Of the variables describing the student's socioeconomic background, the number of books in the childhood home explained variation in the scores achieved in the selected-response section and the total score to a statistically significant degree, while parental education explained variation in the performance task. The group who had had the greatest number of books in their childhood homes (more than 500 books), stood out particularly clearly from the other groups. Regarding parental education, the main division was based on whether the students' parents had completed at least a secondary level

qualification or not. The average points scored in the performance task by students whose parents had no more than a primary level qualification were statistically significantly lower than those of other groups.

Among university students, attitudes towards the CLA+ test explained variation in all subarea scores statistically significantly. What is essential in this respect was that the student had put in at least a moderate amount of effort when taking the test (Table 3, Appendix 7). Regardless of the subarea, their predicted score was in this case more than 100 points higher than the score of a student who made little or no effort.

In universities of applied sciences, the average difference between the scores of initial stage and final stage students (Table 4, Appendix 7) was slightly greater than among university students. This was mainly due to the better performance of final stage students, not only in the performance task (difference of 32 points with background variables controlled) but also in the selected-response questions (difference of 22 points with background variables controlled) than initial stage students, even if the latter difference was not statistically significant. However, the difference in the total scores between initial and final stage UAS students was statistically significant, unlike for university students. Statistically significant differences between fields of study only emerged in the performance task, in which students in the fields of Health or Welfare and Business and Law scored the highest point, with other background variables controlled. Women did better in the performance task than men (difference of 26 points in favour of women), while the result was the opposite in the selected-response questions (difference of 36 points in favour of men). These differences were also calculated with background variables controlled.

The association between mother tongue grade in the matriculation examination and the CLA+ scores was even stronger for UAS than university students. The mean scores went up systematically as the grade improved. In all subareas, the predicted scores of students who had been awarded the grade E or L were more than 100 points higher than those of students who had not taken the mother tongue test at all. The CLA+ scores also correlated strongly with the number of books in the childhood home and student effort in the CLA+ test, especially in the performance task. Experiencing the CLA+ tasks as engaging was further associated with better than average scores for UAS students, both in the performance task and the selected-response questions.

4.3 Development of generic skills during higher education studies

A cross-sectional design naturally sets its limitations on investigating the development of students' generic skills (section 3.8). In order to examine the development of skills, a follow-up study would have been needed, in which the same students would have been studied both at the initial and final stage. Comparing the differences in the mastery levels of generic skills between initial and final stage students belonging to different cohorts in the Kappas! project can, however, give indications of what happens during higher education from the perspective of generic skills. Here, we describe this development by comparing the differences between initial and final stage students' skills in each task type by examining the magnitude of the differences in the light of effect size, identifying transitions of mastery levels, and comparing the differences between initial and final stage students to the reference data.

As noted in section 4.1.1, there were differences in the mastery of generic skills between initial stage and final stage students, of which the ones related to the total score and performance task score were statistically significant, with final stage students showing a higher level of generic skills than initial stage students. In the selected-response questions, differences between the two student groups were generally smaller than in the performance task. For example, the difference between final stage and initial stage university students in the selected-response questions was as little as one point, which was not statistically significant. The corresponding difference (33 points) observed between initial and final stage UAS students was also not statistically significant (Figure 3). Consequently, the data appear to give some indications of students' skills measured by the performance task (analysis and problem solving, writing effectiveness and writing mechanics) having improved more during higher education studies than the skills measured by the selected-response questions (scientific and quantitative reasoning, critical reading and evaluation, critiquing an argument).

The average difference between the scores of initial stage and final stage students can be measured by effect size. When comparing two groups, the effect size can be measured with Cohen's d statistic, which in this case is the difference between the mean scores of initial stage and final stage students divided by the standard deviation of the initial stage students. The usual interpretation of this indicator is that an effect size exceeding 0.80 is considered large, while an effect size falling between 0.50 and 0.80 is considered moderate. The effect size is small if it is below 0.50, and very small if it is below 0.20. The effect sizes of the performance task score, selected-response section score and the total score calculated for the Kappas! data set were 0.36, -0.01 and 0.20 for university students. For UAS students, they were 0.34, 0.20 and 0.33 respectively. As a comparison, we can note that the effect sizes in the US reference data were 0.31, 0.29 and 0.36 respectively.

The effect sizes indicate that the differences in the generic skills measured by the CLA+ test between initial stage and final stage students were small. The effect sizes were slightly larger in the performance task than in the selected-response section. Examined in this way, Finnish students' generic skills would appear to improve little, and this improvement is more or less on par with the corresponding indicator in the US data.

Cohen's d values for the data set were also calculated by controlling the intervening effects of the background variables examined earlier (field of study, age, gender, educational background, socioeconomic background and attitude to the CLA+ test, see section 4.2). In other words, the possibility that the samples of initial stage and final stage students differed from each other in relation to these background variables was taken into account when determining the effect size. However, controlling the background variables did not essentially change the effect size values: the controlled d values of university students were 0.37 in the performance task, -0.06 in the selected response questions and 0.15 in the total score. These figures for UAS students were 0.30, 0.21 and 0.32. This observation suggests that the considered background variables do not explain the (minor) differences in the skills of initial stage and final stage students. Instead, there are other factors behind these differences. For example, final stage students have completed higher education studies, while their initial stage counterparts have had little time to do so.

When the generic skills of initial stage and final stage students are examined by mastery level distributions (Table 6), we observe that UAS students transition from below basic mastery levels towards the basic and proficient levels. University students similarly transition from a proficient level towards an accomplished level of mastery. This observation, too, contributes to confirming that something happens to improve generic skills during higher education studies.

To obtain a slightly better idea of the improvement in generic skills, we can compare Finnish students' scores to the US reference data. We then notice that the average total scores (1,116 and 1,138) of initial stage and final stage university students are statistically significantly higher than the corresponding reference values. The scores of UAS students (1,027 for initial stage and 1,063 for final stage students), on the other hand, are significantly lower than the reference values. However, the average differences between initial stage and final stage students in Finland are smaller across the board than the corresponding differences in the reference data. For university students, the difference was 21 points in the total score and 45 points in the performance task score, whereas it was only one point in the selected-response questions, which exceptionally was in favour of initial stage students. For UAS students, these differences were 36, 40 and 32 points respectively. The reference values obtained from the US data were 53 to 54 points. In the light of the CLA+ test, there is less improvement in Finnish students' generic skills than in the US data which is, above all, explained by the lower than expected scores of final stage

students. There was no statistically significant difference between the scores of initial stage students in Finland and the reference data.

This can be summed up by saying that the differences in mastery of generic skills between initial stage and final stage higher education students were associated particularly with the skills measured by the performance task: analysis and problem solving, writing effectiveness and writing mechanics. Consequently, it would appear that these are the specific skills that improve during higher education studies. On the other hand, the differences between initial stage and final stage students are not large, which in turn indicates that generic skills improve little during the studies.

5 Higher education institutions' and students' experiences of the Kappas! project

Kaisa Silvennoinen

- According to higher education institutions, particularly well-functioning aspects of the project implementation included project coordination, support offered for arranging the tests, the different materials provided and the support group's work.
- Many higher education institutions planned to use the findings to support teaching development, and it was also felt that the project made generic skills more visible in higher education institutions.
- Aspects that higher education institutions experienced as especially challenging in the project implementation were associated with student recruitment, implementing the test in practice and the limitations of the institution-specific findings.
- The interviewed students' experiences of the test and test situation were mainly positive.
- Experiences of the test's level of difficulty varied slightly, however, and students had experienced time pressure when taking the test.
- To a large extent, the interviewed students intended to use their test results for self-development, while some may also use them when seeking for employment.
- Taking the test and receiving feedback increased the interviewed students' awareness of their generic skills.

5.1 Higher education institutions' experiences

Higher education institutions' experiences of the Kappas! project were mapped by means of an electronic feedback survey (Webropol survey) between September and December 2020. A link to the survey was sent to the project's support group members in all participating higher education institutions. In total, six universities of applied sciences and nine universities responded. While higher education institutions were allowed to coordinate their responses as they preferred, each higher education institution was asked to only provide a single response. The survey was available both in Finnish and Swedish.

It contained seven open-ended questions (Appendix 9) about such aspects as the implementation of the Kappas! project, use of the findings in teaching development, and potential proposals for improvements from the institutions' perspective. At the end of the survey, the institutions were asked to indicate the higher education sector in which they operate. Responding to the survey was voluntary for the institutions. The notes attached to the survey provided information on the purpose of the survey and its voluntary and confidential nature. The project's support group was also offered the opportunity to comment on the questions before the survey was sent out.

In this section, we take a closer look at higher education institutions' experiences of the project. The examination follows the order of the survey questions.

What motivated higher education institutions to participate in the project?

The first question in the feedback survey was about what motivated higher education institutions to participate in the Kappas! project. Higher education institutions to a great extent emphasised factors related to the findings and their utilisation in their responses, including the possibility of obtaining information on the levels and development of students' generic skills. In general, higher education institutions wished to receive feedback on their students' generic skills, but gaining knowledge of the mastery levels in different fields of study and among students who were at different stages of their studies were also mentioned separately. The institutions were also interested in finding out how the students' skills develop during their higher education studies. In connection with these themes, the responses also emphasised using the research evidence and institution-specific findings produced by the project in developing the institution's activities, including teaching and guidance.

The importance of generic skills was also brought up in several feedback responses in more general terms. The responses showed that higher education institutions had recognised the need for generic skills. In addition, they had taken or intended to take generic skills into account better, for example in various policies on the education and teaching delivered by higher education institutions. On the other hand, one of the responses pointed out that generic skills may sometimes also be overshadowed by the development of competence specific to the field of study, even if their importance is recognised better than before in higher education institutions.

A few responses mentioned the approach of the study and the test instrument. For example, they noted that this was a new approach to assessing the competence produced by Finnish higher education. A respondent also pointed out that measuring the level and development of students' generic skills is difficult without a study of this type. Higher education institutions felt that this information was important for them, however. One response also mentioned interest in the CLA+ test. The possibility of using the test in entrance examinations was a particular object of interest.

Based on a few responses, participation in the project was partly motivated by the strong role and political will of the institution's top management or the Ministry of Education and Culture. One higher education institution mentioned that it also wished to participate in the project to increase the coverage of national evaluation and improve the reliability of its results.

Higher education institutions' experiences of project implementation

The institutions were asked in the survey about what worked well and, on the other hand, what was challenging in the Kappas! project implementation from the perspective of higher education institutions. The institutions provided diverse feedback on such aspects as project coordination, test implementation, test feedback for higher education institutions and students, and recruiting students.

Many of the factors the institutions found effective were relevant to project coordination, in particular. The responses emphasised cooperation and communication between project coordinators and higher education institutions as an example of the aspects that worked well. Project communication was experienced as effective. The instructions and training events related to the project as well as the support and communication materials sent to the higher education institutions were regarded as useful. The project's support group was also mentioned in many responses. Its operation was experienced as smooth and useful, for example from the perspective of access to information or planning the testing at the respondent's institution.

Many of the aspects experienced as effective in project implementation were related to the practical arrangements and implementation of the tests. For example, respondents were happy with the support provided by the project for organising the test. The internal resources of higher education institutions, including well-functioning IT support or support provided by the institution's management, as well as the organisation of the tests in cooperation with different actors also came up in several responses. Many higher education institutions additionally found that the electronic test system was mainly reliable and easy to use. They also said that the translations of the test and the test environment worked well.

On the other hand, the responses also brought up challenges related to implementing the tests. For example, respondents noted that the organisation and coordination of the tests took up a great deal of the staff's working time and resources. The fact that several project phases related to the organisation and implementation of the tests were sometimes underway at the same time was also experienced as a challenge. Respondents mentioned that the coronavirus situation hampered the testing in spring 2020. Additionally, some higher education institutions felt that the test system offered poor usability. A respondent said that the test itself proved challenging for some students, and some did not finish it.

Some respondents also experienced challenges associated with the test feedback for higher education institutions and students and the sampling of the study. Firstly, some felt that the usability of the institution-specific findings was poor due to such factors as the small number of participants. The usability of the badge intended for students was also questioned. Respondents additionally noted that students had received their test feedback late, some more than six months after taking the test. Secondly, respondents found that the sample was too fragmented, for example from the perspective of an individual higher education institution. There were also ambiguities in the guidelines for preparing the sampling plan for the individual institution.

Institutions' responses indicate that recruiting students for the tests was found a particularly challenging and labour-intensive part of project implementation. This theme came up in one way or another in almost all responses. Motivating students and persuading them to participate in the test on a voluntary basis was found difficult. Many higher education institutions found it particularly challenging to reach final stage students and to motivate them to participate in the tests. It was also mentioned that motivating teachers to participate in organising the tests was challenging. However, integrating the tests with courses was seen as a way of ensuring a higher student participation rate in the test, and teachers were regarded as playing an important role in this.

A respondent pointed out that piloting the project would have been useful from the perspective of identifying potential challenges. According to several responses, on the other hand, institutions found that the project as a whole was well planned and implemented.

Using the results in the development of teaching

One of the survey questions was whether the institutions had been able to use the findings of the assessment project, or intended to use them, to develop their teaching. They were also asked to explain in greater detail how the results will be used or, potentially, why they cannot be used. Based on the responses, most higher education institutions intended to use the findings in some way in their work. For example, the responses mentioned many activities related to the planning and development of teaching, such as curriculum work or pedagogical development, in which institutions intended to use the results. In several higher education institutions, the institution-specific results were also discussed by the management teams for teaching and education. In addition to teaching development, the responses brought up the possibility of using the findings in personnel development (including teachers' pedagogical training) and potentially also in student guidance, including personal study plan discussions and career guidance. A few responses also referred to the organisation of the institution-specific Kappas! webinar for the staff (section 3.7.2). Among other things, they hoped that the webinar would provide more ideas for using the results.

The responses also highlighted the fact that the Kappas! project has in general made generic skills visible in higher education institutions and sparked discussion on their role in teaching. One respondent also mentioned the personal feedback provided to students and the fact that the project increased understanding of how generic skills are important for learning.

Some higher education institutions also mentioned various factors that make using the results challenging. The challenges were mainly related to the limitations of the institution-specific results. For example, the fact that students from all fields of study in an individual institution were not tested was experienced as something that impaired the usability of the institution-specific results. Secondly, students' participation rate in some higher education institutions was rather low, which is why the possibilities of using the results directly were not considered particularly good. A respondent noted that the final report of the project would probably be more useful for their higher education institution and the entire higher education field. The fact that the findings were provided at higher education institution level was also experienced as a limitation, and drilling down to the degree programme level, for example, would require a great deal of effort of the institution. To improve the usability of the findings, comparison data based on the respondent's higher education sector was needed, rather than on the entire data set.

Respondents also mentioned the general challenges related to the organisation and processes of HEI development work, which make the project's results difficult to use. A respondent additionally pointed out that the HEI does not intend to use the results further in teaching development, as other well-functioning processes already exist for this purpose.

Development proposals concerning the project

In addition to questions about project implementation and making use of the findings, the respondents were also asked about how the project could be improved from the perspective of higher education institutions. The development proposals made by institutions concerned project implementation, practical-level test arrangements, and the test feedback for higher education institutions and students in various ways. The proposals largely mirrored the previously mentioned issues that were considered challenging from the institutions' perspective.

In particular, integrating the tests into teaching was highlighted in proposals for developing project implementation. Based on the responses, recruiting students had been successful when, for example, the test was integrated into a course and the teacher was involved in organising it. Many of the responses mentioned that the implementation of the assessment should be taken into account in such contexts as higher education

institutions' curriculum work in the future, and it should thus be more closely integrated into teaching. This would encourage students' participation in the tests and enable higher education institutions to make better use of the findings.

Individual proposals for improving project implementation were also received that concerned sampling and the planning and scheduling of the project as a whole. For example, respondents hoped that sampling could lend better support for the motivation of both students and teachers. Secondly, the duration of the project was considered rather long, and respondents felt it should be made less cumbersome as a whole. They would also have preferred a schedule in which the different stages of testing would clearly take place at different times.

The responses included several suggested improvements for the practical implementation of the tests. Many higher education institutions found the current testing method rather labour-intensive for the organising and supervising staff. To facilitate the practical arrangements, respondents suggested that students could take the test remotely, using their personal devices and at a time that suited them. Respondents would also like to see improvements in the usability and functionality of the actual test system. In addition, higher education institutions had at times found organising the tests somewhat challenging. On the other hand, a respondent also pointed out that repeating the tests would not be a problem, now that the institution has already organised them once.

A few HEIs gave proposals for developing the test feedback. Respondents would prefer slightly simpler institution-specific reports, and the report should be more detailed, for example at the degree programme level. Better usability of the badge intended for students was also called for, and it was suggested that it should provide more detail about the student's generic skills. Ensuring that students receive feedback on the test sooner was also mentioned as an area of development.

Overall, it appeared that higher education institutions would be interested in and have a need for assessing the level and development of students' generic skills. Many higher education institutions intended to use the assessment project's findings to support the development of teaching, and their experience was that the project had also given more general visibility for generic skills. Most of the responding higher education institutions also stated that they would possibly or probably participate in the Kappas! project again. From the perspective of planning and implementing the assessment, however, many important proposals for improvements were received from the institutions, especially in relation to test implementation, student recruitment and the feedback report for higher education institutions.

5.2 Students' experiences

Student interviews were conducted to find out about students' experiences of the Kappas! project. A total of four university students from several different fields of study who participated in the project were interviewed. Three of the interviewees had taken the test in their first year of study, while one was in their third year.

E-mails were sent out to invite students to the interviews. The project team sent an invitation to the interview to a total of 67 UAS and university students who had taken the test. In addition, the members of the project's support group were asked to forward the invitation to students who had taken the test at their higher education institutions. Reaching students to be interviewed was challenging, although the original aim had also been to only interview a small number of participating students for the final report. The students who were interested in being interviewed contacted the project team themselves to arrange a time for a personal interview.

All interviews were conducted on a one-off basis via a remote connection (Zoom) in September and October 2020. The duration of the interviews was 15 to 30 minutes. By the interviewees' consent, the interviews were recorded to enable the further processing of the responses. All four interviewees were informed both in writing and orally about the purpose of the interview and its voluntary and confidential nature. The interviewed students also had an opportunity to read the section on student experiences before the report was published. No comments on this section were received from them by the deadline.

The interview questions (Appendix 10) focused on three themes on which the researchers wished to have student feedback: the test and test situation; test results and feedback; and generic skills in general. While the same set of questions was used in each interview, their order varied slightly based on each interview situation and the interviewees' responses. If necessary, more detailed questions related to their responses were also put to the interviewees.

The interviewed students' experiences of the Kappas! project are discussed in detail in the following sections. The sections follow the order of the interview themes.

Students' experiences of the test and test situation

At the beginning of the interview, the students were asked when they had taken the test and why they had decided to participate in it. They were also asked how they were invited to take the test. Three of the interviewees had taken the test in autumn 2019 and one in spring 2020. As the most important factor that had motivated the students to take

part in the test the interviewees emphasised obtaining information about their skills and competence. Their personal scores and, in particular, the level of their generic skills compared to other higher education students interested several of the interviewees. A student also said that they had wished to help with the research in general and contribute to it. Almost all of the interviewees had received an invitation to the test by e-mail through their higher education institutions. Some additionally recalled that they had possibly been informed of the test at the beginning of a lecture or on some other information channel of the institution rather than by e-mail.

The students were asked how they experienced the test situation and the test. The interviews indicate that the students' experiences of the test situation, and especially their practical arrangements and implementation, were positive. They described the test situations as having been pleasant, calm and well prepared. In this context, some of the interviewees referred to good and clear test instructions. However, one interviewee also described feeling nervous about the external setting of the test situation (including a venue which they had not previously visited).

Some of the interviewees were also nervous about the test itself and especially their performance in it. One of the students said that they did not really know what to expect in the test. They mainly found taking the test rather pleasant, however. For example, some of the interviewees described the contents of the test as comprehensible and meaningful. One of them mentioned that the test was an opportunity to demonstrate your skills, also to yourself.

However, the interviewees had somewhat varied experiences of the test's level of difficulty. One of the students did not find the test too difficult, while another had slightly mixed feelings about whether the test was easy or difficult. One student said that the test was interesting but that the selective-response section, in particular, was challenging. Almost all of the interviewees had also experienced time pressure in the test. Some mentioned that they had run out of time in the selective-response tasks, while others felt under time pressure in the performance task, in which they had to use a number of sources to compile the response. On the other hand, one of the students reflected that there probably was a reason for putting the test together in a certain way and limiting the time available for it.

Students' experiences of test results and feedback

The aims of the interviews also included exploring the students' experiences of the CLA+ test results and feedback. The students were asked how they experienced the feedback provided to them, whether or not the feedback was sufficient, and how the feedback should be improved to help them build up their generic skills further.

Based on the interviews, the students were mainly happy with the feedback they received on the test. In particular, different factors related to the content and structure of the feedback report were emphasised in the interviews. The students found that the feedback was clear and nicely structured into different sections. Several of the interviewees also welcomed the fact that the report explained verbally how the scores should be interpreted and what they mean (for example, a more detailed description of the mastery level contents). Some of the students also mentioned that the report provided them with information about their results in comparison to other students' scores. On the other hand, they would have liked more detailed feedback on the different subareas of the test. One of the students was unsure about the type of skills the individual subareas targeted and, to support the development of their skills, would have liked more detailed information about what the weaknesses and strengths in each area could consist of, for example. Providing more concrete feedback on individual test subareas emerged as the only actual development proposal concerning the feedback report in the interviews.

In the context of the test feedback, one of the interviewees pointed out that higher education studies differ from secondary level studies, and they felt that their results gave them self-confidence and the feeling that they could cope with higher education studies. Another student said that they had already received feedback similar to that provided by the report on their level of mastery through good performance in their studies. In other words, it appeared that the test feedback was to some extent also associated with the interviewees' ideas of their academic success and themselves as (higher education) students.

The students were asked to consider in the interview if the feedback came too late after the time at which they had taken the test. Almost every interviewee noted that the delay before they received the feedback was long. Some of the interviewees said that the test was still fairly fresh in their memory when they received the feedback, but especially some of the students who had taken the test in the autumn semester said they had forgotten all about it by the time they received the results. A student said that the feedback came too late also in the sense that they had already started losing interest in the matter. An interviewee also recalled having read somewhere that the test feedback would come earlier than it actually did but, on the other hand, said they had been prepared for the eventuality that the results could take a long time.

The interviewees were also asked about their experiences of the support material sent to them with the feedback report (attachment with advice for developing generic skills) and the possible badge. Two of the interviewees said that they had read the attachment. They described it as having been useful and illustrating the content of the report even better. On the other hand, some of the interviewed students had not read the attachment carefully; for example, they had only glanced at their results. They mainly felt that the

badge was a fun addition to the results, and some of the interviewees considered possibly using it when applying for jobs. On the other hand, they did not experience it as particularly useful, however. A student said that the badge may not be very well known in Finland, and they would not add the image directly to such documents as their CV as they felt that the badge needed to be complemented with some explanation.

The students were also asked what they intend to do with the test feedback they received. All interviewees intended to use the test results for self-development in one way or another. The interviewees mentioned that the feedback enabled them to engage in self-reflection and pay attention to their development areas in generic skills. They also felt that their (good) test results improved their self-esteem and served as a necessary reminder of their skills to them. Additionally, they said that the feedback supported their self-image as learners and employees. Some of the interviewees mentioned that they may use, or have already used, the feedback they received when applying for jobs. The test feedback was mainly considered useful for the students themselves, however. One observation made in the interviews was that students do not necessarily know how such feedback could be utilised.

In addition to questions about personal test feedback, the interviewees were asked about how feedback on an assessments of this type could be utilised in higher education institutions from the student's perspective. The interviewees highlighted the viewpoint of using the feedback to support learning and studying. Gaining knowledge of their generic skills, also in relation to other students, was considered useful for students. They felt that this made it possible to study more efficiently. For example, one of the interviewees said that if everyone took a test of this type, you could perhaps determine the level of your skills in relation to others, and thus consider the areas on which you should spend more time in your studies. An interviewee also pointed out that at a personal level, they would otherwise not have paid any attention to this issue, and that they had not really even heard about generic skills before or thought that these skills could be measured and developed.

The students also considered the assessment of generic skills in higher education in more general terms. One of the interviewees wondered if these assessment could be used in entrance examinations, for example. Including generic skills and their assessment in the studies was also brought up in the interviews. One of the students reflected on the relevance of different skills, such as critiquing arguments and reading skills, to degree programmes in general. They had found the texts used in the test interesting specifically in relation to critiquing arguments and wondered if more of such activities could be integrated into the studies. An interviewee additionally said that the possibility of participating in the test was a good thing and hoped that testing of this type would become more widespread and mandatory for all students.

Importance of learning generic skills

The students were also asked in more general terms how important they consider generic skills and how they felt that higher education studies have supported the learning and development of their generic skills. The interviewees found generic skills important in general and considered them necessary for supporting both their studies and participation in working life. The importance of generic skills in today's world also emerged in the interviews in a way. For example, skills in evaluating information, media literacy and the ability to recognise and critique different arguments were considered highly important and trending skills.

The interviewed students' views of how higher education studies have supported the learning of generic skills varied. Some of the interviewees felt that their studies had encouraged the development of skills. One student also reported having noticed that the test showed an improvement in their skills, even if they had never thought to pay attention to them previously. Some of the interviewees, on the other hand, felt that their studies had provided rather insufficient support for learning generic skills. For example, one student said that their studies had involved learning a high number of individual pieces of information and called for stronger links between learning generic skills and the practice.

The interviewees were also asked if their perception of their generic skills changed after completing the test and receiving the feedback. Some students said that rather than being changed, their ideas of their generic skills were reinforced. By this they meant that they already experienced the skills measured by the test (such as writing) as personal strengths, and the test results reinforced this perception of their skills further. On the other hand, some of the interviewees emphasised the fact that the study increased their awareness and understanding of different generic skills.

6 Conclusions – read, write and think!

Jani Ursin, Heidi Hyytinen, Auli Toom & Katri Kleemola

- Learning generic skills plays a key role already at lower levels of education and in learning environments outside the school.
- The possibility of taking generic skills into consideration in student admissions should be examined.
- Generic skills should be developed in line with the objectives of UAS and university education.
- The learning of generic skills should be supported in a goal-oriented manner in higher education teaching.
- A longitudinal study would be needed to obtain information about the development of the students' generic skills during higher education studies.

6.1 Key findings

The purpose of this study was to find out about higher education students' mastery of generic skills, the factors associated with the level of mastery, and the extent to which generic skills develop during higher education studies. The target group of the study were initial and final stage university and UAS students in Bachelor's degree programmes. The performance task of the study measured analysis and problem solving skills, writing effectiveness and writing mechanics, while the selected-response section focused on scientific and quantitative reasoning, critical reading and evaluation, and critiquing an argument. The findings of this study were also compared to reference data collected by the same method in the United States.

Level of generic skills. Almost 60% of the participating students reached at most the basic level of generic skills. On the other hand, approx. 40% reached at least the proficient level, even if students who achieved the highest mastery level (advanced skills) were few and far between. A more detailed examination reveals differences in the level of mastery between UAS and university students. University students had better generic skills across the board than UAS students. This was explained by university and UAS students' dissimilar

family and educational backgrounds. For example, the students who had completed the matriculation examination had a clearly better mastery level of generic skills than those who had not. The majority of university students but only two out of three UAS students had sat the matriculation examination. University students' parents also had a higher level of education than UAS students' parents, which partly explained the differences in skills.

When looking at the mastery of generic skills by skill category, we notice that higher education students reached a high level in writing mechanics: 88% of the students were placed at the top three levels of mastery. This means that the students had a relatively good command of the established practices of written language. They also had good levels of mastery in writing effectiveness (81% at the top three levels) as well as in analysis and problem solving (79% at the top three levels). Finnish students had clearly higher levels of mastery in these three subareas than the students in the US reference data. When we examine the generic skills measured by the selected-response section, or scientific and quantitative reasoning, critical reading and evaluation as well as critiquing an argument, Finnish students' skills are on par with the US reference data. In addition, the differences in these skills between university and UAS students were smaller than in the skills measured by the performance task. For example, final stage university students, who typically had the best mastery of generic skills of all respondent groups, only had better skills than final stage UAS students in critical reading and evaluation, whereas their levels of mastery in scientific and quantitative reasoning and critiquing an argument were the same.

Link between generic skills and background factors. When we look at the link between generic skills and field of study, age, gender, educational background, socioeconomic background and attitude towards the test, we observe that age and field of study do not systematically correlate to the mastery of generic skills. The findings indicate that women appear to have a better level of mastery in skills measured by the performance task, or analysis and problem solving, writing effectiveness and writing mechanics. On the other hand, men had a better level of mastery in the skills measured by the selected-response section, or scientific and quantitative reasoning, critical reading and evaluation and critiquing an argument. This gender difference was also apparent in an examination by field of study: in female-dominated fields, the students showed better mastery of analysis and problem solving, writing effectiveness and writing mechanics than students in male-dominated fields. In male-dominated fields, on the other hand, the students had a better level of mastery in scientific and quantitative reasoning, critical reading and evaluation and critiquing an argument. However, the key factors explaining the mastery of generic skills were related to the student's educational and socioeconomic background and attitude to the test. In particular, the student's performance in the mother tongue test in the matriculation examination strongly explained mastery of the generic skills assessed in this study. Another strong explanator was the number of books in the student's childhood home, which reflects a culture of reading and learning. The more books there were in

the student's childhood home, the better generic skills they had. The third important explanator was associated with the test situation: the amount of effort the student put in when taking the test was strongly linked to their success in it.

Development of generic skills. While a cross-sectional design naturally does not enable a detailed examination of how generic skills develop, it does provide indications of changes in these skills during the studies. When we compare initial stage and final stage students' levels of generic skills, we observe that the skills measured by the performance task have improved more than those measured by the selected-response section during the studies. Consequently, it would appear that especially analysis and problem solving skills, writing effectiveness and writing mechanics improve during the studies. On the other hand, the findings indicate that the development of generic skills is negligible when examined as differences between the skills of initial stage and final stage students. The same observation is made when comparing this difference in skills with the reference data from the United States: the differences in skills between initial stage and final stage Finnish students are smaller than those recorded in the United States. This is another indication of Finnish students' generic skills developing only to a modest degree during their studies (cf. Badcock et al. 2010; Arum & Roksa 2011; Evens et al. 2013).

6.2 Development proposals

The following development proposals may be given based on the project's findings:

1. *To lay a foundation for continuous learning, attention should be paid to generic skills already at lower levels of education and in learning environments outside the school.*

While the subjects of this project were university students, the findings show that the foundation of generic skills is already laid at the previous levels of education, and especially at the secondary level. The general upper secondary school as a form of general education appears to produce the strongest generic skills – at least regarding those skills that were studied in this project. Higher education students who had completed a vocational upper secondary qualification had lower mastery levels of generic skills than those who had completed the general upper secondary school. Consequently, the findings of this project indicate a need to consider the extent to which generic skills should also be emphasised more at the secondary level, especially in vocational education and training.

The findings of this study also partly emphasise the importance of learning environments outside the school for the mastery of generic skills. The key factors explaining higher education students' mastery of generic skills were the mother tongue grade in the

matriculation examination and the number of books in the student's childhood home. Both of these factors demonstrate the importance of a growth environment in which a culture of learning and reading is emphasised.

The findings of this project are also consistent with numerous previous studies (e.g. Leino et al. 2017) which emphasise the importance of reading and encouraging it from a very young age. To put generic skills within the reach of every citizen, their development should be supported from very early on and in a variety of learning environments. This also lays a foundation for the continuous learning skills needed by citizens throughout their lives.

2. *Exploring the possibility of taking generic skills into consideration in student admissions*

Student admissions have been reformed in recent years, and the focus has shifted from entrance examinations measuring field-specific knowledge to selections based on grades in the matriculation examination certificate. The emphasis that should be placed on different subjects has sparked a lively discussion in the public forum and within higher education institutions. The Kappas! project's findings lend support for selections based on certificates. There is a strong association between initial stage students' mother tongue and mathematics grades in the matriculation examination on the one hand and their generic skills on the other; the latter in turn facilitate the learning of field-specific knowledge and skills (cf. Arum & Roksa 2011; Halpern 2014). In the future, the possibility of emphasising these subjects of general nature even more in student admissions should be explored. Whereas selection based on certificates appears to be an effective method in student admissions in the light of the observations of this project it is, of course, not possible to give up entrance examinations completely. A pathway to higher education must be secured for those eligible applicants who have not completed the matriculation examination. The reform of student admissions has brought in more generic entrance examinations: at universities of applied sciences, the entrance examinations already strongly emphasise generic skills, and broader entrance examinations have also been discussed in the context of universities (e.g. Talman, Borodavkin, Kanerva & Haavisto 2018). Prior to a thorough renewal of entrance examinations, a discussion across the entire higher education sector is needed on which generic skills are the most important. It is also crucial to consider what the aims of the entrance examination are in each field of study. Longitudinal research on the development of generic skills in Finnish higher education will additionally be needed to underpin the reforms. It is also important to take into account the challenges of assessing skills: while selected-response questions make higher education institutions' work easier, they are not suitable for assessing all skills, and developing performance tasks with open-ended responses and assessing the responses require a great deal of resources and special expertise (Chapter 2; Ercikan & Pellegrino 2017; Hyytinen & Toom 2019). It is also essential to remember that the students' initial level of mastery tells us nothing about their potential to improve their generic skills during higher education studies.

3. *Generic skills should be developed in line with the objectives of UAS and university education.*

The findings of the study highlighted the differences between university and UAS students' mastery of generic skills, with university students displaying better skills than UAS students. This observation naturally tells us nothing about the quality of higher education (cf. Ashwin 2020); primarily, it indicates that each higher education sector attracts different students. This naturally is as it should be, as universities and universities of applied sciences also have different statutory missions: universities of applied sciences emphasise vocational competence (Universities of Applied Sciences Act 2014) whereas universities focus on research (Universities Act 2009). It should also be emphasised that the generic skills measured in this study were skills that general upper secondary education appeared to improve more than vocational upper secondary education. Consequently, if the project had measured (generic) skills in which professional capabilities are more prominent, the results could also have been different. From the perspective of improving and developing generic skills it is, in other words, important to acknowledge the fact that students in each higher education sector are different to begin with, and to consider the consequences of this for pedagogical solutions. Striving to develop the generic skills that are needed in each field of study and in the tasks of the field is at least equally important. However, this naturally does not mean that the type of generic skills that this project examined (critical thinking and written communication skills), should be overlooked. On the contrary, they are important skills throughout life, regardless of the field of study (e.g. Shavelson 2010; Arum & Roksa 2011; Tuononen et al. 2017; Virtanen & Tynjälä 2018; Hyytinen et al. 2019), and they also help students reflect on their personal relationship with who they are and what they can do in this world – an issue that many critical education researchers highlight as the most important goal of higher education (Ashwin 2020). Ultimately, this is a question of how the teaching and learning of different generic skills can be integrated into the intended learning outcomes and assessment practices of each field of study, whether at a university of applied sciences or a university.

4. *The learning of generic skills should be supported in a goal-oriented manner in higher education teaching.*

The findings of the Kappas! project indicate that, as assessed by the CLA+ test, Finnish higher education students' mastery levels of generic skills are surprisingly low. However, generic skills are not permanent characteristics of an individual. They can be taught and learned as part of higher education studies (Halpern 1998, 2014; Heijltjes et al. 2014; Arum & Roksa 2011). The findings of the Kappas! project also indicate that students can improve their generic skills during their higher education studies.

Mastering generic skills is one of the objectives of higher education studies, in addition to building expertise in the student's special field and learning key knowledge and skills of the field of study. It is natural to perceive teaching field-specific knowledge and skills as the mainstay of higher education; on the other hand, the teaching of generic skills at different stages of studies and during courses throughout a degree programme may be experienced as challenging (Hyytinen et al. 2019; Tuononen, Hyytinen, Hailikari & Toom 2020). From the perspective of improving skills, it would be beneficial to pay attention to learning generic skills in individual courses, more extensively in the curricula, and at the level of teacher–student interactions alike. This requires communality, cooperation and interaction between teachers, those making decisions on the curriculum and students in higher education institutions (Tuononen et al. 2020). Pedagogical leadership is also needed, which can ensure that the objectives are implemented in practice in curriculum work and the daily teaching work. Attention should also be paid to the attitudes and perceptions of teachers or the higher education community (see Barrie 2007; Jääskelä, Nykänen & Tynjälä 2018).

Addressing generic skills at the course level. In terms of learning generic skills, it is essential that teachers have adequate pedagogical competence to support the students' learning and reflection. The teachers' competence is what ultimately determines what happens in the teaching situation, which teaching methods the teacher uses, what kind of feedback they give and how they assess the students' learning. If field-specific courses are simultaneously expected to teach generic skills, higher education teachers may find this difficult (see Tuononen et al. 2020).

A precondition for combining generic skills with courses specific to the field of study often is applying student-centered approach to teaching as well as mastering and using many different teaching and assessment methods during a course (see Virtanen & Tynjälä 2018). If the aim is to learn cooperation skills, for example, it must be ensured when planning and implementing the course that the students work together in a versatile manner during it and complete at least some of the learning assignments together. The learning of cooperation skills should also be taken into account in the assessment of the course (Hyytinen et al. 2019; Tuononen et al. 2020).

Generic skills can be integrated into field-specific contents in small steps. It is important that the generic skills included in the course are made visible as part of the course's objectives, contents and assessment. For example, if the objective is that the students learn to make arguments, the objectives, teaching methods and tasks of the course should consistently support the achievement of this objective (Hyytinen et al. 2019; Tuononen et al. 2020). It is equally essential that the selected assessment methods guide the learning of the desired generic skills.

Addressing generic skills in curricula. It is only possible to focus on a limited number of generic skills during a single course. Developing skills also takes time. This is why the integration of generic skills into individual courses is a good start, but the skills that are essential in terms of in-depth learning should be comprehensively and consistently addressed at the curriculum level (Virtanen & Tynjälä 2018; Tuononen et al. 2020). In practice, the precondition for this is clearly integrating the generic skills into the teaching methods, contents and assessment in the individual field of study throughout the student's study path. Incorporating generic skills diversely in field-specific studies supports the student's interest and motivation by offering them an authentic context in which to learn generic skills meaningfully (see Halpern 1998; Heijltjes et al. 2014; Virtanen & Tynjälä 2018; Hyytinen et al. 2019).

It is important to define intended learning outcomes that are appropriate in terms of the students' stage of studies and their level of knowledge and skills. The teaching and assessment methods should be designed to support the learning of the desired generic skills on courses specific to the field of study in a variety of ways (Halpern, 1998, 2014; Virtanen & Tynjälä 2018). The more the programme requires and expects of the students at the curriculum level, the more they are likely to learn (Arum & Roksa 2011).

One way to get started in defining development targets is curriculum mapping, in which the curriculum and all of its courses are analysed as a whole, ensuring that they cover the knowledge and skills expressed in the intended learning outcomes in terms of content, teaching methods and assessment. In addition, it may be necessary to monitor the students' competence and analyse student feedback as well as to identify development areas in the curriculum on this basis.

Students as active agents in learning generic skills. At best, higher education creates a favourable environment for the development of generic skills (e.g. Halpern 1998, 2014; Heijltjes et al. 2014; Virtanen & Tynjälä 2018; Tuononen et al. 2020). However, the students are ultimately responsible for their own learning and participation as well as for the goals they set and the way in which they direct their learning. Research on generic skills has found that the development of skills takes time and requires of the student an effort, active input and reflection on learning (e.g. Halpern 1998; Arum & Roksa 2011; Evens et al. 2013; Tuononen et al. 2019b). In order to develop generic skills, it is important to regularly do things in which these skills are genuinely used. For example, you become a better writer by writing (Appendix 5). Argumentation skills, on the other hand, can be learned by building arguments related to different contexts, including claims and justifications. Interpreting texts written by others and analysing the arguments presented in them also help to develop argumentation skills. There is no shortcut to learning generic skills.

It has sometimes been found that higher education students struggle to identify and name their generic capabilities (Tuononen et al. 2017, 2019b). To improve their generic skills, it is important that students reflect on, are aware of and regulate their thinking and learning processes (Halpern 1998; 2014; Hyytinen et al. 2020b). The feedback provided by the Kappas! project and participation in the test enabled students to gain new perspectives on their mastery of generic skills. Feedback received from teachers or peers during the studies may also help students identify their knowledge and skills and personal development needs (see Halpern 1998, 2014). Students experience interaction with their peers as a central and meaningful resource for their learning (Toom, St. Petersburg, Soini & Pyhältö 2017), and supporting it is also important in learning generic skills. Diverse generic skills are the key to working in expert positions and to drawing on the student's field-specific competence, which is why learning them actively and in a goal-oriented manner during the studies is essential.

6.3 Lessons learned from the Kappas! project

The Kappas! project produced new and important information on the level of and differences in higher education students' generic skills. In addition, valuable information was obtained on how a project of this type could be implemented even better in the future. Key development areas are related to student recruitment, better resource allocation to implementing the tests in higher education institutions, and developing the feedback reports for higher education institutions and students. From the time the sampling is planned, it is important to consider how students' active participation in the study could be promoted without compromising on its scientific objectives. For example, this could mean selecting a limited number of fields in higher education institutions and trying to test all students in the target group. Additionally, and most importantly, testing should be integrated into a course or unit, in which one session of the course would be set aside for students to take the test. This naturally also requires commitment from the teacher of the relevant course. If data collection is developed in this way, it will require sufficient resource allocation by the higher education institutions and close interaction with the national implementer. This is why it is important to consider the extent to which national resources could be allocated to higher education institutions for conducting the tests. Naturally, it should also be considered if the test could be taken and supervised remotely, which would be a resource-wise method of collecting data. The feedback on the Kappas! study received by higher education institutions and students was considered important. However, the reports for higher education institutions and students should be made more comprehensible and reader-friendly while reinforcing the principles of enhancement-led evaluation. To discuss the higher education institutions' results, webinars were organised, which proved to be effective venues for discussing generic skills in a broader context as part of higher education studies. Also offering a similar solution to students should be considered to support the development of their generic skills.

The Kappas! project has made the importance of generic skills more visible in higher education studies. At the Kappas! webinars, it was found that many participating higher education institutions have launched or are already making efforts to address generic skills better in teaching and curriculum work. For example, attention has been paid to teachers' pedagogical competence, and curricula are being developed further. Different teaching experiments and projects are also under way in higher education institutions. While projects and experiments are highly valuable for the future of generic skills, increasing the level of mastery will require more long-term and goal-oriented investments and resources of higher education institutions to mainstream better practices and attitudes that promote the development of generic skills in higher education. To support this, higher education institutions need more research evidence. More specific information on how students' generic skills develop during higher education studies, in particular, would help the institutions to understand the added value of the studies. Longitudinal research data will be needed to determine this.

Longitudinal research would make it possible to study in greater detail not only the development of skills but also the way in which generic skills are reflected in the student's learning, quality of learning and progress of studies on their study path. The research could also produce important additional information to support student admissions and the development of teaching at the levels of education preceding higher education. It could also advance understanding of the background factors that influence the development of generic skills and the possibilities of higher education to support the improvement of these skills. Such additional research would also help support continuous learning (see Virtanen, Postareff & Hailikari 2015).

Appendices

Appendix 1. Background information questions of CLA+ International test.

1. What is your gender?

1 = Man | 2 = Woman | 3 = Decline to state

2. Think of the parent or guardian who had the greatest influence on you as you were growing up. What is the highest level of education this parent or guardian has?

1 = Lower primary level (approx. 6 years) | 2 = Higher primary level (approx. 9 years) | 3 = Vocational/general upper secondary education (approx. 12 years) | 4 = Specialist vocational qualification (approx. 14 years) | 5 = Bachelor's degree (approx. 16 years) | 6 = Master's degree (approx. 18 years) | 7 = Licentiate or doctoral degree (approx. 20 years)

3. What is the current employment status of your mother or female guardian?

1 = Working full-time | 2 = Working part-time | 3 = Not working, but looking for work | 4 = Other status (e.g. caring for children at home, retired, in between jobs but not actively looking) | 5 = Unknown or not applicable

4. What is the current employment status of your father or male guardian?

1 = Working full-time | 2 = Working part-time | 3 = Not working, but looking for work | 4 = Other status (e.g. caring for children at home, retired, in between jobs but not actively looking) | 5 = Unknown or not applicable

5. How many books were there in your childhood home, excluding school books? One shelf meter contains on average 30 books.

1 = 0–10 | 2 = 11–25 | 3 = 26–100 | 4 = 101–200 | 5 = 201–500 | 6 = Over 500

6. Is your home language the same as the language of instruction at your higher education institution?

1 = Yes | 2 = No

7. Which of the following describes best your main field of study?

1 = Humanities or Art (visual arts, performing arts, languages, literature, history etc.) | 2 = Social Sciences and Education (educational science, sociology, political history, psychology, anthropology, communication, journalism etc.) | 3 = Business | 4 = Law | 5 = Science (life sciences, physical sciences, bio and environmental sciences, chemistry, information technology etc.) | 6 = Technology, Manufacturing, Architecture or Construction | 7 = Agriculture (including forestry, fishing, veterinary medicine etc.) | 8 = Health or Welfare (including medicine, the social sector and nursing science etc.) | 9 = Services (personal services, transport, environmental and security services etc.) | 10 = Not known or not specified

8. Roughly how many hours do you spend in the classroom in a typical week?

1 = 0 to 2 hours | 2 = 3 to 5 hours | 3 = 6 to 8 hours | 4 = 9 to 11 hours | 5 = 12 to 14 hours | 6 = 15 to 17 hours | 7 = 18 to 20 hours | 8 = Over 20 hours

9. Roughly how many hours do you spend working in a typical week?

1 = 0 to 2 hours | 2 = 3 to 5 hours | 3 = 6 to 8 hours | 4 = 9 to 11 hours | 5 = 12 to 14 hours | 6 = 15 to 17 hours | 7 = 18 to 20 hours | 8 = Over 20 hours

10. Roughly how many hours do you spend reading, processing, or learning study materials outside the classroom in a typical week?

1 = 0 to 2 hours | 2 = 3 to 5 hours | 3 = 6 to 8 hours | 4 = 9 to 11 hours | 5 = 12 to 14 hours | 6 = 15 to 17 hours | 7 = 18 to 20 hours | 8 = Over 20 hours

11. Which of the following best describes the most common type of course you have taken to date at a higher education institution? (Only select one)

1 = Seminars (based on discussions, more typically organised in classrooms than lecture halls) | 2 = Lectures (little or no discussion, more typically organised in lecture halls than classrooms) | 3 = Distance learning or online classes that do not require physical presence on the campus | 4 = Science laboratories (guided practical work, organised on the campus) | 5 = Work in art studios (guided or free) | 6 = Service learning or field work (instructed or free, off the campus) | 7 = Independent study (self-directed written work or a project with little guidance from a teacher or expert) | 8 = Most of my courses do not fit in any of these categories

12. How much effort did you put in to complete the written task?

1 = No effort at all | 2 = A little effort | 3 = A moderate amount of effort | 4 = A lot of effort | 5 = My best effort

13. How engaging did you find the written task?

1 = Not at all engaging | 2 = Slightly engaging | 3 = Moderately engaging | 4 = Very engaging | 5 = Extremely engaging

14. How much effort did you put in to answer the selected-response questions?

1 = No effort at all | 2 = A little effort | 3 = A moderate amount of effort | 4 = A lot of effort | 5 = My best effort

15. How engaging did you find the selected-response questions?

1 = Not at all engaging | 2 = Slightly engaging | 3 = Moderately engaging | 4 = Very engaging | 5 = Extremely engaging

16. How likely would you be to recommend this test to your friends?

1 = Extremely unlikely | 2 = Very unlikely | 3 = A little unlikely | 4 = Unsure | 5 = A little likely | 6 = Very likely | 7 = Extremely likely

17. How do you plan on using your test score? (Only select one)

1= Employment opportunities | 2 = Further study opportunities | 3 = CLA+ CareerConnect service | 4 = Personal enrichment | 5 = Other | 6 = Decline to specify

18. What is your mother tongue?

Finnish | Swedish | other

19. What is your age?

Under 20 years | 20 to 21 years | 22 to 23 years | 24 to 25 years | 26 to 27 years | over 27 years

20. Have you completed the matriculation examination?

No | Yes

21. Do you have some other degree or qualification other than the matriculation examination (select the highest level)?

No | Yes, a vocational upper secondary qualification | Yes, a UAS degree | Yes, a Bachelor's degree | Yes, a Master's degree | Yes, some other degree/qualification

22. What was your mother tongue grade in the matriculation examination?

I did not take the mother tongue test | Improbatur (I) | Approbatur (A) or lubenter approbatur (B) | Cum laude approbatur (C) | Magna cum laude approbatur (M) | Eximia cum laude approbatur (E) or laudatur (L)

23. What was your mathematics grade in the matriculation examination?

I did not take the mathematics test | Improbatur (I) | Approbatur (A) or lubenter approbatur (B) | Cum laude approbatur (C) | Magna cum laude approbatur (M) | Eximia cum laude approbatur (E) or laudatur (L)

24. In the matriculation examination, I took the test in the

Basic syllabus of mathematics | Advanced syllabus of mathematics | I did not take the mathematics test

Consider next the way you worked when solving the written task (and not the selected-response section) and respond to the following statements.

25. I made a plan for solving the problem comprised in the task.

Completely disagree | Disagree | Neither agree or disagree | Agree | Completely agree

26. Before writing my response, I considered what it would take to solve the task.

Completely disagree | Disagree | Neither agree or disagree | Agree | Completely agree

27. I thought of the areas of the task that I still had to solve.

Completely disagree | Disagree | Neither agree or disagree | Agree | Completely agree

28. I went through the steps of my plan for solving the tasks in my mind.

Completely disagree | Disagree | Neither agree or disagree | Agree | Completely agree

29. I constantly reviewed the work I had done while I was working on the task.

Completely disagree | Disagree | Neither agree or disagree | Agree | Completely agree

30. I planned my methods for solving the task carefully.

Completely disagree | Disagree | Neither agree or disagree | Agree | Completely agree

31. I backtracked during the task to check if my response made sense.

Completely disagree | Disagree | Neither agree or disagree | Agree | Completely agree

32. I stopped to reconsider a part of the task I had already solved.

Completely disagree | Disagree | Neither agree or disagree | Agree | Completely agree

33. I backtracked now and then to check that I had solved the task correctly.

Completely disagree | Disagree | Neither agree or disagree | Agree | Completely agree

34. I felt enthusiastic while completing the written task.

Not at all | Somewhat | Moderately | Very | Extremely

35. I felt stressed while completing the written task.

Not at all | Somewhat | Moderately | Very | Extremely

36. I felt calm while completing the written task.

Not at all | Somewhat | Moderately | Very | Extremely

37. I felt frustrated while completing the written task.

Not at all | Somewhat | Moderately | Very | Extremely

Appendix 2. Performance task scoring rubric

Area of assessment	Description	Not applicable	1	2	3	4	5	6
Analysis and problem solving	Making logical conclusions or inferences (or forming an opinion) and using the documents in the electronic library to support it (facts, ideas, calculated values or key properties).	• The response is not relevant to the topic. • The response could have been written without reading the documents in the first place. • The information has been copied directly from the documents, without analysis.	• May express or indicate a solution/ conclusion/ opinion. • Provides a minimal analysis to support the response (e.g. discusses briefly only one of the ideas in the document), or the analysis is completely faulty, illogical or unreliable, or not relevant to the solution/ conclusion/ opinion.	• Expresses or indicates a solution/ conclusion/ opinion. • Provides an analysis that addresses a few supportive ideas, some of which are incorrect, illogical, unreliable or irrelevant to the solution/ conclusion/ opinion.	• Expresses or indicates a solution/ conclusion/ opinion. • The response is justified sensibly but critical information has been omitted or misinterpreted, indicating that the documents have been analysed superficially and only partly understood. • May not explain conflicting information (if any).	• Gives a clear solution/ conclusion/ opinion. • The response is justified sensibly and it discusses several essential and reliable pieces of information in a way that indicates adequate document analysis and understanding. Some critical information has been omitted. • May attempt to explain conflicting information or present alternative solutions/ conclusions/ opinions (if possible).	• Expresses a clear solution/ conclusion/ opinion. • The response is well justified and discusses many relevant and reliable pieces of information in a way that indicates that the documents' content has been very well understood and analysed. • Refutes conflicting information or alternative solutions/ conclusions/ opinions (if possible).	• Expresses a clear solution/ conclusion/ opinion. • The response is comprehensively justified and contains almost all the relevant and reliable information, indicating excellent document analysis and understanding. • Thoroughly refutes conflicting evidence or alternative solutions/ conclusions/ opinions (if possible).

Area of assessment	Description	Not applicable	1	2	3	4	5	6
Writing effectiveness	Forming of structured and logically consistent arguments. Backing up the writer's view by discussing facts or ideas in greater detail (e.g. by explaining how the evidence is related to the problem, giving examples and highlighting particularly convincing evidence).	• Only scored as 'Not relevant' if the same score is given for analysis and problem solving.	• Does not form convincing arguments; the response may be written in an unsystematic and confused manner. • Does not discuss facts or ideas in detail.	• The arguments presented are too few or incorrect, exaggerated or very unclear. Information may have been presented unsystematically or in a way that undermines the writer's views. • More detailed consideration of facts or ideas is vague, irrelevant, inaccurate or unreliable (e.g. based entirely on the writer's opinion); information sources are often unclear.	• Provides few arguments, or the arguments are rather unclear. Provides relevant information in each response but does not build it into arguments. • Discusses facts or thoughts in detail a few times, some of them competently; information sources are sometimes unclear.	• Structure of the response indicates the writer's arguments and their logic but they are not self-evident. • Discusses facts or ideas in detail several times and refers to or quotes information sources.	• Structures the response logically and consistently, making it relatively easy to follow the writer's arguments. • Competently discusses facts or thoughts related to each argument in detail and refers to or quotes information sources.	• Structures the response logically and consistently, making the writer's arguments very easy to follow. • Discusses facts or ideas correctly and comprehensively in detail in relation to each argument and clearly refers to or quotes information sources.

Area of assessment	Description	Not applicable	1	2	3	4	5	6
Writing mechanics	• Ability to follow the rules of written language (congruence, tenses, capitalization, punctuation and spelling) as well as demonstrating command of the Finnish language, including syntax (sentence structure) and vocabulary (choices and use of words).	• Only scored as 'Not relevant' if the same score is given for analysis and problem solving.	• Mastery of grammatical rules is minimal and the response contains plenty of errors that make the text difficult to read or inadequate for assessment purposes. • Writes sentences which contain plenty of repetitions or which are incomplete and sometimes difficult to understand. • Uses simple vocabulary; some words are used inaccurately or in a way that leaves the meaning unclear.	• Mastery of grammatical rules is poor and the response contains many minor and some serious errors. • Consistently writes sentences in which similar structures and lengths are repeated, some of which may be difficult to understand. • Uses simple vocabulary, and some of the words may be used inaccurately or in a way that leaves the meaning unclear.	• Mastery of grammatical rules is moderate and the response contains many minor errors. • Writes sentences that are easy to read but whose structure and length vary little. • Uses vocabulary that conveys ideas adequately but is limited.	• Mastery of grammatical rules is good and there are few errors. Writes well-formed sentences whose structure and length vary to some extent. • Uses vocabulary that conveys ideas clearly but is limited.	• Shows extremely good mastery of grammatical rules. • Consistently writes well-formed sentences with varying structures and lengths. • Uses versatile and sometimes more demanding vocabulary that conveys ideas effectively.	• Shows excellent mastery of grammatical rules. • Consistently writes well-formed and complex sentences with varying structures and lengths. • Uses suitable vocabulary that is accurate, demanding and varied.

Appendix 3. CLA+ mastery levels and their descriptions (Zahner 2013).

Mastery level	Description
Below basic	Students who are below basic do not meet the minimum requirements to merit a basic level of mastery.
Basic	Students at the basic level should be able to demonstrate that they at least read the documents and made a reasonable attempt at an analysis of the details. They are also able to communicate in a manner that is understandable to the reader. They should also show some judgement about the quality of the evidence. Students at the basic level should also know the difference between correlation and causality and be able to interpret graphs.
Proficient	Students at the proficient level are able to use and analyse the documents diversely and provide cohesive arguments. Proficient students should be able to evaluate the reliability of the documents and the conclusions made on their basis. Additionally, students should be able to consider possible counterarguments in their responses. Proficient students have the ability to identify logical fallacies and accurately interpret quantitative evidence. They should be able to determine if a certain source of information is applicable to an argument.
Accomplished	Students at the accomplished level of mastery are able to analyse the information provided in the documents, extract relevant pieces of evidence, and make correct inferences about this information. Accomplished students are able to identify bias, evaluate the credibility of the sources, and craft an original and independent argument. They will refute some of the false claims within the documents and use this information to advance their argument. Accomplished students also have the ability to correctly identify logical fallacies, interpret and analyse qualitative and quantitative evidence, and incorporate this information into their response. Student responses are presented in a cohesive and organized fashion. There may be minor errors in writing fluency and mechanics, but they will not detract from the reader's comprehension of the text.
Advanced	Students at the advanced level demonstrate consistency and completeness, and show a command of the Finnish/Swedish language in their responses. Advanced students are able to synthesize the information across multiple documents and address the ambiguities in the data that are presented, such as outliers. They are able to structure their thoughts and understand causality. They are also able to identify gaps in logic and reasoning, take into account conditions and nuances, express more discreet or specific ideas and observations, and present different conditional conclusions if necessary.

Appendix 4. An example of a student report.

Enni Esimerkki | Sample University | enni@esimerkki.com CLA+
Academic year 2019–2020 Student Score Report www.cae.org/cla

Total Score: 1231
Mastery Level: Accomplished
Percentile Rank: 75

Your CLA+ Scores and Percentile Ranks

	Your Score	Your ranking in score distribution of your field of study (percentile score)	Your ranking in score distribution nationally (percentile score)
Total CLA+ Score	1231	76	75
Performance Task	1220	78	70
Selected-Response Questions	1243	77	80

Note: CLA+ scores range from approximately 400 to 1600.

Your CLA+ Mastery Level

	Your Mastery Level	Average for your field of study	Average for Finland
Total CLA+ Score Mastery Level	Accomplished	Accomplished	Proficient

Note: There are five Mastery Levels: Below Basic, Basic, Proficient, Accomplished, and Advanced.

Your CLA+ Subscores

PERFORMANCE TASK

Analysis & Problem Solving			Writing Effectiveness			Writing Mechanics		
	Average Scores			Average Scores			Average Scores	
Your Subscore	Field of study	Finland	**Your Subscore**	Field of study	Finland	**Your Subscore**	Field of study	Finland
3	3.2	2.8	**3**	2.8	3.1	**2**	3.5	3

SELECTED-RESPONSE QUESTIONS

Scientific & Quantitative Reasoning			Critical Reading & Evaluation			Critique an Argument		
	Average Scores			Average Scores			Average Scores	
Your Subscore	Field of study	Finland	**Your Subscore**	Field of study	Finland	**Your Subscore**	Field of study	Finland
362	370	400	**386**	378	380	**582**	600	415

Note: The Performance Task subscore categories are scored on a scale 1 to 6, and Selected-Response Question subscores are reported on a scale of approximately 200 to 800.

Enni Esimerkki | Sample University | enni@esimerkki.com CLA+
Academic year 2019–2020 Student Score Report www.cae.org/cla

UNDERSTANDING YOUR RESULTS

CLA+, a product of CAE (the Council for Aid to Education), is an assessment designed to measure generic skills.

CLA+ includes two major components: a Performance Task (PT) and a series of Selected-Response Questions (SRQs). The Performance Task presents students with a real-world situation that requires a purposeful written response. Students are asked to address the issue and recommend a course of action to resolve the conflict. They are instructed to support their responses by utilising information provided in a Document Library. This repository contains a variety of reference materials, such as technical reports, data tables, newspaper articles, office memoranda, and emails. Students have 60 minutes to complete this task.

In the second part of the test, students are asked to answer 25 SRQs. Ten questions measure scientific and quantitative reasoning and ten questions measure critical reading and evaluation. Another five questions call for students to critique arguments by identifying logical flaws. Like the PT, the 25 SRQs are document-based and require students to draw information from provided materials. Students have 30 minutes to complete this section of the assessment.

Your CLA+ Scores and Percentile Ranks

The scores in the first table represent your performance on the two sections of CLA+ and on the assessment as a whole.

Your score for the Total CLA+ Score is the average of the PT and SRQ section scores. Each CLA+ section score is determined by your subscores for that section, which are aggregated and converted to scale scores. This process results in a scale score distribution. CLA+ scores typically range roughly from 400 to 1600, though it is possible for individuals with exceptionally high performance on the assessment to receive scores above 1600.

The percentile score, which describes your ranking, shows how well you performed compared to other Finnish students who had completed the test and who were at the same stage of their studies. The first percentile score shows how you did among students of your field of study, while the second one indicates how you performed among all Finnish students at the same stage of the studies as you. For example, if your percentile score is 75, this means that you did better in the test than 75% of the students in your comparison group.

Mastery Level

Your CLA+ mastery level is presented in the second table on page 1. These levels are determined by your Total CLA+ Score. There are five mastery levels for the CLA+: Below Basic, Basic, Proficient, Accomplished, and Advanced. Your mastery level and the average mastery levels of the comparison groups are described in the second table.

Students who are **Below Basic** do not meet the minimum requirements to merit a basic level of mastery.

Students at the **Basic** level should be able to demonstrate that they read the documents and made a reasonable attempt at an analysis of the details. They are also able to communicate in a manner that is understandable to the reader. They should also show some judgement about the quality of the evidence. Students at the Basic level should also know the difference between correlation and causality and be able to interpret graphs.

Students at the **Proficient level** are able to use and analyse the documents diversely and provide cohesive arguments. Proficient students should be able to evaluate the reliability of the documents and the conclusions made on their basis. They should additionally be able to consider possible counterarguments in their responses. Proficient students also have the ability to identify logical fallacies and accurately interpret quantitative evidence. They should be able to determine if a certain source of information is applicable to an argument.

Students at the **Accomplished level** of mastery are able to analyse the information provided in the documents, extract relevant pieces of evidence, and make correct inferences about this information. Accomplished students are able to identify bias, evaluate the credibility of the sources, and craft an original and independent argument. They will refute some of the false claims within the documents and use this information to advance their argument. These students also have the ability to correctly identify logical fallacies, interpret and analyse qualitative and quantitative evidence, and incorporate this information into their responses. Student responses are presented in a cohesive and organized fashion. There may be minor errors in writing fluency and mechanics, but they will not detract from the reader's comprehension of the text.

Students at the **Advanced level** demonstrate consistency and completeness, and show a command of the Finnish language in their responses. Advanced students are able to synthesise the information across multiple documents and address the ambiguities in the data that are presented, such as outliers. They are able to structure their thoughts and understand causality. They are also able to identify gaps in logic and reasoning, take into account conditions and nuances, express more discreet or specific ideas and observations, and present different conditional conclusions if necessary.

Your CLA+ Subscores

Performance Task responses are scored across three skill areas: Analysis & Problem Solving, Writing Effectiveness, and Writing Mechanics. Each of these subscore categories is scored on a scale of 1 through 6. The top section of the Your Subscores table shows your Performance Task subscore and, as a benchmark value, the average subscore of students in your field of study and other Finnish students who took the CLA+ test and who are at the same stage of their studies as you. For information about the scoring system, please refer to CLA+ Performance Task scoring rubric at www.cae.org/claptrubric.

The SRQ section is also scored across three skill areas: Scientific & Quantitative Reasoning, Critical Reading & Evaluation, and Critique an Argument. Subscores for the SRQ section of CLA+ are determined by the number of correct answers in a given section, adjusted for the difficulty of the set of questions received, and then converted to a more interpretable scale using a linear transformation, with an approximate mean of 500 and an approximate standard deviation of 100. This scale results in SRQ subscores that have approximate ranges of 200 to 800. Students must attempt at least half of the questions to receive a score for this section. The lower section of the Your Subscores table shows your subscores and, as a benchmark value, the average subscore of students in your field of study and other students who took the CLA+ test and who are at the same stage of their studies as you.

MOVING FORWARD

Using the results in working life or further studies

You can use your CLA+ results for different purposes. For example, you can use them for self-development, or discuss them with your Personal Tutor or various learning experts. Instructions for developing generic skills are appended to the score report.

If you intend to apply for a place in further studies or move on to working life after graduation, the CLA+ test results are a document that provides employers or postgraduate study programmes with information on your mastery level in generic skills and thus the skills needed in working life and further studies. The percentile score you achieved in the CLA+ test shows how your score compares with the scores of others who took the test. The mastery level provides detailed information on the skills you demonstrated. You can use the results to supplement your applications for internships, jobs or places in further studies and add your Total CLA+ Score, the percentile score and mastery level to your CV.

If you have scored at Proficient, Accomplished, or Advanced on CLA+, you will also be eligible to receive a verified digital badge. For example, you can add the CLA+ badge to your online CV or LinkedIn profile.

Appendix 5. Additional attachment on developing generic skills sent to students.

You can develop your generic skills

You participated in the KAPPAS study. Based on your test response, you received feedback on how well you master different dimensions of generic skills. The tasks you completed in the study are challenging and require mastery of a wide range of generic skills. **Remember that generic skills can be learned**; they are not permanent characteristics. Learning generic skills takes time and continues throughout your higher education studies and later in working life. You can use the feedback you received in your portfolio and for developing your study skills. In this document, we explain how you can develop your generic skills.

What are generic skills and what are they needed for?

Generic skills include critical thinking, problem solving, evaluation of information sources, justification, self-regulation and writing skills. Generic skills are important in studies, working life and everyday life alike. They help you make use of your expertise and field-specific competence. Solving problems you encounter in life is easier when you are able to evaluate matters from different perspectives, including the effectiveness, logic and consequences of different solutions. The ability to assess and recognise your and other people's beliefs and perceptions is also useful. Writing is an important generic skill in itself but it also enables you to make visible your problem solving, reasoning and argumentation skills by enabling you to communicate your views to others.

How can I develop my generic skills?

Developing your generic skills is very simple: all you need to do is practise them regularly. In other words, you need to repeatedly do things in which different generic skills are used: for example, you become a better writer by writing.

You can improve your generic skills in your studies, at work or independently, alone or with other students. You can improve your skills by using different study methods that help you understand, apply and evaluate information better. It is also important to learn to challenge and recognise your own way of acting, allowing you to improve them. You can use the exercises presented below to develop your skills.

Exercises for improving generic skills

Analyse a task and plan how to do it

In this exercise, a task refers to any task related to studies or work, including an essay, an exam answer or a thesis. It may also be a task assigned to you by your supervisor at the workplace.

1. Analyse the task: think about what is required and what it takes to complete the task.
2. Based on the analysis, define your goals and plan how you can complete the task and achieve the goal.
3. Specify a timeline, starting with an overall schedule: is it a question of hours, days, weeks, or months? Sometimes you have a given deadline, at other times you need to determine the schedule yourself. In case of an exam answer, you are talking about minutes and hours, whereas the schedule for your thesis is based on weeks and months. The next step is to add detail to the schedule. To this end, divide the task into parts and think about what you need to have completed in an hour, by Monday, this week or this month. You should also determine if keeping to the overall schedule is possible when you schedule the parts of the task in this way.
4. As you proceed with the task, assess your achievements in the light of the goals, both in the middle and in the end. Based on your assessment, correct and complement your answer as needed.

This is not only a good exercise for developing generic skills but also an excellent strategy whenever you encounter different tasks and problems.

Put together and recognise an argument

First make sure you understand what an argument is. An argument seeks to persuade the listener and prove that a claim is correct, or to refute objections. The argument comprises at least a claim and arguments, but it may also contain counterarguments and underlying assumptions. Consequently, a claim alone does not add up to an argument. The elements of an argument are 1) a claim, or the matter on which you seek to persuade the listeners, 2) justification(s) stating why the listeners should accept the claim, 3) a counter-argument, or an alternative or opposite perspective on the claim, 4) background assumptions, or the basic premises behind the claim and the justifications that link the claim to the justifications and are often not stated expressly. In a good argument, the claim and the

justifications are credible and clearly relevant to the matter at hand. The justifications of a good argument are adequate to legitimise the claim, they do not conflict with the claim, and they are not based on personal opinions or emotions.

Write down the claim. Then consider ways to legitimise or justify it. Try to come up with an effective counter-argument. Finally consider the underlying assumptions of your argument.

You can also practise argumentation by analysing arguments in other texts, using the definition of argument above.

Read and discuss!

Read a large and diverse selection of texts and discuss them with others. Remember that all types of texts can help develop your generic skills: newspaper articles, non-fiction, literary fiction or online texts. It is useful to familiarise yourself with matters from different perspectives through different texts. Discuss what you have read with others. Are your interpretations different or similar? How do you evaluate the reliability of different texts?

When reading, try to interpret and analyse the text and explain it in your own words. Whilst reading, answer the following questions: What does the text say? What is the main claim of the text? What could the text mean? Is the text important in its own context or more broadly? How does the text relate to what you have learned before or to your own life? Also consider the underlying assumptions, different perspectives and possible inconsistencies in the text. When evaluating a text, useful questions include 'what, why, how, where, in what way'.

Write!

Write different texts regularly. You can find countless writing exercises on the Internet and in the list of references below. Make use of them. Practise summing up a text you have read, finding your own words rather than repeating those used in the text. Plan a preliminary structure for your text and edit it as the writing progresses. Learn how to tolerate incompletion: the text does not need to be ready immediately. Also remember that you do not need to write the text in order: you can start with the easiest part and proceed to the more challenging ones. The most important thing is to write in a versatile manner, whether the text is short or long, fact or fiction.

Ask and give feedback

Ask for feedback on your text, for example from a teacher or other students. It is easy to become blind to your own writing, and it is easier for outsiders to identify areas in need of improvement. Take the feedback you receive into account when you work on your text: structure it, eliminate repetition and other unnecessary content, select content and expand the views. Also learn to give constructive feedback in return. Reading other writers' texts will also help you in your own work. If getting feedback is not possible for some reason, get some distance to your text, for example by leaving it aside for a day or two. When you come back to it, you will see it with new eyes.

Evaluate the reliability of information

When reading a text, evaluate its reliability and consider the impact reliability has on the conclusions you draw based on the text. You can use the following checklist when evaluating its reliability:

1. Who wrote the text and for what purpose, and in what context does it appear?
2. Where was the text published? Was it on a website, in a newspaper or a local paper, a special interest or other magazine, a scientific journal, etc.? What conclusions can you make on the reliability of the text based on its publication channel?
3. What is the information presented in the text based on? For example, is it based on a survey, the writer's opinion or an individual's experiences?
4. What justification does the writer provide for their claims? Does the author rely on sources to justify the claims? Do the justifications given in the text support the claims?
5. Is there anything else in the text that could affect your evaluation of its reliability?

Recognise your own way of acting and learning outcomes

Stop to think about things and also assess your studies and learning objectives. What is the aim of your studies, what kind of competence do you seek and by what means? Are you trying to understand things, or do you perhaps attempt to learn things by rote? It is worth remembering that understanding things and their contexts may be difficult if your aim is to learn individual things by heart. Whereas rote learning usually only produces short-term learning outcomes, learning which aims at understanding will develop your competence and expertise more extensively and permanently.

REFERENCES

Each-hankkeen blogi. Kirjoituksia kriittisestä ajattelusta. https://blogs.helsinki.fi/each-project/.
Kniivilä, Sonja, Sari Lindblom-Ylänne, and Anne Mäntynen. Tiede ja teksti: tehoa ja taitoa tutkielman kirjoittami- seen. Helsinki: Gaudeamus, 2017.
Mälkki, Kaisu, Mansikka-aho, Anette. Kasvatustieteen Taskutuutori: Raketti ajattelun avaruuteen! Tampere: University of Tampere, 2020. http://urn.fi/URN:ISBN:978-952-03-1438-5
Silfverberg, Anu, Hirvonen, Elina. Sata sivua. Tekstintekijän harjoituskirja. Helsinki: Avain, 2010.
Svinhufvud, Kimmo. Kokonaisvaltainen Kirjoittaminen. Helsinki: Tammi, 2007.
Svinhufvud, Kimmo. Gradutakuu. Helsinki: Art House, 2015.

Appendix 6. Example of an institutional report

Academic Year 2019-2020 CLA+ Results **Institutional Report**

Sample Institution

CLA+ is a valuable tool that measures critical thinking and written-communication skills of students in higher education. Institutions use CLA+ to estimate institutional and individual student growth of these essential skills, measure the efficacy of curricular and other programs, and demonstrate individual, class, and institutional proficiency. CLA+ results give individual students an opportunity to better understand their strengths and areas for improvements in order to master the skills necessary for post-collegiate success.

CLA+ Digital Badging gives students who are proficient and beyond an opportunity to communicate these skills directly to employers. CLA+ results are a tool to measure growth on these skills and determine how your institution compares to other Finnish universities and universities of applied sciences as well as international benchmarks using CLA+.

For entering students, Sample Institution has a mean CLA+ Total Score of ##. A score of ## demonstrates ## mastery of the critical thinking and written communication skills measured by CLA+.

For exiting students, Sample Institution's mean CLA+ Total Score is ##. A score of ## signifies ## mastery of the skills measured by CLA+.

Mastery Levels

CLA+ Mastery Levels allow distinctions in student performance relative to students' proficiency in critical thinking and written communication. These levels contextualize CLA+ scores by interpreting test results in relation to the qualities exhibited by examinees. Each Mastery Level—Below Basic, Basic, Proficient, Accomplished, and Advanced—corresponds to specific evidence of critical-thinking and written- communication skills.

CLA+ Subscores

In addition to total scores, there are six subscores reported across CLA+. The Performance Task—an essay- based section of the exam—is scored in three skill areas: Analysis and Problem Solving, Writing Effectiveness, and Writing Mechanics. Students receive subscores for each skill category based on key characteristics of their written responses. Selected-Response Questions are also scored in three areas: Scientific and Quantitative Reasoning, Critical Reading and Evaluation, and Critique an Argument. These subscores are scored based on the number of correct responses that students provide.

Growth Estimates

The institutional report uses effect size as a measure of the amount of growth shown across classes. Effect size is reported in standard deviation units. (Standard deviation is a measure of the distance between the mean, or average, and all other values in a score set.) Effect size is calculated by subtracting the mean score of the entering students from the mean score of the exiting students and dividing this difference by the standard deviation of the entering students' scores.

Benchmark Data

This report contains benchmark data at both the national and international level. The Finland national benchmark data includes 1468 entering students and 829 exiting students. The international benchmark includes 71,403 United States university freshmen and 54,340 United States university seniors.

Academic Year 2019-2020

SECTION 1A: SUMMARY RESULTS, ENTERING

Number of Entering Students Tested

Number of Students

Summary of CLA+ Results

Total Scores

	Mean Score	Standard Deviation	25th Percentile Score	75th Percentile Score
Your Institution				
Finland	1071	117	996	1153
International Benchmark	1056	149	950	1163

PT Scores

	Mean Score	Standard Deviation	25th Percentile Score	75th Percentile Score
Your Institution				
Finland	1062	132	989	1148
International Benchmark	1049	170	933	1165

SRQ Scores

	Mean Score	Standard Deviation	25th Percentile Score	75th Percentile Score
Your Institution				
Finland	1081	156	966	1195
International Benchmark	1059	186	922	1196

Academic Year 2019-2020

SECTION 1B: SUMMARY RESULTS, EXITING

Number of Exiting Students Tested

Number of Students

Summary of CLA+ Results

Total Scores

	Mean Score	Standard Deviation	25th Percentile Score	75th Percentile Score	Effect-size, vs Entering
Your Institution					
Finland	1101	118	1024	1186	0.26
International Benchmark	1110	148	1009	1217	0.36

PT Scores

	Mean Score	Standard Deviation	25th Percentile Score	75th Percentile Score	Effect-size, vs Entering
Your Institution					
Finland	1100	132	1021	1180	0.29
International Benchmark	1102	171	998	1224	0.31

SRQ Scores

	Mean Score	Standard Deviation	25th Percentile Score	75th Percentile Score	Effect-size, vs Entering
Your Institution					
Finland	1103	161	986	1224	0.14
International Benchmark	1113	183	985	1250	0.29

Academic Year 2019-2020

SECTION 1C: SUMMARY RESULTS, BY FIELD OF STUDY

Total Mean Scores, by Field of Study

	Your Institution	Finland, Entering	Finland, Exiting
Humanities or Arts		1078	1123
Social Sciences and Education		1100	1138
Business & Law		1080	1104
Science		1104	1134
Engineering, Manufacturing, Architecture, or Construction		1055	1076
Agriculture		1046	1029
Health or Welfare		1072	1109
Services		1032	1078
Not Known or Specified		1062	1113

PT Mean Scores, by Field of Study

	Your Institution	Finland, Entering	Finland, Exiting
Humanities or Arts		1092	1132
Social Sciences and Education		1102	1164
Business & Law		1086	1114
Science		1077	1123
Engineering, Manufacturing, Architecture, or Construction		1030	1059
Agriculture		1033	1002
Health or Welfare		1080	1127
Services		1022	1069
Not Known or Specified		1029	1100

Academic Year 2019-2020

SECTION 1C: SUMMARY RESULTS, BY FIELD OF STUDY

SRQ Mean Scores, by Field of Study

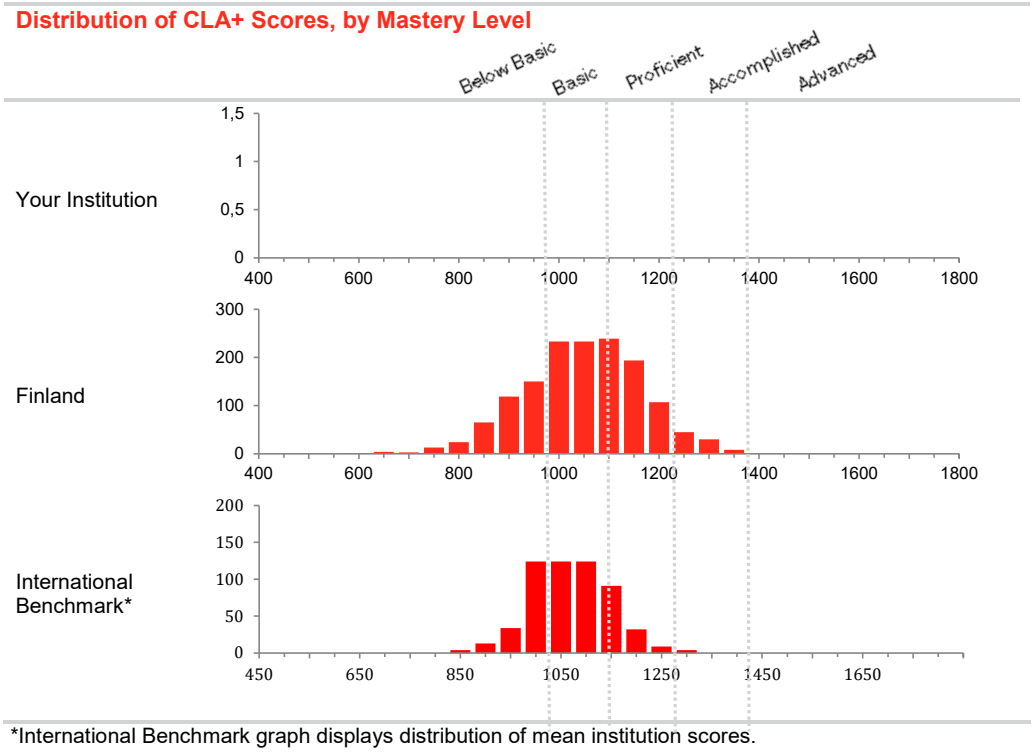
	Your Institution	Finland, Entering	Finland, Exiting
Humanities or Arts		1063	1115
Social Sciences and Education		1098	1112
Business & Law		1074	1093
Science		1131	1146
Engineering, Manufacturing, Architecture, or Construction		1080	1093
Agriculture		1059	1057
Health or Welfare		1065	1091
Services		1041	1088
Not Known or Specified		1095	1126

Academic Year 2019-2020

SECTION 2: DISTRIBUTION OF CLA+ SCORES, ENTERING AND EXITING STUDENTS

Distribution of CLA+ Scores, Entering

Dashed lines show approximate Mastery Level cut scores.



*International Benchmark graph displays distribution of mean institution scores.

Mastery Levels, Entering

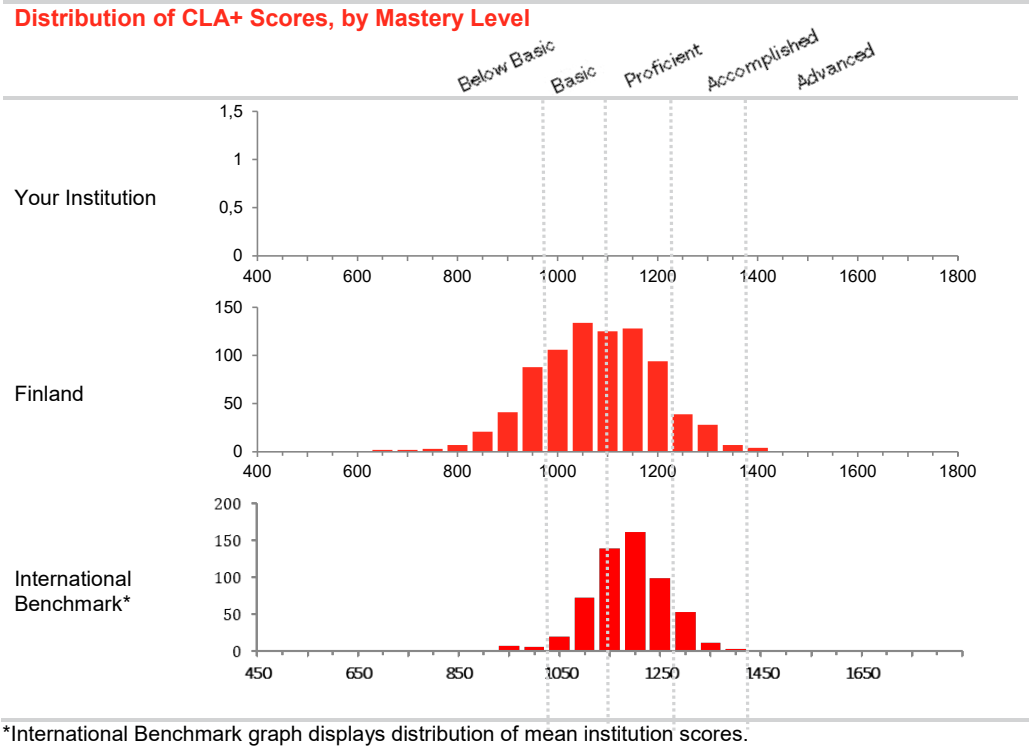
Class	Mean Score	Mean Mastery Level	Percent Below Basic	Percent Basic	Percent Proficient	Percent Accomplished	Percent Advanced
Your Institution							
Finland	1071	Basic	18%	39%	34%	9%	<1%

Academic Year 2019-2020

SECTION 2: DISTRIBUTION OF CLA+ SCORES, ENTERING AND EXITING STUDENTS

Distribution of CLA+ Scores, Exiting

Dashed lines show approximate Mastery Level cut scores.



*International Benchmark graph displays distribution of mean institution scores.

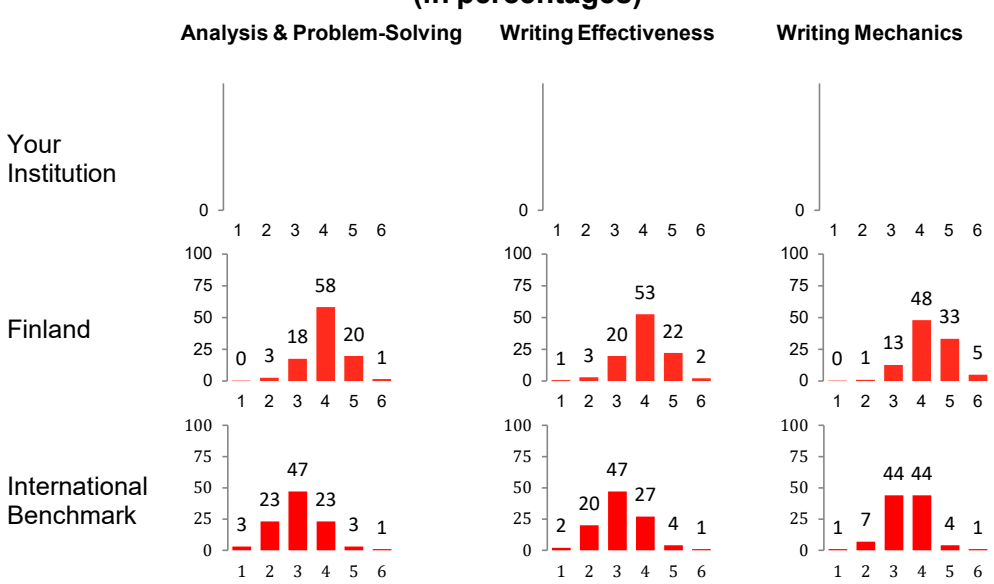
Mastery Levels, Exiting

Class	Mean Score	Mean Mastery Level	Percent Below Basic	Percent Basic	Percent Proficient	Percent Accomplished	Percent Advanced
Your Institution							
Finland	1101	Proficient	12%	36%	36%	15%	1%

Academic Year 2019-2020

SECTION 3A: CLA+ SUBSCORES, ENTERING

Performance Task Subscores (in percentages)



NOTE: The Performance Task subscore categories are scored on a scale of 1 through 6.

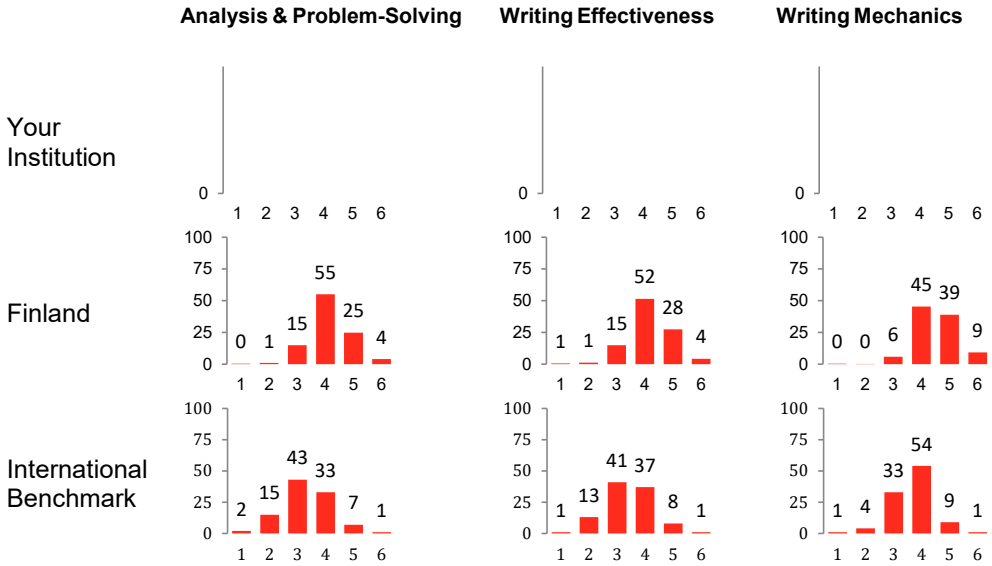
Selected-Response Question Subscores

	Scientific and Quantitative Reasoning			Critical Reading and Evaluation			Critique an Argument		
	Mean Score	25th Percentile	75th Percentile	Mean Score	25th Percentile	75th Percentile	Mean Score	25th Percentile	75th Percentile
Your Institution									
Finland	515	458	562	519	450	582	521	452	602
International Benchmark	508	438	593	507	429	574	510	429	591

Academic Year 2019-2020

SECTION 3B: CLA+ SUBSCORES, EXITING

Performance Task Subscores (in percentages)



NOTE: The Performance Task subscore categories are scored on a scale of 1 through 6.

Selected-Response Question Subscores

	Scientific and Quantitative Reasoning			Critical Reading and Evaluation			Critique an Argument		
	Mean Score	25th Percentile	75th Percentile	Mean Score	25th Percentile	75th Percentile	Mean Score	25th Percentile	75th Percentile
Your Institution									
Finland	532	458	615	525	450	582	520	452	602
International Benchmark	533	457	613	530	466	607	529	454	600

Academic Year 2019-2020

SECTION 3C: CLA+ SUBSCORES, BY FIELD OF STUDY

Analysis and Problem-Solving, by Field of Study

	Your Institution	Finland, Entering	Finland, Exiting
Humanities or Arts		4.1	4.3
Social Sciences and Education		4.2	4.6
Business & Law		4.1	4.1
Science		4.1	4.3
Engineering, Manufacturing, Architecture, or Construction		3.9	4.0
Agriculture		3.9	3.5
Health or Welfare		4.1	4.3
Services		3.8	4.0
Not Known or Specified		3.8	4.1

Writing Effectiveness, by Field of Study

	Your Institution	Finland, Entering	Finland, Exiting
Humanities or Arts		4.1	4.2
Social Sciences and Education		4.2	4.5
Business & Law		4.2	4.3
Science		4.1	4.4
Engineering, Manufacturing, Architecture, or Construction		3.8	3.9
Agriculture		3.9	3.7
Health or Welfare		4.1	4.3
Services		3.8	3.9
Not Known or Specified		3.8	4.2

Academic Year 2019-2020

SECTION 3C: CLA+ SUBSCORES, BY FIELD OF STUDY

Writing Mechanics, by Field of Study

	Your Institution	Finland, Entering	Finland, Exiting
Humanities or Arts		4.5	4.7
Social Sciences and Education		4.6	4.8
Business & Law		4.4	4.6
Science		4.3	4.5
Engineering, Manufacturing, Architecture, or Construction		4.1	4.3
Agriculture		4.1	4.1
Health or Welfare		4.3	4.6
Services		4.1	4.3
Not Known or Specified		4.2	4.6

Scientific & Quantitative Reasoning, by Field of Study

	Your Institution	Finland, Entering	Finland, Exiting
Humanities or Arts		494	534
Social Sciences and Education		514	535
Business & Law		506	520
Science		548	558
Engineering, Manufacturing, Architecture, or Construction		517	524
Agriculture		509	502
Health or Welfare		508	539
Services		498	512
Not Known or Specified		524	544

Academic Year 2019-2020

SECTION 3C: CLA+ SUBSCORES, BY FIELD OF STUDY

Critical Reading & Evaluation, by Field of Study

	Your Institution	Finland, Entering	Finland, Exiting
Humanities or Arts		522	531
Social Sciences and Education		528	530
Business & Law		525	530
Science		531	547
Engineering, Manufacturing, Architecture, or Construction		521	522
Agriculture		506	503
Health or Welfare		503	507
Services		496	532
Not Known or Specified		526	528

Critique-an-Argument, by Field of Study

	Your Institution	Finland, Entering	Finland, Exiting
Humanities or Arts		515	529
Social Sciences and Education		540	523
Business & Law		512	512
Science		536	515
Engineering, Manufacturing, Architecture, or Construction		509	521
Agriculture		510	527
Health or Welfare		531	516
Services		516	514
Not Known or Specified		516	537

Academic Year 2019-2020

SECTION 4A: ENTERING STUDENT EFFORT AND ENGAGEMENT

Student Effort Survey Responses (in percentages)

PERFORMANCE TASK

	No Effort At All	A Little Effort	A Moderate Amount of Effort	A Lot of Effort	My Best Effort
Your Institution					
Finland	1%	11%	46%	36%	5%
International Benchmark	1%	6%	37%	34%	21%

SELECTED-RESPONSE QUESTIONS

	No Effort At All	A Little Effort	A Moderate Amount of Effort	A Lot of Effort	My Best Effort
Your Institution					
Finland	4%	21%	52%	21%	3%
International Benchmark	3%	16%	45%	24%	11%

Student Engagement Survey Responses (in percentages)

PERFORMANCE TASK

	Not At All Engaging	Slightly Engaging	Moderately Engaging	Very Engaging	Extremely Engaging
Your Institution					
Finland	7%	25%	46%	18%	3%
International Benchmark	9%	19%	40%	26%	6%

SELECTED-RESPONSE QUESTIONS

	Not At All Engaging	Slightly Engaging	Moderately Engaging	Very Engaging	Extremely Engaging
Your Institution					
Finland	14%	35%	37%	11%	2%
International Benchmark	20%	30%	34%	13%	3%

Academic Year 2019-2020

SECTION 4B: EXITING STUDENT EFFORT AND ENGAGEMENT

Student Effort Survey Responses (in percentages)

PERFORMANCE TASK

	No Effort At All	A Little Effort	A Moderate Amount of Effort	A Lot of Effort	My Best Effort
Your Institution					
Finland	<1%	8%	42%	42%	7%
International Benchmark	1%	6%	35%	33%	25%

SELECTED-RESPONSE QUESTIONS

	No Effort At All	A Little Effort	A Moderate Amount of Effort	A Lot of Effort	My Best Effort
Your Institution					
Finland	2%	19%	45%	29%	5%
International Benchmark	3%	13%	42%	26%	16%

Student Engagement Survey Responses (in percentages)

PERFORMANCE TASK

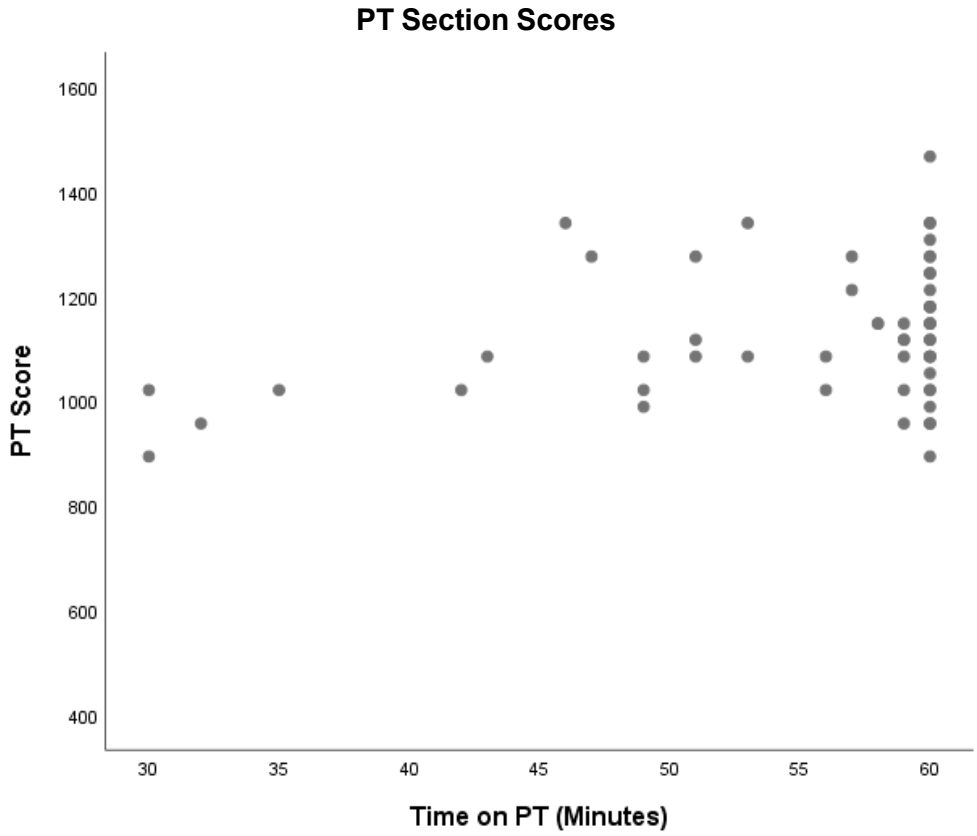
	Not At All Engaging	Slightly Engaging	Moderately Engaging	Very Engaging	Extremely Engaging
Your Institution					
Finland	7%	23%	44%	22%	4%
International Benchmark	9%	17%	38%	30%	7%

SELECTED-RESPONSE QUESTIONS

	Not At All Engaging	Slightly Engaging	Moderately Engaging	Very Engaging	Extremely Engaging
Your Institution					
Finland	14%	34%	35%	14%	3%
International Benchmark	17%	26%	37%	17%	4%

Academic Year 2019-2020

SECTION 5: PT SCORES BY TIME, EFFORT, AND ENGAGEMENT



PT Sections Scores by Effort on PT

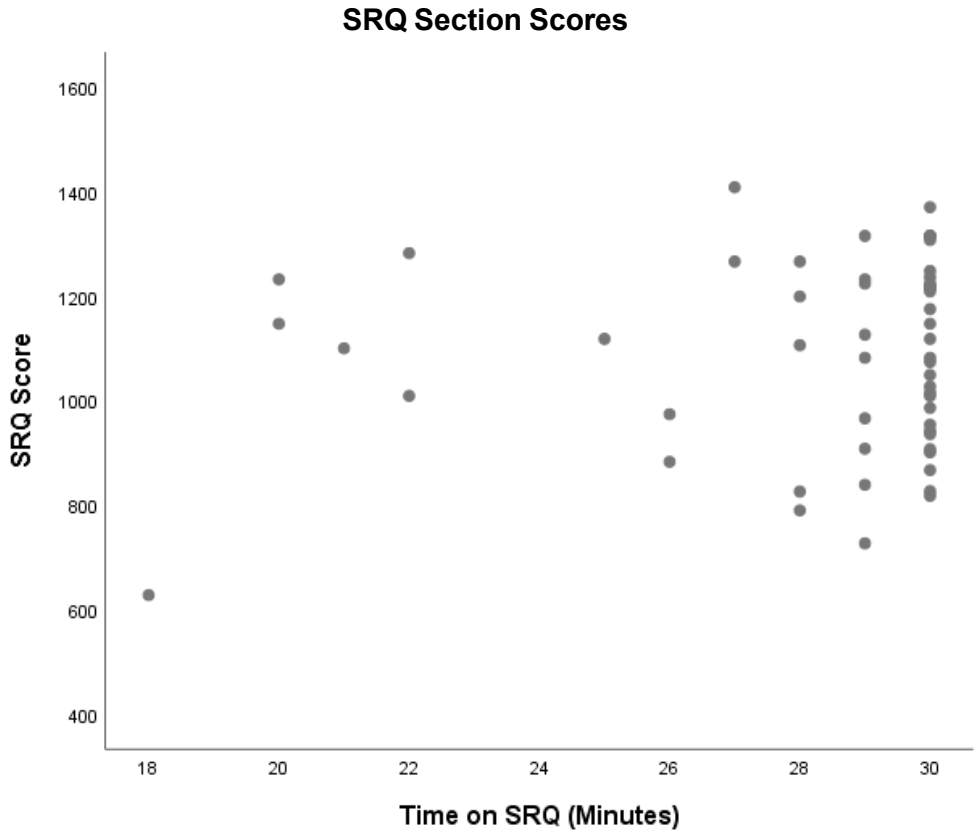
	No Effort At All	A Little Effort	A Moderate Amount of Effort	A Lot of Effort	My Best Effort
Entering					
Exiting					

PT Sections Scores by Engagement on PT

	Not At All Engaging	Slightly Engaging	Moderately Engaging	Very Engaging	Extremely Engaging
Entering					
Exiting					

Academic Year 2019-2020

SECTION 6: SRQ SCORES BY TIME, EFFORT, AND ENGAGEMENT



SR Sections Scores by Effort on SR

	No Effort At All	A Little Effort	A Moderate Amount of Effort	A Lot of Effort	My Best Effort
Entering					
Exiting					

SR Sections Scores by Engagement on SR

	Not At All Engaging	Slightly Engaging	Moderately Engaging	Very Engaging	Extremely Engaging
Entering					
Exiting					

Academic Year 2019-2020

SECTION 7: STUDENT SAMPLE SUMMARY

		Entering		Exiting	
Demographic Characteristic		N	%	N	%
Gender	Male				
	Female				
	Decline to State				
Parental Education	Primary/elementary				
	Lower secondary/middle school				
	Upper secondary/high school				
	Post-secondary non-tertiary/associate's degree/technical college				
	Bachelor's or equivalent				
	Master's or equivalent				
	Doctoral/professional or equivalent				
Mother/Female Guardian Employment Status	Working full-time				
	Working Part-time				
	Not working, but looking for a job				
	Other status (e.g., home duties, retired, in between jobs but not actively looking)				
	Unknown or not applicable				
Father/Male Guardian Employment Status	Working full-time				
	Working Part-time				
	Not working, but looking for a job				
	Other status (e.g., home duties, retired, in between jobs but not actively looking)				
	Unknown or not applicable				
Books in Home	0-10				
	11-25				
	26-100				
	101-200				
	201-500				
	More than 500				

Academic Year 2019-2020

SECTION 7: STUDENT SAMPLE SUMMARY

		Entering		Exiting	
Demographic Characteristic		N	%	N	%
Home Language	Same As Language of Instruction				
	Different from Language of Instruction				
Field of Study	General Program				
	Humanities or Arts				
	Social Sciences and Education				
	Business				
	Law				
	Science				
	Engineering, Manufacturing, Architecture, or Construction				
	Agriculture				
	Health or Welfare				
	Services				
	Not Known or Not Specified				
Hours Spent in Classroom per Week	0-2 Hours				
	3-5 Hours				
	6-8 Hours				
	9-11 Hours				
	12-14 Hours				
	15-17 Hours				
	18-20 Hours				
	More than 20 Hours				
Hours Spent Working per Week	0-2 Hours				
	3-5 Hours				
	6-8 Hours				
	9-11 Hours				
	12-14 Hours				
	15-17 Hours				
	18-20 Hours				
	More than 20 Hours				

Academic Year 2019-2020

SECTION 7: STUDENT SAMPLE SUMMARY

		Entering		Exiting	
Demographic Characteristic		N	%	N	%
Hours Spent Studying or Revising per Week	0-2 Hours				
	3-5 Hours				
	6-8 Hours				
	9-11 Hours				
	12-14 Hours				
	15-17 Hours				
	18-20 Hours				
	More than 20 Hours				
Most Common Course Type	Seminars				
	Lectures				
	Distance Learning/Online Classes				
	Science Laboratories				
	Art Studios				
	Service Learning/Field Work				
	Independent Study				
	Other				
Would You Recommend This Test to a Friend?	Extremely Unlikely				
	Very Unlikely				
	A Little Unlikely				
	Unsure				
	A Little Likely				
	Very Likely				
	Extremely Likely				
How Do You Plan on Using Scores	Employment Opportunities				
	Graduate School Opportunities				
	CLA+ CareerConnect				
	Personal Enrichment				
	Other				
	Decline to Specify				

Academic Year 2019-2020

APPENDIX A: INTRODUCTION

In 2002, the Collegiate Learning Assessment (CLA) was introduced as a major initiative of the Council for Aid to Education (CAE). Since its launch, the CLA has offered institutions a value-added approach to the measurement of higher-order thinking skills. The carefully designed questions in this examination require students to analyze, evaluate, and synthesize information as they demonstrate their ability to think critically and solve problems. Hundreds of institutions and hundreds of thousands of students have participated in the CLA testing program to date.

Initially, the CLA focused on helping institutions estimate their contributions to the development of students' higher-order thinking skills. As such, the institution rather than the student was the primary unit of analysis. In 2013, CAE expanded this scope with the introduction of CLA+. This enhanced version of the examination provides useful and reliable information about educational growth at the student level as well as the institutional level. Other features new to CLA+ include subscores for scientific and quantitative reasoning, critical reading and evaluation, and critiquing an argument. The addition of mastery levels also supports the reporting of criterion-referenced results in relation to skill proficiency.

CLA+ includes two major components: a Performance Task (PT) and a series of Selected-Response Questions (SRQs).

The Performance Task presents students with a real-world situation that requires a purposeful written response.
Students are asked to address an issue, propose the solution to a problem, or recommend a course of action to resolve a conflict. They are instructed to support their responses by utilizing information provided in a Document Library. This repository contains a variety of reference materials, such as technical reports, data tables, newspaper articles, office memoranda, and emails. A full PT includes four to nine documents in the library. Students have 60 minutes to complete this constructed-response task.

In the second part of the examination, students are asked to answer 25 Selected-Response Questions. Ten questions measure scientific and quantitative reasoning and ten measure critical reading and evaluation. Another five questions call for students to critique arguments by identifying logical flaws and questionable assumptions. Like the PT, the 25 SRQs are document-based and require students to draw information from provided materials. Students have 30 minutes to complete this section of the assessment.

CLA+ is a powerful assessment tool created to help teachers and students meet their educational objectives. The examination supports programmatic change, particularly in regard to higher-order thinking skills. It shows faculty members, school administrators, and other interested individuals the skill areas requiring attention on an institutional level to strengthen instruction and maximize learning. CLA+ also provides students with direct, formative feedback they can use to evaluate and reflect on their development on a personal level.

Institutions may wish to use CLA+ results to provide independent corroboration of competency-based learning, or to recognize students who have exhibited the higher-order thinking skills required for success in twenty-first century careers. Students may choose to share their results with potential employers or graduate schools as well to provide evidence of the skills they have acquired at their college or university. A single test cannot serve as the benchmark for all student learning within higher education, but there are certain skill areas deemed important by most educators across virtually all institutions. The higher- order thinking skills that CLA+ measures fall into this crucial category.

CLA+ allows institutions to benefit from a model of continuous improvement that positions educators as central actors in the relationship between assessment, instruction, and the learning process. Significantly, it provides educators with a frame of reference for determining the status of skill achievement within their institutions as well as the progress their students have made relative to the development of students at other colleges and universities. That said, CLA+ does not rank institutions; rather, it highlights differences between them that can identify opportunities for educational improvements. Similarly, CLA+ does not rank students but instead highlights areas where individuals excel or may need to focus more effort. CLA+ is an instrument designed to make a meaningful contribution to the improvement of teaching and learning. In this respect, it is in a league of its own.

Academic Year 2019-2020

APPENDIX B: METHODS

Introduction

CLA+ uses innovative questions and tasks to evaluate students' higher-order thinking skills. Each test form includes one Performance Task (PT) and 25 Selected-Response Questions (SRQs). The PT section measures three domains: analysis and problem solving, writing effectiveness, and writing mechanics. The SRQ section measures three domains as well: scientific and quantitative reasoning, critical reading and evaluation, and critiquing an argument, which involves the identification of logical flaws and questionable assumptions. Students have 90 minutes to complete the two sections of the assessment—60 minutes for the PT and 30 minutes for the SRQs.

Test results for CLA+ are delivered to institutions after administration windows have closed. Your institutional report presents scoring information for each section of the examination as well as total CLA+ performance. The report includes analyses of the PT score, the SRQ score, and the Total CLA+ score.

Section Scores

PT and SRQ scores indicate the mean, or average, performance of all students who completed each section. PT mean scores are calculated by adding three raw subscores—for analysis and problem solving, writing effectiveness, and writing mechanics—and converting the sum using a common scale. SRQ mean scores are also calculated by adding three raw subscores—for scientific and quantitative reasoning, critical reading and evaluation, and critique an argument—and converting this sum using a common scale. Total CLA+ scores are then calculated by averaging the PT and SRQ mean scores. For more information about the scaling process, please see the appendix on Scaling Procedures.

In addition to mean scores, your report includes 25th and 75th percentile scores, which characterize the score values earned by 25% and 75% of your students, respectively. For example, a 25th percentile score of 974 for the total CLA+ would inform you that 25% of your students earned 974 or less. Similarly, a 75th percentile score of 1096 would let you know that 75% of your students earned 1096 or less. The values that fall between the 25th and 75th percentile scores thus tell you the score values earned by 50% of your students. To extend the previous example, the 25th and 75th percentile scores reported would let you know that 50% of your students earned CLA+ Total Scores between 974 and 1096.

Effect-Sizes

Additionally, your institutional report shows effect sizes, which serve as estimates of growth between entering and exiting students in your institution.

Effect sizes characterize the amount of growth evident across classes by relating the CLA+ performance of entering students to that of exiting students. Please note that these statistics are available based on your students' participation in CLA+ testing. They do not take into account the performance of students at other institutions.

Effect sizes are calculated by subtracting the mean scores of the entering students from the mean scores of each subsequent class and dividing this difference by the standard deviation of the entering students' scores. (Standard deviation is a measure of the distance between the mean, or average, and all other values in a score set.) Effect sizes are reported in standard deviation units. By comparing effect sizes, you can gauge student growth over time and begin to analyze patterns of teaching and learning at your institution.

Academic Year 2019-2020

APPENDIX C: EXPLANATION OF YOUR RESULTS

This appendix provides guidance on interpreting the institutional results presented in your report. To ensure that the results in your report are based on a consistent sample, your students must complete all sections of the assessment, including the Performance Task, Selected-Response Questions, and the accompanying survey.
Please also note that CAE may set a minimum value on the number of students testing per class level that provide complete information in order to populate that class level in your institution's reports.

The results discussed in this appendix include various statistics, which relate performance in your institution to performance at other CLA+ universities and universities of applied sciences.

Summary of Results

In this section, summary statistics are presented for entering and exiting students in your institution. The results are also broken down by field of study. The first table in this section of the report is titled Number of Students Tested. This table specifies the number of students summarized in this report.

The next table, Summary of CLA+ Results presents a statistical overview of the students in your sample. It provides mean scores, standard deviations, and 25th and 75th percentiles for each class level tested (i.e. entering and exiting students). The effect sizes measuring difference between entering and exiting students are given in the table of exiting students.

The Mean Score column lists the average CLA+ scores for students in your sample. These scores are also considered your institutional CLA+ scores. The Standard Deviation column gives an estimate of the amount of scatter, or dispersion, in the scores at that class level, around the mean score. Higher standard deviations indicate that scores were further from the mean score at that class level and lower standard deviations indicate that scores were closer to the mean score at that class level.

The 25th Percentile Score column indicates which score separates the bottom 25% of students in that class from the top 75% of students in that class. Similarly, the 75th Percentile Score column indicates which score separates the bottom 75% of students in that class from the top 25% of students in that class. Additionally, 50% of students in the given class score between the 25th and 75th Percentile Scores.

The Effect Size vs. Entering column presents the growth estimates your institution. Effect sizes relate the performance of exiting students to that of entering students, allowing you to evaluate student learning outcomes over time relative to your own entering students. Effect sizes are reported in units of standard deviation established by the performance of entering students within your institution. An effect size of 0 indicates no difference in the performance of entering and exiting students, positive effect sizes indicate improved performance, and negative effect sizes indicate worse performance. Larger magnitudes of effect sizes (i.e., effect sizes further away from 0) indicate a greater effect and smaller magnitudes (i.e., effect sizes closer to 0) indicate more negligible changes in performance.

Distributions of Scores and Mastery Levels, by Entering and Exiting Students

This section of your institutional report focuses on Mastery Levels, which are criterion-referenced indicators of performance. On individual reports, Mastery Levels are determined by students' CLA+ Total Scores. On institutional reports, they are determined by each class level's mean CLA+ Total Score. There are five Mastery Levels: Below Basic, Basic, Proficient, Accomplished, and Advanced. Please see the appendix on Mastery Levels for a detailed description of these categories and the process through which they were derived. The first part of this section gives the distribution of CLA+ Total Scores for entering and exiting students, and includes vertical reference lines showing the approximate location of the cut scores for each Mastery Level. Note that although CAE never reveals the exact values of the Mastery Level cut scores, one can see approximate locations from these graphs. The second part of this section provides a summary of Mastery Levels by entering and exiting students. For each class level tested, the first column provides the mean CLA+ Total Score, the second column provides the Mastery Level corresponding to that mean score, and the third through seventh columns provide a frequency table for the Mastery Levels attained by the students entering and exiting students.

Academic Year 2019-2020

CLA+ Subscores, by Class

In this section, CLA+ Subscores are presented for entering and exiting students in your institution. The results are also broken down by field of study. Your report includes Total CLA+ scores as well as scores for the Performance Task (PT) and Selected-Response Questions (SRQs). These section scores based on item type are further divided into subscores based on skill categories. The three subscores for the PT indicate performance in Analysis and Problem Solving, Writing Effectiveness, and Writing Mechanics. The three subscores for the SRQs indicate performance in Scientific and Quantitative Reasoning, Critical Reading and Evaluation, and Critique an Argument, which involves the identification of logical flaws and questionable assumptions.

The first part of this section is Performance Task: Distribution of Subscores (in percentages). The charts indicate the distribution of subscores for each of the three skill categories by class level. The charts present the percentage of your students at each score value. Ranging from 1 to 6, each value is associated with a specific set of response characteristics. For more information about the scoring rubric, please see the appendix on Scoring CLA+.

The second part, Selected-Response Questions: Mean Subscores, provides summary statistics for the three skill categories measured in the SRQ section. The scores in this CLA+ section are determined by the number of correct responses and adjusted based on item difficulty. Each subscore is reported on a scale of approximately 200 to 800.

Mean Scores in this table reflect the average score received by each class for each of the three skill categories. The 25th Percentile Scores indicate the score values at or below which 25% of your students scored (again, by class level). The 75th Percentile Scores indicate the score values at or below which 75% of your students scored. By comparing results in the 25th and 75th columns, you can determine the range in which 50% of your students scored. Finally, the PT and SRQ subscore statistics are also presented by field of study.

Student Effort and Engagement, by Entering and Exiting Students

CLA+ ends with a set of survey questions, two of which are related to the assessment. One question asks students how much effort they put into completing the Performance Task (PT) and 25 Selected-Response Questions (SRQs). The other question asks students how engaging they found each section of the assessment to be. Students indicate their answers on a likert scale, ranging from "No effort at all" to "My best effort" and "Not at all engaging" to "Extremely engaging." The table in this section, Student Effort and Engagement Survey Responses, provides the percentage of students who selected each answer option by entering and exiting students.

The survey questions are designed to help institutions consider the role that effort and engagement may play in student performance on CLA+. Survey results may also be consulted when evaluating the impact that recruitment efforts have on student motivation.

Student Sample Summary

The final section of your institutional report includes a Student Sample Summary, which provides the number and percentage of students within your sample who meet various characteristics. These characteristics include, among others: gender, home language, field of study, and various parental and other study characteristics.. The characteristics are provided by students as part of a the post-assessment survey.

Academic Year 2019-2020

APPENDIX D: CLA+ TASKS

CLA+ includes one Performance Task (PT) and 25 Selected-Response Questions (SRQs). All items are administered online. Each PT consists of an open-ended prompt that asks students to provide a constructed response. Every SRQ presents students with four options and asks them to choose a single answer. The SRQs are further organized into three sets, each focusing on a different skill area.

Questions that appear on CLA+ call on students to use critical-thinking and written-communication skills as they perform cognitively demanding tasks. The integration of these skills mirrors the requirements of serious thinking and writing faced outside of the classroom.

Overview of the CLA+ Performance Task

Each PT asks students to answer an open-ended question about a hypothetical yet realistic situation. The prompt requires students to integrate analytical reasoning, problem solving, and written-communication skills as they consult materials in a Document Library and use them to formulate a response. The library includes a range of informational sources, such as letters, memos, summaries of research reports, newspaper articles, maps, photographs, diagrams, tables, charts, and interview notes or transcripts. Each PT is typically accompanied by four to nine documents, and students have 60 minutes to prepare their responses.

The first screen of each PT contains general instructions and an introduction to the scenario. The second screen is split. On the right side, students have a list of the informational sources in the Document Library. By using the pull-down menu, they can select and view each document. On the left side of the screen, students can read the question in the PT and enter their response in a field that has no word limit. An example of the split screen is shown on the following page.

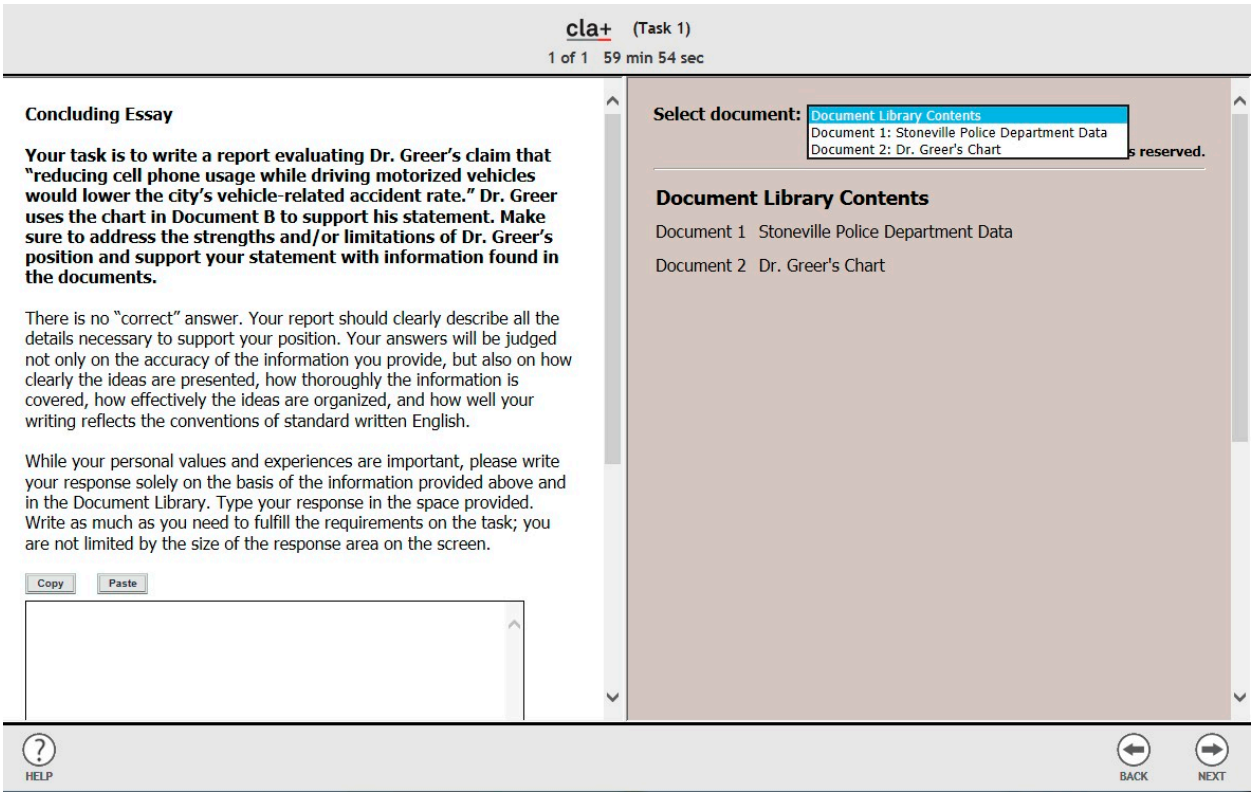
Each PT assesses a unique combination of skills—no two are exactly the same. Some PTs ask students to identify, compare, and contrast the strengths and limitations of alternate hypotheses, points of view, courses of action, etc. Other PTs ask students to review a collection of materials and choose amongst a set of options to solve a problem or propose a new solution to the problem. Still other PTs ask students to suggest or select a course of action that resolves conflicting or competing strategies and to provide a rationale for their decision, explaining why one approach is better than another.

For example, students may be asked to anticipate potential difficulties or hazards associated with different ways of addressing a problem, propose likely short- and long-term consequences of these strategies, and defend one or more of these approaches.

PTs require students to utilize higher order thinking skills, more specifically, to

- recognize information that is relevant and not relevant to the task at hand;
- analyze and understand data in tables and figures;
- evaluate the credibility of various documents;
- distinguish rational arguments from emotional ones;
- determine the difference between fact and opinion;
- identify questionable or critical assumptions;
- deal with inadequate, ambiguous, or conflicting information;
- spot deception, possible bias, and logical flaws in arguments;
- identify additional information that would help resolve issues;
- weigh different types of evidence;
- organize and synthesize information from several sources; and
- marshal evidence from different sources in a written response.

To view a sample PT, please visit the Sample Tasks section of CAE's website at www.cae.org/cla.



Overview of the CLA+ Selected-Response Questions

Like the PT, the 25 SRQs measure an integrated set of critical-thinking skills. Students utilize these skills to answer three sets of questions. The first measures scientific and quantitative reasoning, the second measures critical reading and evaluation, and the third (critique an argument) measures students' ability to identify logical fallacies and questionable assumptions. This final set requires students to detect logical flaws and questionable assumptions. Also like the PT, each question set is document-based and includes one to three informational sources of varying natures. Students are instructed to use these materials when preparing their answers within the 30 minutes provided.

The first two question sets require students to draw on the information and arguments provided in accompanying materials. Each set contains 10 questions, for a total of 20 questions.

Supporting documents for the Scientific and Quantitative Reasoning set discuss real-life research results. To answer questions in this section, students must apply critical- thinking skills that include

- making inferences and hypotheses based on given results,
- drawing a conclusion or deciding on a course of action to solve a problem,
- evaluating alternate conclusions, and
- recognizing when a text has open issues that require additional research.

Supporting documents for the Critical Reading and Evaluation set present debates, conversations, and literary or historical texts with opposing views on authentic issues. To answer questions in this section, students apply critical-thinking skills that include

- supporting or refuting a position, analyzing
- logic,
- identifying assumptions in arguments,
- evaluating the reliability of information,
- identifying connected and conflicting information, and
- making justifiable inferences.

In the Critique an Argument set, students are presented with a brief argument about an authentic issue and asked to analyze the argument. To answer the five questions in this section, students must apply critical-thinking skills that include

- evaluating the reliability of information, including

- evaluating the reliability of information (such as experimental design or data collection methodology),
- identifying information or quantitative data that is connected and conflicting,
- detecting questionable assumptions (such as implications of causation based on correlation),
- supporting or refuting a position,
- potential biases or conflicts of interest;
- detecting logical flaws and questionable assumptions;
- addressing additional information that could strengthen or weaken the argument; and
- evaluating alternate conclusions.

To view sample SRQs, please visit the Sample Tasks section of CAE's website at www.cae.org/cla.

Assessment Development

CAE has a team of experienced writers who work with educational researchers and editorial reviewers to generate ideas and design carefully constructed performance tasks (PTs), selected-response questions (SRQs), and supporting documents. Each group contributes to the development and revision of these materials.

Throughout development, writers, researchers, and reviewers refine materials to ensure that each PT can support a variety of different approaches. The prompt must be sufficiently focused to guide students purposefully while providing them with the flexibility to demonstrate independent thinking.
Questions must further be structured so students need to analyze and evaluate multiple sources of information from the Document Library to draw conclusions and justify their arguments.

Accompanying documents must present information in various formats and text types (e.g., tables, figures, news articles, editorials, emails, etc.). They must also provide enough information for students to formulate a number of reasonable arguments in response to the prompt. To achieve these goals, the development team drafts and revises a list of the intended content within each document. The list is used to check that each piece of information is clearly provided in the documents and that unwanted information is not embedded. During the editorial process, information is added and removed from the documents to ensure that students can reach approximately three to four different conclusions.
Typically, some conclusions are better supported by available evidence than others.

The document list also serves as a starting point for scorer training and is used in alignment with analytic descriptions in the PT scoring rubrics. After several rounds of revisions, the most promising PTs are selected for piloting. During this stage, student responses are examined to identify any lack of clarity in the prompt or any unintentional ambiguity or unuseful information in the accompanying documents. After revisions are made, PTs that meet expectations by eliciting a full range and variety of responses become operational.

The development process for SRQs is similar to the one used for PTs. Writers create documents that are based on real-life data and topics and can support questions measuring higher-order thinking skills. When crafting these documents, writers present valid and invalid assumptions and conclusions, devise alternate hypotheses and conclusions, incorporate flawed arguments, and leave some issues intentionally unanswered. These characteristics serve as a foundation for the creation of SRQs.

When reviewing item sets, editors work with writers to confirm that correct answer options are in fact correct based on information provided in the documents. Editors and writers also ensure that incorrect answer options are not potentially plausible. Throughout this process, the development team also checks to make sure that questions assess the intended critical-thinking skills.

After several rounds of revision, the most promising SRQs are selected for piloting. During this stage, student responses are examined to identify any errors or lack of clarity in questions and answer options. Responses are also reviewed to check whether accompanying documents contain unintentional ambiguity or unuseful information.

After revisions are made, SRQs that function well—questions that are of appropriate difficulty and that effectively discriminate between high- and low-performing students— become operational.

Academic Year 2019-2020

APPENDIX E: SCORING CLA+

Student responses to Performance Tasks are scored in three skill areas: Analysis and Problem Solving, Writing Effectiveness, and Writing Mechanics. Students receive criterion-referenced subscores for each skill category based on key characteristics of their written responses. These characteristics are described in detail within the Performance Task rubric, available on CAE's website at www.cae.org/cla.

Selected-Response Questions are scored based on the number of correct responses that students provide. Each of three question sets represents a skill area: Scientific and Quantitative Reasoning (10 questions), Critical Reading and Evaluation (10 questions), and Critique an Argument (5 questions). Because some question sets may be more difficult than others, the subscores for each category are adjusted to account for these differences and reported on a common scale. See the appendix on Scaling Procedures for more information about the scaling process.

The Scoring Process

During the piloting of Performance Tasks (PTs), all student responses are double-scored. Human scorers undertake this process, and the documentation they assemble is later used to train more scorers and program the machine-scoring engine for operational test administrations.

The rigorous training that candidates undergo to become certified CLA+ scorers further promotes the validity and reliability of the scoring process. Training sessions include an orientation to the prompts, scoring guides, and rubrics; extensive feedback and discussion after the evaluation of each student response; and repeated practice grading a wide range of student responses.

To ensure the continuous calibration of human scorers, CAE has also developed the E-Verification system for its online scoring interface. This system calibrates scorers by having them evaluate previously-scored responses, or "Verification Papers," throughout the scoring process. Designed to improve and streamline scoring, the E-Verification system periodically substitutes student responses with Verification Papers. These papers are not flagged for the scorers, and the system does not indicate when scorers have successfully evaluated them. However, if a scorer fails to assess a series of Verification Papers accurately, that scorer is targeted for additional coaching in a remediation process or is permanently removed from scoring.

Each student response receives three subscores in Analysis and Problem Solving, Writing Effectiveness, and Writing Mechanics. The subscores are assigned on a scale of 1 (lowest) to 6 (highest). Blank responses or responses unrelated to the task (e.g., what a student had for breakfast) are flagged for removal from test results.

Students also receive three subscores for the Selected-Response Questions (SRQs), one for each of the sets, which measure Scientific and Quantitative Reasoning, Critical Reading and Evaluation, and Argument Critique. Unless a student fails to start the section or is unable to finish due to a technical glitch or connection error, any unanswered SRQs are scored as incorrect. However, if a student does not attempt at least half of the SRQs, the student will not receive a score for the section. Subscores are determined by the number of correct responses, adjusted based on item difficulty, and reported on a common scale. The adjustment ensures that scoring is consistent, for example, whether a student answers seven questions correctly in an easier set or six in a more difficult one.

Scores are equated so that each subscore category has the same mean and standard deviation and all test forms are comparable. Score values range from approximately 200 to 800 for each SRQ section.

Academic Year 2019-2020

APPENDIX F: MASTERY LEVELS

Setting Standards for CLA+

Following the creation of CLA+, a standard-setting study was conducted to establish fair and defensible levels of mastery for the new and improved assessment. This formal study was held at CAE headquarters in New York City on December 12, 2013. Twelve distinguished panelists, representing a variety of educational and commercial sectors, were invited to participate. The table below lists each panelist.

During the standard-setting study, panelists defined descriptions of three mastery levels: Basic, Proficient, and Advanced. A fourth level, Accomplished, was added in November 2014 using the same methodology and the same panelists. Panelists' discussions were based on the CLA+ scoring rubric as well as the knowledge, skills, and abilities required to perform well on CLA+. The purpose of this activity was to develop consensus among the judges regarding each mastery level and to create a narrative profile of the knowledge, skills, and abilities necessary for CLA+ students.

During subsequent rating activities, panelists relied on these consensus profiles to make item performance estimates. Judges broke into three groups of four, and each group evaluated characteristics related to one mastery level. The groups then reconvened and reported their findings to the group at large so they could form final consensus on student performance at each of the mastery levels.

CLA+ Standard-Setting Study Participant List and Institutional Affiliation

Participant	Institution
Aviva Altman	Johnson & Johnson
Jon Basden	Federal Reserve
Mark Battersby	Capilano University (Canada)
Paul Carney	Minnesota State Technical and Community College
Anne Dueweke	Kalamazoo College
Terry Grimes	Council of Independent Colleges
Sonia Gugga	Columbia University
Marsha Hirano-Nakanishi	California State University System
Rachel L. Kay	McKinsey & Company
Michael Poliakoff	American Council of Trustees and Alumni
Elizabeth Quinn	Fayetteville State University
Paul Thayer	Colorado State University

CLA+ Mastery Levels

CAE uses outcomes from the 2013 standard-setting study to distinguish between CLA+ students with varying knowledge, skills, and abilities as measured by the assessment. On individual reports, Mastery Levels are determined by students' Total CLA+ scores. On institutional reports, they are determined by each class level's mean Total CLA+ score.

Institutions should not use mastery levels for purposes other than the interpretation of test results. If an institution wishes to use the attainment of CLA+ mastery levels as part of a graduation requirement or the basis for an employment decision, the institution should conduct a separate standard-setting study with this specific purpose in mind.

The following table summarizes each level of mastery and provides a description of students below the basic level of mastery.

Mastery Level	Description
Below Basic	Students who are below basic do not meet the minimum requirements to merit a basic level of mastery.
Basic	Students at the basic level should be able to demonstrate that they at least read the documents, made a reasonable attempt at an analysis of the details, and are able to communicate in a manner that is understandable to the reader. Students should also show some judgment about the quality of the evidence. Students at the basic level should also know the difference between correlation and causality. They should be able to read and interpret a bar graph, but not necessarily a scatter plot or comprehend a regression analysis. Tables may be out of reach for basic students as well.
Proficient	Students at the proficient level should be able to extract the major relevant pieces of evidence provided in the documents and provide a cohesive argument and analysis of the task. Proficient students should be able to distinguish the quality of the evidence in these documents and express the appropriate level of conviction in their conclusion given the provided evidence. Additionally, students should be able to suggest additional research and/or consider the counterarguments. Minor errors in writing need to be defined rigorously. Proficient students have the ability to correctly identify logical fallacies, accurately interpret quantitative evidence, and distinguish the validity of evidence and its purpose. They should have the ability to determine the truth and validity of an argument. Finally, students should be able to know when a graph or table is applicable to an argument.
Accomplished	Students at the accomplished level of mastery should be able to analyze the information provided in the documents, extract relevant pieces of evidence, and make correct inferences about this information. Accomplished students should be able to identify bias, evaluate the credibility of the sources, and craft an original and independent argument. When appropriate, students will identify the need for additional research or further investigation. They will refute some, but not all of the counterarguments within the documents and use this information to advance their argument. Accomplished students also have the ability to correctly identify logical fallacies, accurately interpret and analyze qualitative and quantitative evidence (e.g., graphs and charts), and incorporate this information into their argument. Students will be able to correctly identify false claims and other sources of invalid information and integrate this information in their responses. Student responses are presented in a cohesive and organized fashion. There may be infrequent or minor errors in writing fluency and mechanics, but they will not detract from the reader's comprehension of the text.
	Students at the advanced level demonstrate consistency, completeness, and show a command of the English language in their response. They have a level of sophistication that is not seen in the proficient or basic levels. Advanced students create and synthesize the provided evidence, are comfortable with ambiguity, are able to structure their thoughts,

Advanced	understand causality, add new ideas, and introduce new concepts in order to create or seek new evidence. They think about conditions and nuances and express finer points and caveats by proposing a conditional conclusion. The students at this level display creativity and synthesis, while understanding the finer points in the documents. For example, advanced students will be able to synthesize the information across multiple documents and address the ambiguities in the data that are presented, such as outliers and knowing how sample size affects outcomes. Advanced students will also be able to identify and highlight gaps in logic and reasoning.

Academic Year 2019-2020

APPENDIX G: DIAGNOSTIC GUIDANCE

Interpreting CLA+ Results

CLA+ test results can be used to evaluate an institution's overall performance on tasks measuring higher-order thinking skills. Test results can also be used to determine an individual student's areas of relative strength and weakness.

Examining performance across both CLA+ sections can serve as a comprehensive diagnostic exercise since the combination of necessary knowledge, skills, and abilities differs for the Performance Task (PT) and the Selected-Response Questions (SRQs). The PT measures Analysis and Problem Solving, Writing Effectiveness, and Writing Mechanics, while the SRQs measure Scientific and Quantitative Reasoning, Critical Reading and Evaluation, and Critique an Argument (the detection of logical flaws and questionable assumptions).

SRQ subscores are assigned based on the number of questions answered correctly; this value is then adjusted to account for item difficulty, and the adjusted value is converted to a common scale. Established in relation to the test performance of entering students in the fall of 2013, the scale has a mean of 500 and a standard deviation of 100. SRQ subscores thus range from approximately 200 to 800.

PT subscores are assigned on a scale of 1 (lowest) to 6 (highest). Unlike the SRQ subscores, PT subscores are not adjusted for difficulty. These subscores remain as is because they are intended to facilitate criterion-referenced interpretations. For example, a score of "4" in Analysis and Problem Solving signifies that a response has certain qualities (e.g., "Provides valid support that addresses multiple pieces of relevant and credible information..."). Any adjustment to the score would compromise this interpretation.

The ability to make a claim such as, "Our students seem to be doing better in Writing Effectiveness than in Analysis and Problem Solving," is clearly desirable. These types of observations can be made by comparing the distributions for each subscore in Section 3 of your institutional report. Please examine these test results in combination with the PT scoring rubric as well, available on CAE's website at www.cae.org/claptrubric.

CLA+ Mastery Levels further contextualize PT and SRQ subscores by interpreting test results in relation to the qualities exhibited by examinees. Each Mastery Level corresponds to specific evidence of critical-thinking and written-communication skills. Please see the appendix on Mastery Levels for detailed information about each Mastery Level.

Academic Year 2019-2020

APPENDIX H: SCALING PROCEDURES

Converting CLA+ Scores to a Common Scale

To provide CLA+ Total Scores, CAE converts PT and SRQ Scores to a common scale of measurement. This process allows us to combine score values from the two different assessment tasks.

For each Performance Task (PT), the three subscores are added to produce a raw section score. The raw section score is then converted to a common scale of measurement, yielding a scaled section score. The conversion produces scale scores that maintain comparable levels of proficiency across performance tasks and test forms. So, for example, a CLA+ PT Score would indicate the same percentile rank regardless of the task a student received.

For the PT, CAE uses a linear transformation when converting raw scores to scale scores. The process creates a scale score distribution for CLA+ entering students that has the same mean and standard deviation as their combined SAT Math and Critical Reading (or converted ACT) scores. The transformation was defined using data from students in the norm sample (i.e. those who took CLA+ between 2013 and 2018). This type of scaling preserves the shape of the raw score distribution and maintains the relative standing of students. For example, the student with the highest raw score on a PT will also have the highest scale score for that task, the student with the next highest raw score will be assigned the next highest scale score, and so on.

This scaling practice ensures that a very high PT raw score (not necessarily the highest possible score) corresponds approximately to the highest SAT (or converted ACT) score. Similarly, a very low PT raw score would be assigned a scale score value close to the lowest SAT (or converted ACT) score. On rare occasions when students earn exceptionally high or low raw PT scores, their scale scores may fall outside the normal SAT Math and Critical Reading score range of 400 to 1600.

For the Selected-Response Questions (SRQs), the raw subscores (for the three skill categories measured by the three question sets) are determined based on the number of correct responses. These raw subscores are first equated and then placed on a common scale. This process adjusts the subscores based on the difficulty of the item sets so the subscores have the same mean and standard deviation across all question sets. Comparisons can then be made across test forms.

Using a linear transformation, CAE then converts the equated subscores to a more interpretable scale with a mean of 500 and standard deviation of 100, again, based on data from the norm sample. This scale produces SRQ subscores ranging from approximately 200 to 800, similar to the subsections of the SAT.

The weighted average of the SRQ subscores (i.e., the average after accounting for number of items per set) is then transformed again, using the same scaling parameters as the PT. As before, the process creates a scale score distribution for CLA+ entering students that has the same mean and standard deviation as their combined SAT Math and Critical Reading (or converted ACT) scores. The transformation is based on data from the norm sample. The application of common parameters places both CLA+ section scores on the same scale.

Finally, CLA+ Total Scores are calculated by taking the average of the CLA+ PT and SRQ Scores. Thus, students who do not complete or provide scorable responses for both sections of the assessment do not receive CLA+ Total Scores.

Academic Year 2019-2020

APPENDIX I: MOVING FORWARD

The information presented in your institutional report is designed to help you better understand the contributions your institution has made toward student learning. Yet, the report alone provides only a snapshot of student performance. By combining it with other tools and services that CLA+ has to offer, the institutional report can become part of a powerful evaluation and enrichment strategy. It can help you and your institution target specific areas of improvement and align teaching, learning, and assessment effectively to enhance student performance over time.

We encourage institutions to examine CLA+ performance closely and review the results carefully with their educators. Institutions can extend these analyses by linking student-level CLA+ outcomes with other data sources and pursuing in- depth sampling. Collaboration with peer institutions and participation in professional development opportunities can support institutions and their educators further by showing how research findings can inform teaching practices and help improve student learning.

Using your Student Data File, you can relate student-level CLA+ results to data you collect on course-taking patterns, grade achievement, and other topics of inquiry. CLA+ subscores in Analysis and Problem Solving, Writing Effectiveness, Writing Mechanics, Scientific and Quantitative Reasoning, Critical Reading and Evaluation, and Critique an Argument can contribute to analyses of portfolios, student surveys, and other sources by helping you focus on specific areas that may benefit from improvement. Internal analyses conducted through in-depth sampling can help you generate hypotheses and develop a basis for additional research.

CLA+ can offer peer group comparisons, but the true strength of peer learning comes through collaboration. CAE facilitates cooperative relationships among CLA+ institutions by encouraging the formation of consortia. Moreover, CAE hosts web conferences that periodically feature campuses engaged in promising work with CLA+.

CAE also provides workshops geared toward helping institutions maximize the utility of their Student Data Files. In these sessions, CAE researchers work with institutional staff, showing them ways to dig deeper into student results so they can answer questions about performance on CLA+ and identify areas of strength or weakness. To reserve one of these sessions for your institution, please email clateam@cae.org.

Finally, our professional development services shift the focus from assessment outcomes to pedagogical tools in Performance Task Academies. These two-day, hands-on training workshops offer faculty members guidance in the creation of their own performance tasks. Modeled on the structure of CLA+ tasks and designed to support the teaching objectives of individual courses, faculty-developed tasks can be used as classroom exercises, homework assignments, or even local-level assessments. To learn more about Performance Task Academies, please consult the Events page on the CAE website.

We encourage institutions to explore a system of continuous improvement driven by the diagnostic potential of CLA+. When used in combination, our programs and services reinforce the belief that institutions must connect teaching, learning, and assessment in authentic and meaningful ways to strengthen and advance their students' higher-order thinking skills.

Without your contributions, CLA+ would not be on the exciting path it is on today. We thank you for your participation and look forward to your continued involvement!

Council for Aid to Education
215 Lexington Avenue
Floor 16
New York, NY 10016

CAE

Appendix 7. Descriptive analyses and multivariate regression models for the data set.

Table 1. CLA+ mean scores (av.) of participating higher education students and standard error (s.e.) and standard deviation of (s.d.) of the mean in the entire data set as well as by stage of studies and higher education sector. Number of observations = n

	Total score				PT score				SRQ score			
	n	av.	s.e.	s.d.	n	av.	s.e.	s.d.	n	av.	s.e.	s.d.
Entire data set	2,300	1,075	7	118	2,379	1,074	8	133	2,315	1,072	8	159
Initial stage students	1,469	1,061	10	117	1,520	1,054	10	129	1,479	1,063	12	158
Final stage students	831	1,090	8	116	859	1,095	10	135	836	1,082	9	158
University students	1,227	1,127	5	111	1,265	1,132	7	123	1,231	1,118	9	162
UAS students	1,073	1,044	9	110	1,114	1,039	10	127	1,084	1,045	10	150
Initial stage university students	785	1,116	6	108	811	1,111	7	125	788	1,118	10	154
Final stage university students	442	1,138	9	113	454	1,156	9	116	443	1,117	16	172
Initial stage UAS students	684	1,027	12	108	709	1,020	12	118	691	1,030	15	152
Final stage UAS students	389	1,063	10	109	405	1,060	13	133	393	1,061	12	146

Table 2. CLA+ mean scores (av.) of participating higher education students by gender in the entire data set and by stage of studies and higher education sector. Number of observations = n.

	Total score				PT score				SRQ score			
	Men		Women		Men		Women		Men		Women	
	av.	n	av.	n	av.	n	av.	n	av.	n	av.	n
Entire data set	1,066	1,139	1,082	1,108	1,042	1,169	1,098	1,152	1,088	1,145	1,062	1,113
Initial stage students	1,050	664	1,067	769	1,021	681	1,077	800	1,080	667	1,052	773
Final stage students	1,081	475	1,099	339	1,064	488	1,122	352	1,096	478	1,074	340
University students	1,125	628	1,128	572	1,108	644	1,147	594	1,140	630	1,105	573
UAS students	1,034	511	1,052	536	1,006	525	1,066	558	1,060	515	1,035	540
Initial stage university students	1,112	361	1,119	406	1,083	371	1,126	422	1,140	362	1,107	407
Final stage university students	1,138	267	1,138	166	1,134	273	1,170	172	1,140	268	1,102	166
Initial stage UAS students	1,016	303	1,032	363	986	310	1,043	378	1,047	305	1,017	366
Final stage UAS students	1,052	208	1,075	173	1,026	215	1,092	180	1,073	210	1,056	174

Table 3. Multivariate regression models estimated for university students' CLA+ test scores: regression coefficients and their standard errors (in brackets). Statistically significant ($p < 0.05$) coefficients in bold

	University students		
	PT score	SRQ score	Total score
Standard term	**857.1 (34.6)**	**885.0 (51.6)**	**1,060.7 (29.2)**
Stage of studies			
Initial stage	reference group		
Final stage	**32.7 (9.2)**	-9.6 (14.6)	13.8 (9.0)
Field of study			
Humanities or Arts	reference group		
Social Sciences and Education		-2.6 (27.2)	
Business and Law		-13.1 (21.5)	
Science (including ICT)		**48.6 (16.3)**	
Engineering, Manufacturing, Architecture or Construction		**35.3 (17.1)**	
Agriculture		**78.2 (22.5)**	
Health or Welfare		40.2 (26.4)	
Services		**48.2 (13.9)**	
Gender			
Male	reference group		
Female		**-49.7 (14.1)**	
Mother tongue grade in the matriculation examination			
No matriculation examination	reference group		
A or B	12.7 (24.9)	-13.0 (28.8)	5.8 (23.8)
C	18.4 (20.4)	-2.2 (21.4)	3.1 (21.2)
M	**65.9 (18.5)**	26.1 (23.8)	36.5 (19.8)
E or L	**99.5 (19.5)**	**68.6 (26.6)**	**78.4 (23.0)**

University students			
Other degree/qualification besides matriculation examination			
No other degree/qualification	reference group		
Vocational upper secondary qualification	-4.2 (14.0)		**-22.8 (11.2)**
Higher education degree	**41.6 (12.9)**		21.1 (1.2)
Number of books in the childhood home			
0 to 10 books	reference group		
11 to 25 books		10.3 (46.7)	10.4 (28.5)
26 to 100 books		29.4 (35.5)	11.1 (21.2)
101 to 200 books		64.0 (34.9)	38.0 (22.6)
201 to 500 books		62.1 (40.4)	39.3 (24.4)
Over 500 books		**101.0 (39.3)**	**64.5 (21.8)**
Parental education			
At most primary	reference group		
Secondary	**46.5 (19.3)**		
Specialist vocational qualification or similar	**61.8 (21.1)**		
Bachelor's degree	**60.2 (19.4)**		
Master's degree	**62.3 (17.9)**		
Scientific postgraduate degree	**74.4 (26.9)**		
Student effort in the CLA+ test			
Little or no effort	reference group		
A moderate amount of effort	**124.7 (20.9)**	**122.2 (49.5)**	**113.6 (44.3)**
A lot of effort	**147.0 (20.9)**	**179.7 (44.2)**	**156.8 (41.7)**
My best effort	**129.0 (24.0)**	**181.7 (48.0)**	**164.7 (43.3)**
Student engagement with the CLA+ test			
Not at all or slightly engaging	reference group		
Moderately engaging	-2.0 (9.5)		
Very engaging	**32.2 (11.6)**		
Extremely engaging	**37.0 (17.7)**		

Table 4. Multivariate regression models estimated for UAS students' CLA+ test scores: regression coefficients and their standard errors (in brackets). Statistically significant ($p < 0.05$) coefficients in bold

	UAS students		
	PT score	SRQ score	Total score
Standard term	**743.7 (45.4)**	**902.1 (29.8)**	**870.2 (21.9)**
Stage of studies			
Initial stage	reference group		
Final stage	**32.3 (6.8)**	22.3 (14.6)	**26.3 (9.8)**
Field of study			
Humanities or Arts	reference group		
Social Sciences and Education	-24.8 (41.8)		
Business and Law	**37.3 (15.4)**		
Science (including ICT)	37.9 (23.8)		
Engineering, Manufacturing, Architecture or Construction	10.8 (16.6)		
Agriculture	29.3 (30.5)		
Health or Welfare	**44.9 (16.3)**		
Services	26.8 (19.9)		
Gender			
Male	reference group		
Female	**25.8 (10.6)**	**-35.9 (11.4)**	
Mother tongue grade in the matriculation examination			
No matriculation examination	reference group		
A or B	**28.1 (10.6)**	26.0 (14.6)	**28.9 (9.2)**
C	**42.3 (8.9)**	**51.7 (15.3)**	**50.4 (8.9)**
M	**54.5 (8.8)**	**95.9 (12.8)**	**75.0 (8.8)**
E or L	**106.9 (15.0)**	**137.1 (17.0)**	**132.2 (11.4)**

	UAS students		
Number of books in the childhood home			
0 to 10 books	reference group		
11 to 25 books	-2.8 (13.0)	4.2 (18.7)	13.7 (15.0)
26 to 100 books	**27.1 (11.1)**	16.8 (19.7)	30.2 (15.3)
101 to 200 books	23.9 (12.5)	36.7 (19.2)	**42.4 (14.6)**
201 to 500 books	**28.4 (11.4)**	29.3 (21.9)	**39.5 (16.2)**
Over 500 books	**73.2 (27.6)**	**58.8 (24.1)**	**71.3 (22.7)**
Student effort in the CLA+ test			
Little or no effort	reference group		
A moderate amount of effort	89.7 (55.7)	**60.1 (27.0)**	**64.7 (23.1)**
A lot of effort	**144.1 (56.2)**	**93.7 (32.2)**	**98.2 (20.6)**
My best effort	**184.9 (53.9)**	**95.0 (30.4)**	**125.6 (21.6)**
Student engagement with the CLA+ test			
Not at all or slightly engaging	reference group		
Moderately engaging	21.3 (18.1)	-1.6 (14.1)	
Very engaging	**40.4 (19.5)**	16.9 (14.2)	
Extremely engaging	**46.2 (20.7)**	**42.4 (18.8)**	

Appendix 8. Age distributions of higher education students who participated in the study.

Table 1. Age distributions of initial stage university and UAS students in the data set

Age class	Universities		UASs	
	n	%	n	%
Under 20 years	174	22	88	12
20 to 21 years	354	44	243	34
22 to 23 years	128	16	127	18
24 to 27 years	83	10	104	15
Over 27 years	64	8	146	21
Total	803	100	708	100

Table 2. Age distributions of final stage university and UAS students in the data set

Age class	Universities		UASs	
	n	%	n	%
Under 22 years	78	17	34	8
22 to 23 years	211	47	146	36
24 to 27 years	118	26	113	28
Over 27 years	45	10	111	27
Total	452	100	404	100

Appendix 9. Questions in the feedback survey addressed to higher education institutions.

FEEDBACK QUESTIONNAIRE FOR HIGHER EDUCATION INSTITUTIONS ON THE KAPPAS! PROJECT

1. What motivated your higher education institution to take part in the Kappas! assessment project?
2. What did you think worked well in the implementation of the Kappas! assessment project from the perspective of your institution?
3. What did you find challenging in the implementation of the Kappas! assessment project from the perspective of your institution?
4. Has your higher education institution been able to/does it intended to use the findings of the Kappas! assessment project in the development of teaching? If yes, how? If no, why?
5. How would you improve the Kappas! assessment project from the perspective of your institution?
6. Would your higher education institution participate in the Kappas! assessment project again?
7. Do you have any other comments about the Kappas! assessment project from the perspective of your institution?
8. Your institution is
   - a university
   - a university of applied sciences
   - do not wish to state

Appendix 10. Questions of student interviews conducted for the final report.

Test and test situation

- When did you take the test?
- Why did you decide to take the test?
- How were you invited to the test?
- How did you experience the test and the test situation?

Test results and feedback (report, attachment, possible badge)

- How did you feel about receiving the feedback? (Did the feedback come too late?)
- Did you think the feedback was adequate?
- What did you think of/did you go through the support material?
- (How did you feel about being awarded the badge?)
- What do you intend to do with the feedback?
- How would you develop the feedback so that it could support you in improving your generic skills further?
- How could feedback received through assessments of this type be used in higher education institutions from the student's perspective?

Generic skills

- How important do you find generic skills?
- How do you feel your higher education studies have supported you in learning and improving your generic skills?
- Have your ideas about generic skills changed since you completed the test and received feedback? /Did your idea of your own generic skills change?
- Can you think of anything else that has not yet been covered in this interview?

REFERENCES

Arffman, I. & Nissinen, K. 2015. Lukutaidon kehitys PISA-tutkimuksissa. In J. Välijärvi, P. Kupari, A. K. Ahonen, I. Arffman, H. Harju-Luukkainen, K. Leino, M. Niemivirta, K. Nissinen, K. Salmela-Aro, M. Tarnanen, H. Tuominen-Soini, J. Vettenranta & R. Vuorinen. Millä eväillä osaaminen uuteen nousuun? PISA 2012 -tutkimustuloksia. Ministry of Education and Culture publications 2015:6, 28–49.

Arum, R. & Roksa, J. 2011. Academically adrift: Limited learning on college campuses. Chicago: University of Chicago Press.

Ashwin, P. 2020. Transforming university education: a manifesto. London: Bloomsbury Academic.

Badcock, P. B. T., Pattison, P. E. & Harris, K-L. 2010. Developing generic skills through university study: a study of arts, science and engineering in Australia. Higher Education 60 (4), 441–458. https://doi.org/10.1007/s10734-010-9308-8

Barrie, S. C. 2007. A conceptual framework for the teaching and learning of generic graduate attributes. Studies in Higher Education 32 (4), 439–458. DOI: 10.1080/03075070701476100

Bloom, B. S., Englehart, M. B., Furst E.J., Hill, W. H. & Krathwohl, D. R. 1956. Taxonomy of Educational Objectives: The Classification of Educational Goals. New York: McKay.

Borowiec, K. & Castle, C. 2019. Using rater cognition to improve generalizability of an assessment of scientific argumentation. Practical Assessment, Research, and Evaluation 24 (8). DOI: https://doi.org/10.7275/ey9d-p954

De Graaf, P. M. 1986. The impact of financial and cultural resources on educational attainment in the Netherlands. Sociology of Education 59 (4), 237–246.

Ercikan, K. & Pellegrino, J. 2017. Validation of Score Meaning Using Examinee Response Processes for the Next Generation of Assessments. In K. Ercikan & J. Pellegrino (eds.) Validation of score meaning for the next generation of assessments. The use of response processes. NY: Routledge, 1–8.

European Commission. 2013. High Level Group on the Modernisation of Higher Education. Report to the European Commission on improving the quality of teaching and learning in Europe's higher education institutions. Luxembourg: Publications Office of the European Union.

European Commission. 2019. Key competences for lifelong learning. Luxembourg: Publications Office of the European Union.

Evens, M., Verburgh, A. & Elen, J. 2013. Critical thinking in college freshmen: The impact of secondary and higher education. International Journal of Higher Education 2 (3), 139–151. https://doi.org/10.5430/ijhe.v2n3p139

Finnish Education Evaluation Centre. 2020. Kehittävä arviointi kansallisessa koulutuksen arviointikeskuksessa. Helsinki: FINEEC

Halpern, D. F. 1998. Teaching critical thinking for transfer across domains: Disposition, skills, structure training, and metacognitive monitoring. American Psychologist 53 (4), 449–455. https://doi. org/10.1037/0003-066X.53.4.449

Halpern, D. F. 2014. Thought and knowledge: An introduction to critical thinking. New York, London: Psychology Press.

Hambleton, R. K. & Patsula, L. 1998. Adapting tests for use in multiple languages and cultures. Social Indicators Research 45, 153–171.

Heijltjes, A., Tamara van Gog, T., Leppink, J. & Paas, F. 2014. Improving critical thinking: Effects of dispositions and instructions on economics students' reasoning skills. Learning and Instruction 29, 31–42.

Hiltunen, J. & Nissinen, K. 2018. Erinomaiset matematiikan osaajat vuoden 2015 PISA-tutkimuksessa. Teoksessa J. Rautopuro & K. Juuti (eds.) PISA pintaa syvemmältä. PISA 2015 Suomen pääraportti. Kasvatusalan tutkimuksia 77. Jyväskylä: Finnish Educational Research Association, 213–234.

Hyytinen, H. 2015. Looking Beyond the Obvious: Theoretical, Empirical and Methodological Insights into Critical Thinking. Doctoral dissertation. University of Helsinki. Studies in Educational Sciences 260. https://helda.helsinki.fi/bitstream/handle/10138/154312/LOOKINGB.pdf?sequence=1

Hyytinen, H., Holma, K., Shavelson, R. J. & Lindblom-Ylänne, S. 2014. The complex relationship between students' critical thinking and epistemological beliefs in the context of problem solving. Frontline Learning Research 6, 1–15.

Hyytinen, H., Nissinen, K., Ursin, J., Toom, A. & Lindblom-Ylänne, S. 2015. Problematising the equivalence of the test results of performance-based critical thinking tests for undergraduate students. Studies in Educational Evaluation 44, 1–8. https://doi.org/10.1016/j.stueduc.2014.11.001

Hyytinen, H., Postareff, L. & Lindblom-Ylänne, S. 2020a. Challenges in exploring individual's conceptions of knowledge and knowing: Examples of research on university students. In E. K. Kallio (ed.) Development of Adult Thinking: Interdisciplinary Perspectives on Cognitive Development and Adult Learning. London: Routledge, 177–190.

Hyytinen, H. & Toom, A. 2019. Developing a performance assessment task in the Finnish higher education context: conceptual and empirical insights. British Journal of Educational Psychology 89, 551–563. https://doi.org/10.1111/bjep.12283.

Hyytinen, H., Toom, A. & Postareff, L. 2018. Unraveling the complex relationship in critical thinking, approaches to learning and self-efficacy beliefs among first-year educational science students. Learning and Individual Differences 67, 132–142.

Hyytinen, H., Toom, A. & Shavelson, R. J. 2019. Enhancing scientific thinking through the development of critical thinking in higher education. In M. Murtonen & K. Balloo (eds.) Redefining scientific thinking for higher education. Cham: Palgrave Macmillan, 59–78. DOI: https://doi. org/10.1007/978-3-030-24215-2_3.

Hyytinen, H., Ursin, J., Silvennoinen, K., Kleemola, K. & Toom. A. 2020b. The dynamic relationship between cognitive response processes and metacognition in critical thinking assessments. Manuscript undergoing assessment.

Jääskelä, P., Nykänen, S. & Tynjälä, P. 2018. Models for the development of generic skills in Finnish higher education. Journal of Further and Higher Education 42 (1), 130–142. DOI: 10.1080/0309877X.2016.1206858

Karjalainen, A. 2002. Mitä benchmark-arviointi on? In K. Hämäläinen & M. Kaartinen-Koutaniemi (eds.) Benchmarking korkeakoulujen kehittämisvälineenä. Publications of the Finnish Higher Education Evaluation Council 13:2002. Helsinki: Edita, 10–19.

Kleemola, K., Hyytinen, H. & Toom, A. 2021. Exploring internal structure of a performance-based critical thinking assessment for new students in higher education. Assessment and Evaluation in Higher Education. https://doi.org/10.1080/02602938.2021.1946482

Leighton, J. 2017. Using think-aloud interview and cognitive labs in educational research. NY: Oxford University Press.

Leighton, J. P. 2019. The risk–return trade-off: Performance assessments and cognitive validation of inferences. British Journal of Educational Psychology 89 (3), 441–455.

Leino, K., Nissinen, K., Puhakka, E. & Rautopuro, J. 2017. Lukutaito luodaan yhdessä. Kansainvälinen lasten lukutaitotutkimus (PIRLS 2016). University of Jyväskylä, Finnish Institute for Educational Research.

Lemons, P. P. & Lemons, J. D. 2017. Questions for Assessing Higher-Order Cognitive Skills: It's Not Just Bloom's. Life Sciences Education 12 (1), 47–58.

McClelland, D. C. 1973. Testing for competence rather than testing for "intelligence". American Psychologist 28 (1), 1–14. https://doi.org/10.1037/h0034092

Messick, S. 1994. Alternative modes of assessment, uniform standards of validity. A Research Report. Princeton: Educational Testing Service.

Miettinen, R. 2019. 21. vuosisadan kompetenssit. OECD kasvatuksen kielen uudistajana. Kasvatus (50) 3, 203–215.

Ministry of Education and Culture. 2017. Working together for the world's best education. Policies on promoting internationality in higher education and research 2017–2025. Ministry of Education and Culture publications 2017:11

Muukkonen, H., Lakkala, M., Lahti-Nuuttila, P., Ilomäki, L., Karlgren, K. & Toom, A. 2020. Assessing the Development of Collaborative Knowledge Work Competence: Scales for Higher Education Course Contexts. Scandinavian Journal of Educational Research 64 (7), 1071–1089. DOI: 10.1080/00313831.2019.1647284

OECD. 2019. PISA 2018 results (Volume II): Where all students can succeed. PISA, OECD Publishing, Paris.

Opinion of the European Economic and Social Committee on Digitalisation, AI and Equity – How to strengthen the EU in the global race of future skills and education, while ensuring social inclusion. 2020. Official Journal of the European Union, C 14/46.

Shavelson, R. J. 2010. Measuring college learning responsibly: Accountability in a new era. Stanford, CA: Stanford University Press.

Shavelson, R. J., Zlatkin-Troitschanskaia, O. & Marino, J. 2018. International performance assessment of learning in higher education (iPAL): Research and development. In O. Zlatkin-Troitschanskaia, M. Toepper, H. A. Pant, C. Lautenbach & C. Kuhn (eds.) Assessment of learning outcomes in higher education – Cross-national comparisons and perspectives. Wiesbaden, Germany: Springer, 193–214.

Sikora, J., Evans, M. & Kelley, J. 2019. Scholarly culture: How books in adolescence enhance adult literacy, numeracy and technology skills in 31 societies. Social Science Research 77, 1–15.

Sulkunen, S. & Nissinen, K. 2012. Heikot lukijat Suomessa. In S. Sulkunen & J. Välijärvi (eds.) Kestääkö osaamisen pohja? PISA 09. Ministry of Education and Culture publications 2012:12, 46–61.

Suorsa, O. & Sainio, J. 2020. Osaamisesta yliopistojen maisteriseurantakyselyiden 2017–2019 pohjalta. Aarresaari. Universities' working life and career services network.

Talman K., Borodavkin M., Kanerva A. & Haavisto E. 2018. Ammattikorkeakoulujen uuden digitaalisen valintakokeen kehittäminen – määrittelyvaiheen tulokset : Tutkimusraportti. Metropolia University of Applied Sciences. http://urn.fi/URN:ISBN:978-952-328-119-6

Toom, A., Pietarinen, J., Soini, T. & Pyhältö, K. 2017. How does the learning environment in teacher education cultivate first year student teachers' sense of professional agency in the professional community? Teaching and Teacher Education 63, 126–136. DOI: 10.1016/j.tate.2016.12.013

Tremblay, K., Lalancette, D. & Roseveare, D. 2012. Assessment of higher education learning outcomes AHELO. Feasibility study report. Volume 1: design and implementation. OECD.

Tuononen, T. 2019. Employability of university graduates: The role of academic competences, learning and work experience in the successful transition from university to working life. Helsinki Studies in Education 46. Helsinki: University of Helsinki. http://urn.fi/URN:ISBN:978-951-51-5123-0

Tuononen, T., Hyytinen, H., Hailikari, T. & Toom, A. 2020. Akateemisten asiantuntijataitojen integroimisen tukeminen koulutusohjelma- ja kurssitasolla. In A. Virtanen, J. Helin & P. Tynjälä (eds.) Työelämäpedagogiikka korkeakoulutuksessa. Jyväskylä: Finnish Institute for Educational Research, 96–100. https:// www.tyopeda.fi/julkaisut/tyoelamapedagogiikka-korkeakoulutuksessa.pdf.

Tuononen, T., Kangas, T., Carver, E. & Parpala, A. 2019a. Yliopisto-opintojen anti viisi vuotta valmistumisen jälkeen - Tukivatko yliopisto-opinnot työelämätaitojen kehittymistä työuran näkökulmasta?
Yliopistopedagogiikka 1. Työelämäteemanumero osa II. https://lehti.yliopistopedagogiikka.fi/2019/02/08/yliopisto-opintojen-anti-tyoelamataitojen-kehittyminen/.

Tuononen, T., Parpala, A. & Lindblom-Ylänne, S. 2017. The transition from university to working life - An exploration of graduates perceptions of their academic competences. In E. Kyndt, V. Donche, K. Trigwell & S. Lindblom-Ylänne. (eds.) Higher Education Transitions - Theory and Research. Routledge: Taylor & Francis Group, 238–253.

Tuononen, T., Parpala, A. & Lindblom-Ylänne, S. 2019b. Graduates' evaluations of usefulness of university education, and early career success – a longitudinal study of the transition to working life. Assessment & Evaluation in Higher Education, 1–14.

Universities Act. 2009. Section § (28.12.2018/1367) Mission. Referred to on 1 December 2020. https://www.finlex.fi/fi/laki/ ajantasa/2009/20090558.

Universities of Applied Sciences Act. 2014. Section 4. (28.12.2018/1368). Mission. Referred to on 1 December 2020. https://www.finlex.fi/fi/ laki/ajantasa/2014/20140932.

Ursin, J. 2014a. Learning outcomes in Finnish higher education from the perspective of comprehensive curriculum framework. In H. Coates (ed.) Higher education learning outcomes assessment. International perspectives. Frankfurt am Main: Peter Lang, 159–173.

Ursin, J. 2014b. AHELO – Korkeakouluopiskelijoiden oppimistulosten kansainvälinen arviointi. Kieli, koulutus ja yhteiskunta. April/2014. Web publication.

Utriainen, J., Marttunen, M., Kallio, E. & Tynjälä, P. 2017. University applicants' critical thinking skills: The case of the Finnish Educational Sciences. Scandinavian Journal of Educational Research 61 (6), 629–649.

Vettenranta, J., Hiltunen, J., Nissinen, K., Puhakka, E. & Rautopuro, J. 2016. Lapsuudesta eväät oppimiseen: neljännen luokan oppilaiden matematiikan ja luonnontieteiden osaaminen: kansainvälinen TIMSS-tutkimus Suomessa. University of Jyväskylä, Finnish Institute for Educational Research.

Vipunen – statistical service of the education administration. 2019. Higher education students. Data of the information service agreement between the Finnish National Agency for Education and Statistics Finland 2.8. https://vipunen.fi/fi-fi/_layouts/15/xlviewer. aspx?id=/fi-fi/Raportit/Korkeakoulutuksen%20opiskelijat_A1.xlsb

Virtanen, A. & Tynjälä, P. 2013. Kohti työelämätaitoja kehittävää yliopistopedagogiikkaa - opiskelijoiden näkökulma. Yliopistopedagogiikka 20 (2), 2–10. https://lehti.yliopistopedagogiikka.fi/2013/10/02/ kohti-tyoelamataitoja-kehittavaa-yliopistopedagogiikkaa-opiskelijoiden-nakokulma/

Virtanen, A. & Tynjälä, P. 2018. Factors explaining the learning of generic skills: a study of university students' experiences. Teaching in Higher Education 24 (7), 880–894. https://doi.org/10.1080/13562517.2018.1515195.

Virtanen, V., Postareff, L. & Hailikari, T. 2015. Millainen arviointi tukee elinikäistä oppimista? Yliopistopedagogiikka 22 (1), 3–11. https://lehti.yliopistopedagogiikka.fi/2015/03/27/ millainen-arviointi-tukee-elinikaista-oppimista/

Zahner, D. & Ciolfi, A. 2018. International comparison of a performance-based assessment in higher education. In O. Zlatkin-Troitschanskaia, M. Toepper, H. Anand Pant, C. Lautenbach & C. Kuhn (eds.) Assessment of Learning Outcomes in Higher Education. Cross-National Comparisons and Perspectives. Wiesbaden, Germany: Springer, 215–244.

Zahner, D. 2013. Standard-Setting Study Final Report: CLA+. Unpublished.

Zlatkin-Troitschanskaia, O., Shavelson, R., Schmidt, S. & Beck, K. 2019. On the complementarity of holistic and analytic approaches to performance assessment scoring. British Journal of Educational Psychology 89, 468–484.

Zlatkin-Troitschanskaia, O., Toepper, M., Molerov, D., Buske, R., Brückner, S., Pant, H., Hofmann, S. & Hansen-Schirra, S. 2018. Adapting and Validating the Collegiate Learning Assessment to Measure Generic Academic Skills of Students in Germany: Implications for International Assessment Studies in Higher Education. In O. Zlatkin-Troitschanskaia, M. Toepper, H. Anand Pant, C. Lautenbach & C. Kuhn (eds.) Assessment of Learning Outcomes in Higher Education. Cross-National Comparisons and Perspectives. Wiesbaden, Germany: Springer, 245–266

Zoller, U., & Tsaparlis, G. 1997. Higher and lower-order cognitive skills: The case of chemistry. Research in Science Education 27, 117–130.

OPETUS- JA KULTTUURIMINISTERIÖ
UNDERVISNINGS- OCH KULTURMINISTERIET
MINISTRY OF EDUCATION AND CULTURE

ISBN 978-952-263-901-1 (PDF)
ISSN 1799-0351 (PDF)