KANSALLINEN
KOULUTUKSEN
ARVIOINTIKESKUS

MATEMATIIKKAA COVID-19-PANDEMIAN VARJOSSA II

Menetelmälliset ratkaisut matematiikan 9. luokan arvioinnissa keväällä 2021

Jari Metsämuuronen | Saara Nousiainen (Toim.)

JULKAISUT 5:2023

MATEMATIIKKAA COVID-19-PANDEMIAN VARJOSSA II

Menetelmälliset ratkaisut matematiikan 9. luokan arvioinnissa keväällä 2021

Jari Metsämuuronen
Saara Nousiainen
(Toim.)

KARVI

Kansallinen koulutuksen arviointikeskus
Julkaisut 5:2023

JULKAISIJA Kansallinen koulutuksen arviointikeskus

KANSI JA ULKOASU Juha Juvonen (org.) & Ahoy, Jussi Aho (edit)
TAITTO PunaMusta

ISBN 978-952-206-787-6 pdf
ISSN 2342-4184 (verkkojulkaisu)

Tiivistelmä

Julkaisija
Kansallinen koulutuksen arviointikeskus (KARVI)

Julkaisun nimi
MATEMATIIKKAA COVID-19-PANDEMIAN VARJOSSA II
Menetelmälliset ratkaisut matematiikan 9. luokan arvioinnissa keväällä 2021

Tekijät
Jari Metsämuuronen & Saara Nousiainen (Toim.)

Kansallinen 9. luokan matematiikan oppimistulosarviointi toteutettiin COVID-19-pandemian varjossa vuoden 2021 keväällä; osaamisen arviointi tehtiin noin vuosi sen jälkeen, kun Suomessa oli siirrytty pandemian tartuttamiseksi pääosin etäopiskeluun. Keväällä 2020 lähes kaikki oppilaat olivat etäopetuksessa ja syksyn 2020 ja kevään 2021 aikana oppilaat olivat vaihtelevia jaksoja alue- ja koulukohtaisesti osin etäopetuksessa ja osin lähiopetuksessa. Aineistosta on jo aiemmin raportoitu keskeiset tasa-arvoon liittyvät tulokset. Tämä raportti kuvaa mittauksen teknisiä ratkaisuja ja tulosten luotettavuutta. Raportissa julkaistaan yksi käytetyistä tehtäväsarjoista osioparametreineen. Tätä osioparametritietoa voivat käyttää halutessaan muut toimijat omien mittareidensa kehittämisessä ja vaikeustasojen kalibroinnissa. Tehtäväsarjan yhteydessä julkaistaan myös karkea arvosanaehdotus niitä opettajia varten, jotka haluavat käyttää testiä tai sen osioita omien oppilaidensa tason testaamiseen.

Samanaikaisesti erillisenä julkaisuna julkaistaan vuoden 2014 kansallisen opetussuunnitelman perusteiden perusteella valmisteltu ns. matematiikan oppimistulosarviointien viitekehys, jossa matematiikan sisältöalueet ja tavoitteet on purettu helpottamaan tulevien matematiikan arviointien valmistelua. Myös opettajat voivat hyödyntää viitekehystä tehtävien ja kokeiden laadinnassa.

Vuoden 2021 matematiikan arvioinnissa kehitetyt mittarit ovat sekä *osuvia* eli valideja kuvaamaan matematiikan osaamista, asenteita, emootioita ja kiusaamisen kokemuksen intensiteettiä että riittävän *tarkkoja* eli reliaabeleja erottelemaan oppilaat toisistaan ja antamaan siis tarkkoja arvioita matematiikan osaamiseen liittyvistä ilmiöstä. Aiempiin matematiikan mittauksiin nähden mittaristo on laajempi ja perusteellisempi. Kun siis osaamista, asenteita, emootioita ja kiusaamista kuvataan raporteissa, tulokset ovat sekä kattavampia että osaamista monipuolisemmin kuvaavia kuin ennen. Samoin otoskoko on oleellisesti suurempi kuin aiemmissa mittauksissa—noin kolminkertainen ensimmäisiin mittauksiin nähden—mikä mahdollistaa sen, että aineistossa on riittävästi myös erittäin heikosti ja erittäin hyvin menestyviä oppilaita sekä maahanmuuttotaustaisia oppilaita, jotta näitä ryhmiä voidaan myös analysoida aiempia arviointeja tarkemmin.

Tulosten luotettavuutta haastaa kaksi seikkaa: yhtäältä mittauksen digitaalinen aineiston keruu ja tähän liittyvä aiempaa suurempi määrä puuttuvia tietoja ja toisaalta COVID-19-pandemian aiheuttamat haasteet. Näitä on kuitenkin pyritty paikkaamaan aineiston korjaamisella. Normaalistikin osa oppilasvastauksista poistetaan niihin sisältyvän suuren epävarmuuden vuoksi. Nyt myös osa aiempiin arviointeihin yhdistyneistä linkkitehtävistä poistettiin vertailukelpoisemman tuloksen saamiseksi aiempiin mittauksiin nähden, koska vastaustekniikka oli oleellisesti muuttunut digitaalisessa testauksessa verrattuna aiempiin paperi-kynä-mittauksiin. Tämä koski erityisesti tehtäviä, joissa oppilaan piti perustella suorittamansa laskutehtävä. Kokonaan paperilla tehtyihin arviointeihin verrattuna huomattavasti suurempi määrä oppilaita ei perustellut vastaustaan.

Aiemmasta analyysista tiedetään, että kansallinen matematiikan osaamisen taso on ollut laskujohteinen ainakin vuodesta 2000 lähtien. Lisäksi vuoden 2021 arvioinnissa havaittiin, että osaaminen on eriytynyt selvemmin kuin aiemmissa arvioinneissa siten, että se muodostuu nyt kolmesta osaamisryhmästä: heikosti menestyneistä, keskiosaajista ja erittäin hyvin osaavista oppilaista. Tarkentavissa analyyseissa havaittiin, että teknisissä ja puhtaissa laskutoimituksissa ei ollut huomattavia eroja heikompien ja parempien oppilaiden osaamisessa aiempaan verrattuna. Keskeinen ero heikompien ja parempien oppilaiden välillä esiintyy erityisesti sanallissa tehtävissä. Varsinkin arvosanoja 5–7 saavat oppilaat osaavat tehtäviä aiempaa heikommin, jos niihin sisältyy sanallinen ongelma. Epäselväksi tosin jää, onko heikomman suoriutumiseen syynä oppilaiden puutteellinen keskittymiskyky, vai onko ongelmana ehkä näiden tehtävien ymmärtäminen. Näyttää myös siltä, että osassa tehtävistä osaamisen lasku johtuu digitaalisesta testialustasta, sillä oppilaille näytti olevan taitotasosta riippumatta haasteellista tai vaivalloista esimerkiksi kirjoittaa laskutoimituksia matematiikkaeditorilla. Aiemmissa paperi-kynä-testauksissa huomattavasti useampi oppilas perusteli vastauksensa tämän kaltaisissa tehtävissä.

Heikosti menestyneiden oppilaiden lisämallinnusten perusteella Karvi suosittaa, että osaamisen kehittymistä on seurattava jo alemmilla vuosiluokilla, ja aiempaa intensiivisempää tukea on tarjottava heti tuen tarpeen ilmetessä mahdollisimman varhaisessa vaiheessa, jotta voidaan estää liian suuri taidoissa jälkeen jääminen tulevien kouluvuosien aikana. On järkevää suunnata aiempaa intensiivisempää tukea alkuluokkien jälkeiseen aikaan riittävän varhain ja riittävän tehokkaasti, ennen kuin osaamisessa ilmeneviä puutteita on vaikea korjata ylempien luokkien aikana. Oppilas, jonka matemaattiset taidot jäävät puutteellisiksi jo vuosiluokilla 3–6, vaatii huomattavan paljon tukea yläkoulun aikana, jotta hän pystyisi kirimään jo muodostunutta eroa kiinni.

Näyttää myös siltä, että oppilaiden osaaminen erityisesti matematiikan erityissanastossa, ongelmanratkaisutehtävissä ja sanallistetuissa tehtävissä on oleellisesti heikentynyt vuosien varrella. On siis järkevää kiinnittää huomiota yhtäältä oppilaiden sanavaraston ja luku- ja kirjoitustaitojen kehittymiseen ja toisaalta taitoon työskennellä pitkäjänteisesti ongelmaratkaisutehtävien parissa. Näiden taitojen kehittäminen jo alakoulussa helpottaisi monen oppilaan kykyä ratkaista tehtäviä, joissa pitää soveltaa aiemmin opittuja tietoja. Osalla heikosti menestyvistä oppilaista saattaa myös olla numerotaidottomuutta eli dyskalkuliaa, jonka selvittämiseksi testaaminen olisi perusteltua, jotta heille voidaan tarjota riittäviä ja osuvia tukitoimia.

Hakusanat: Matematiikka, Oppimistulokset, Osaamisen heikentyminen, Digitaalinen testaus, Suhtautuminen matematiikkaan, Tunteet matematiikan yhteydessä

Sammanfattning

Utgiven av
Nationella centret för utbildningsutvärdering (NCU)

Publikationens namn
MATEMATIKEN I SKUGGAN AV COVID-19-PANDEMIN II
Metodlösningar i utvärderingen av lärresultaten i matematik i årskurs 9 våren 2021

Författare
Jari Metsämuuronen & Saara Nousiainen (Red.)

Den nationella utvärderingen av lärresultaten i matematik i årskurs 9 genomfördes under våren 2021 i skuggan av COVID-19-pandemin. Utvärderingen av kunskaperna genomfördes cirka ett år efter att man i Finland hade övergått till distansstudier för att stoppa pandemin. Våren 2020 fick alla elever distansundervisning. Under hösten 2020 och våren 2021 fick eleverna under varierande perioder endera distansundervisning eller närundervisning, beroende på region och skola. I rapporten beskrivs först de tekniska lösningarna i utvärderingen och resultatens tillförlitlighet. Dessutom publiceras en av de använda uppgiftsserierna och dess delparametrar. Denna information om uppgifterna kan användas av andra nationella aktörer som vill utveckla sina egna mätare och kalibrera svårighetsnivåerna. För de lärare som vill använda utvärderingens prov eller delar av provet för att testa sina egna elever publiceras också ett förslag till ungefärligt vitsord.

Samtidigt publiceras också en referensram för matematik som en separat publikation. Referensramen har utarbetats med utgångspunkt i de nationella grunderna för läroplanen 2014. I referensramen har innehållsområdena och målen för matematiken plockats isär och slagits samman över årskursgränserna. Det här underlättar förberedelserna inför kommande utvärderingar av ämnet.

Mätarna som utvecklades för utvärderingen i matematik år 2021 är å ena sidan tillräckligt riktade, dvs. valida för att beskriva matematikkunskaper, attityder, emotioner och intensiteten i upplevelsen av mobbning, och å andra sidan tillräckligt noggranna, dvs. reliabla, för att skilja eleverna från varandra och alltså ge noggranna bedömningar av fenomen med anknytning till matematikkunskaper. Jämfört med tidigare mätningar av matematiken är mätarna mer omfattande och grundligare. När såväl kunskaper, attityder, emotioner och mobbning beskrivs i rapporterna är resultaten både mer uttömmande och mer mångsidiga än tidigare. Samtidigt är samplet betydligt större än i tidigare mätningar—cirka tre gånger så stort som i de första utvärderingarna—vilket möjliggör att materialet också omfattar tillräckligt med elever med mycket svag respektive god framgång samt elever med invandrarbakgrund. Med hjälp av ett större sampel kan också analysen av dessa grupper göras noggrannare än i tidigare utvärderingar.

Resultatens tillförlitlighet har påverkats av två olika faktum: den digitala insamlingen och en större mängd saknad information än i tidigare utvärderingar kombinerat med de utmaningar som orsakades av COVID-19-pandemin. Strävan har ändå varit att åtgärda dessa brister genom att i viss mån manipulera materialet. Också normalt avlägsnas vissa elevsvar på grund av att de är förknippade med stor osäkerhet. En del av de uppgifter som återkom från tidigare mätningar avlägsnades för att resultatet skulle bli mer jämförbart i förhållande till tidigare utvärderingar. Orsaken var att svarstekniken har förändrats väsentligt jämfört med tidigare utvärderingar. Det här gällde särskilt uppgifter där eleverna förutsattes motivera svaret på räkneuppgiften. Jämfört med äldre utvärderingar var andelen elever som inte motiverat sitt svar avsevärt större.

I ljuset av tidigare analyser vet vi att nivån på matematikkunskaperna nationellt sett haft en nedåtgående trend åtminstone sedan år 2000. Dessutom observerades i utvärderingen från år 2021 att ytterligheterna var tydligare än tidigare. Eleverna kan nu delas in i tre grupper: svaga elever, medelmåttiga elever och mycket kunniga elever. I de noggrannare analyserna observerades att jämfört med tidigare förekom det i de tekniska och rena räkneoperationerna inte några betydande skillnader mellan de svagare och starkare elevernas kunskaper. Men särskilt i textuppgifter förekom en uppenbar skillnad mellan svagare och starkare elever. Då uppgiften innehåller ytterligare en dimension, nämligen att den ges som en textuppgift, är i synnerhet kunskaperna hos elever som har vitsorden 5–7 sämre än tidigare. Det förblir oklart om orsaken till att eleverna klarar sig sämre är deras bristfälliga koncentrationsförmåga eller om problemet är svårigheter att förstå dessa uppgifter. Det verkar också som om de sämre resultaten i en del av uppgifterna beror på den digitala testplattformen. Det verkade oberoende av kunskapsnivå vara utmanande eller besvärligt för eleverna att till exempel skriva räkneoperationer med matematikeditorn. I tidigare utvärderingar som gjordes med papper och penna var det synbart vanligare att många fler elever motiverade sitt svar i den här typen av uppgifter.

På basis av tilläggsmodelleringen för svagt presterande elever rekommenderar NCU att dessa elever får mer intensivt stöd än hittills i ett så tidigt skede som möjligt för att förhindra eventuell utslagning i matematik under de följande skolåren. Det vore klokt att tillräckligt tidigt och tillräckligt effektivt rikta ett mera intensivt allmänt stöd till de här eleverna efter de lägsta årskurserna – innan bristerna i kunskaperna blir så stora att de är svåra att åtgärda med undervisningslösningar i de högre årskurserna. En elev som ligger efter i de matematiska färdigheterna redan under årskurs 3–6 behöver mycket stöd i de högre klasserna för att komma i kapp de övrigas försprång.

Det verkar också som om elevernas kunskaper särskilt i matematikens terminologi, problemlösningsuppgifter och verbala uppgifter har försämrats väsentligt under årens lopp. Det finns alltså skäl att å ena sidan fästa uppmärksamhet vid utvecklingen av elevernas ordförråd och å andra sidan vid förmågan att arbeta långsiktigt med problemlösningsuppgifter. Om dessa färdigheter utvecklades redan i de lägre klasserna skulle det vara lättare för eleverna att lösa uppgifter där de måste tillämpa tidigare inlärda kunskaper. En del av de svagt presterande eleverna kan också ha svårigheter med att tolka siffror, dvs. dyskalkyli. För att utreda detta är det motiverat att testa eleverna för att de ska kunna erbjudas rätt stödåtgärder i tillräcklig omfattning.

Sökord: Matematik, lärresultat, försämrade kunskaper, digital testning, inställning till matematik, känslor som matematiken väcker

Summary

Publisher
Finnish Education Evaluation Centre (FINEEC)

Title of publication
MATHEMATICS IN THE SHADOW OF THE COVID-19 PANDEMIC II
Methodological solutions of the assessment of mathematics at the 9th grade in spring 2021

Authors
Jari Metsämuuronen & Saara Nousiainen (Eds.)

The national assessment of 9th grade math learning outcomes was implemented in the shadow of the COVID-19 pandemic in spring 2021; the assessment was done about a year after Finland had switched to distance learning solutions in order to tame the pandemic. In the spring of 2020, all students were in distance learning, and during the fall of 2020 and spring of 2021, the students were in varying periods depending on the region and school, partly in distance learning and partly in face-to-face teaching. This report describes, on the one hand, the technical solutions of the measurement and the trustworthiness (validity and reliability) of the assessment results. On the other hand, one of the used test booklets with item parameters is published. This item-wise data can be used by other national actors in developing their own measurement instruments and in calibrating difficulty levels if they wish. In connection with the test, a rough grade proposal is also published for those teachers who want to use the test or its tasks to test their own students.

As a side of the report, a separate framework for mathematics evaluation and assessment is published also. In the framework, the content areas and goals of mathematics have been broken down from the national core curriculum and they are combined across grade levels to facilitate the preparation of future mathematics assessments.

In the 2021 mathematics assessment, the developed measurement instruments are, on the one hand, valid to describe mathematics knowledge, attitudes, emotions, and the intensity of the experience of bullying. On the other hand, the instruments are sufficiently accurate, or reliable, to distinguish students from each other and therefore give accurate assessments of the phenomena related to mathematics knowledge. Compared to previous mathematics measurements, the set of booklets is wider and more thorough. Hence, when skills, attitudes, emotions, and bullying are described in reports, the results are more comprehensive than before. Similarly, the sample size is substantially larger than in previous measurements—about three times the size of the early measurements—which enables the dataset to include enough special populations such as very poorly and very well achieving students or students with an immigrant background. With a larger sample size, something more precise can be said about these special populations than in previous assessments.

The digital testing of mathematics skills and the associated higher number of missing data on the one hand and the challenges caused by the COVID-19 pandemic on the other have undoubtedly affected the accuracy of the results. However, efforts have been made to rectify these with a small manipulation of the material. Some of the linking items were removed in order to obtain a more comparable result compared to previous measurements, because the answer technique had changed significantly compared to previous measurements. This concerned tasks where the students had to express their own calculations. Compared to previous evaluations, significantly more students did not provide these calculations.

It is known from the previous analysis that the national level of mathematics competence has been declining since at least the year 2000. In addition, in the evaluation of 2021, it was observed that, at the extremes of competence, the competence has differentiated from each other more clearly than before; in the dataset, separate populations of lower and higher achieving students have clearly emerged. Initial analysis of the linking items indicates that there were no notable differences in the skills of the lower and higher achieving students in technical and pure calculation assignments when it comes to the previous assessments and the recent one. The key difference between the lower and higher achieving students seems to come especially in verbal tasks. When a verbal problem is added to the task as a new level, students in grades 5–7, especially, are less proficient at these. It is unclear whether the reason is the students' reluctance to focus on solving more difficult tasks or whether the problem is understanding these tasks. It also seems that the decline in competence in some of the tasks may be due to the tool used, because regardless of the skill level, it seemed to be challenging or difficult for the students, for example, to write calculations in the math editor used in the test.

Based on additional modeling of low-achieving students, FINEEC recommends that low-achieving students should be targeted with more intensive support as early as possible in order to prevent possible mathematical marginalization during the coming school years. A student who falls behind in learning math in grades 3–6 requires a significant amount of support during middle school to close the gap. It makes sense to direct enhanced general support to the period after the first grades early enough and effectively enough, before it is difficult to correct the deficiencies in skills with meaningful teaching solutions during the upper grades.

It also seems that the students' competence, especially in the special vocabulary of mathematics, problem-solving tasks, and verbalized tasks, has notably weakened over the years. It is therefore reasonable to pay attention to the development of students' vocabulary on the one hand and to the ability to work long-term on problem-solving tasks on the other. The development of these skills starting from elementary school would facilitate the ability of many students to solve tasks that require the application of already learned information. Some low-achieving students may also have dyscalculia. It makes sense to test low-achieving students for this at an early stage so that support measures can be targeted in the right way on the right thing.

Keywords: Mathematics, Learning outcomes, Deterioration of competence, Digital testing, Attitude towards mathematics, Emotions in connection with mathematics

Sisällys

1

Alkusanat

Jari Metsämuuronen, Karvi
Saara Nousiainen, Karvi

1

1.1 Johdattelua vuoden 2021 matematiikan arvioinnin teemoihin

Keväällä 2021 Kansallinen koulutuksen arviointikeskus (Karvi) järjesti matematiikan oppimistulosarvioinnin perusopetuksen 9. luokan oppilaille (tuonnempana ”arviointi”). Arviointi oli osa perusopetuslain (628/1998) 21§:n edellyttämää koulutuksen arviointia. Karvia koskevan lain (1295/2013) 2§:n mukaan perusopetuksen oppimistulosten arvioinnit ovat osa valtakunnallista koulutuksen arviointijärjestelmää. Oppimistulosten arviointien tarkoituksena on tuottaa luotettavaa tietoa esi- ja perusopetuksen opetussuunnitelmien perusteiden (tuonnempana POPS; OPH 2014) tavoitteiden saavuttamisesta, oppiaineen osaamisen tasosta sekä koulutuksen tasa-arvon toteutumisesta.

Arvioinnin alkuperäinen toteuttamisaika oli Kansallisen koulutuksen arviointisuunnitelman (Karvi, 2020) mukaan keväällä 2020, mutta COVID-19-pandemian vuoksi toteutusta siirrettiin vuodella. Keväällä 2021 osassa kouluja oppilaat olivat vielä etäopetuksessa, jolloin he myös osallistuivat arviointiin etäyhteydellä. Heidän osuutensa oli kuitenkin melko vähäinen (noin 6 % otoksen oppilaista). Lähi- ja etäopetuksessa arvioinnin suorittaneiden osaamiseroja käsitellään lyhyesti tässä raportissa, koska se on ilmeinen tulosten luotettavuutta koskeva seikka.

Arvioinnin keskeisiä tasa-arvotuloksia on käsitelty jo aiemmassa raportissa (Metsämuuronen & Nousiainen, 2021). Tähän käsillä olevaan raporttiin on koottu arvioinnin taustalla olleita teknisiä näkökulmia ja menetelmällisiä ratkaisuja. Myöhemmin julkaistavassa raportissa osaamisen erojen selittäjiä tarkennetaan yhtäältä oppilaiden ja toisaalta opettajien, oppimisympäristöjen ja huoltajien taustatekijöiden suhteen.

Tiedonkeruu oli monella tavalla aiemmista matematiikan arvioinneista poikkeava. Ensiksi itse arvioinnin tehtäväsarjojen rakenteessa ja toteutuksessa päädyttiin kahdeksaan aiemmista arvioinneista poikkeavaan ratkaisuun, joilla pyrittiin entistäkin laadukkaampaan ja laaja-alaisempaan tiedonkeruuseen. Näitä pohtivat luvussa 2 Metsämuuronen ja Nousiainen (2023). Toiseksi taustalla on vasta valmistunut matematiikan oppimistulosarvioinnin viitekehys (ks. Ukkola & Kivistö, 2023), jossa POPS:ssa esitetyt matematiikan opetuksen tavoitteet muunnettiin oppimisen tavoitteiksi eri vuosiluokilla. Kolmanneksi tehtäviä valmisteltaessa 2019 käytettävissä oli syksyllä 2021 käyttöön tulleiden uusien päättöarvioinnin kriteerien luonnos (OPH, 2020), jota hyödynnettiin tehtävien laadinnassa ja valinnassa.

Arviointi toteutettiin Turun yliopiston oppimisanalytiikan keskuksen digitaalisessa ViLLE-ympäristössä. Karvin asiantuntijat ja Turun yliopiston ViLLE-tiimi (Metsämuuronen, Hermonen, Nousiainen, & Laakso, 2023) tarkentavat teknisiin ratkaisuihin liittyviä seikkoja ja pohtivat samalla digitaalisen testauksen tulevaisuuden näkymiä luvussa 3. Luvussa 4 tarkastellaan osaamisen muutosta pitkittäisaineistojen avulla, ja mallitetaan, mikä erottaa toisistaan heikoimmin ja paremmin menestyneet oppilaat (Metsämuuronen & Suomilammi, 2023). Arvioinnin yhteydessä kehitettiin uusi, matematiikkaan oppiaineena ja matematiikan oppimiseen liittyviä tuntemuksia kartoittava mittaristo. Tämän uuden ”emootiot matematiikkaa kohtaan”-mittarin rakennetta ja luotettavuutta tarkastellaan luvussa 5 (Salonen, 2023).

Aiemmista arvioinneista poiketen käsillä oleva arviointi suunniteltiin niin, että yksi yhdeksästä käytössä olleista tehtäväsarjoista oli mahdollista julkaista (raportin Liite 1). Tämän ansiosta opettajat ja muut toimijat voivat halutessaan arvioida omaa arvosanan antamisen linjaansa suhteessa kansalliseen keskiarvoon. Mittariversion eri osioihin liittyvät ominaisuudet ja mittarin pistemääriin linkittyvät arvosanaehdotukset julkaistaan raportin liitteessä 2. Samalla avataan käytetyn arvosanaehdotuksen peruslogiikka liitteessä 3.

Osa raportin artikkeleista on käynyt läpi tieteellisen käytänteen mukaisen vertaisarviointiprosessin. Kaikki artikkelit kävivät läpi kriittisen vertaisarvion, mutta näistä erityisesti Metsämuuronen, Hermonen, Nousiainen, & Laakso (2023) ja Salonen (2023) arvioitiin perinteistä vertaisarviointiproseduuria käyttäen niin, että asiantuntija-arvioija ei tiedä kirjoittajaa eikä kirjoittaja arvioijaa.

Helsingissä ja Lappeenrannassa 1.3.2023

Jari Metsämuuronen
Saara Nousiainen

1.2 Lähteet

Karvi (2020). *Koulutuksen arviointisuunnitelma 2020–2023*. Kansallinen koulutuksen arviointikeskus. https://karvi.fi/app/uploads/2020/04/Koulutuksen_arviointisuunnitelma_2020-2023.pdf

Metsämuuronen, J., Hermonen, A., Nousiainen, S., & Laakso, M.-J. (2023). Digitaaliseen testaamiseen liittyviä erityiskysymyksiä kansallisen oppimistulosarvioinnin näkökannalta. Teoksessa J. Metsämuuronen & S. Nousiainen (toim.), *Matematiikkaa COVID-19-pandemian varjossa II. Menetelmälliset ratkaisut matematiikan 9. luokan arvioinnissa keväällä 2021* (ss. 83–126). Julkaisut 5:2023. Kansallinen koulutuksen arviointikeskus.

Metsämuuronen, J. & Nousiainen, S. (2021). *Matematiikkaa COVID-19-pandemian varjossa. Matematiikan osaaminen 9. luokan lopussa keväällä 2021*. Julkaisut 27:2021. Kansallinen koulutuksen arviointikeskus. https://karvi.fi/wp-content/uploads/2021/12/KARVI_2721.pdf

Metsämuuronen, J. & Nousiainen, S. (2023). Yleiset menetelmäratkaisut matematiikan oppimistulosten arvioinnissa vuonna 2021. Teoksessa J. Metsämuuronen & S. Nousiainen (toim.), *Matematiikkaa COVID-19-pandemian varjossa II. Menetelmälliset ratkaisut matematiikan 9. luokan arvioinnissa keväällä 2021* (ss. 21–82). Julkaisut 5:2023. Kansallinen koulutuksen arviointikeskus.

Metsämuuronen, J. & Suomilammi, M. (2023). Matematiikan osaamisen eriytyminen ja osaamisen heikentymistä selittäviä tekijöitä. Teoksessa J. Metsämuuronen & S. Nousiainen (toim.), *Matematiikkaa COVID-19-pandemian varjossa II. Menetelmälliset ratkaisut matematiikan 9. luokan arvioinnissa keväällä 2021* (ss. 127–172). Julkaisut 5:2023. Kansallinen koulutuksen arviointikeskus.

OPH (2014). *Perusopetuksen opetussuunnitelman perusteet 2014*. Määräykset ja ohjeet 2014:96. Opetushallitus. https://www.oph.fi/sites/default/files/documents/perusopetuksen_opetussuunnitelman_perusteet_2014.pdf

OPH (2020). *Perusopetuksen päättöarvioinnin kriteerit*. Opetushallituksen määräys OPH-5042-2020. Opetushallitus. https://www.oph.fi/sites/default/files/documents/Perusopetuksen%20p%C3%A4%C3%A4tt%C3%B6arvioinnin%20kriteerit%2031.12.2020.pdf

Salonen, R. V. (2023). Tunteiden mittaaminen matematiikan arvioinnissa—Tunnemittari, uskomukset ja kontrolli–arvo-teoria. Teoksessa J. Metsämuuronen & S. Nousiainen (toim.), *Matematiikkaa COVID-19-pandemian varjossa II. Menetelmälliset ratkaisut matematiikan 9. luokan arvioinnissa keväällä 2021* (ss. 173–189). Julkaisut 5:2023. Kansallinen koulutuksen arviointikeskus.

Ukkola, A. & Kivistö, A. (toim.) (2023). *Matematiikan viitekehys kansallisen koulutuksen arviointikeskuksen oppimistulosarviointiin vuosiluokilla 1–9*. Julkaisut 6:2023. Kansallinen koulutuksen arviointikeskus.

2

Yleiset menetelmäratkaisut matematiikan oppimistulosten arvioinnissa vuonna 2021

Jari Metsämuuronen
Saara Nousiainen

2

- Kahdeksas kansallinen 9. luokan matematiikan oppimistulosarviointi toteutettiin COVID-19-pandemian varjossa; osaamisen kartoitus tehtiin noin vuosi sen jälkeen, kun Suomessa oli siirrytty pandemian taltuttamiseksi alueittain ja kouluittain poikkeaviin etäopiskeluratkaisuihin.

- Tämä raportti kuvaa tiedonkeruun menetelmällisiä ratkaisuja.

- Tiedonkeruu oli laajempi kuin aiemmat päättövaiheen arvioinnit: otoskoko oli laajempi, mittaristot olivat monipuolisemmat ja matematiikan osaamista mitattiin laaja-alaisemmin kuin aiemmin.

- Osaamismittareiden validiteetti eli osuvuus ja reliabiliteetti eli tarkkuus ovat korkeita. Testit siis mittasivat oikeaa asiaa tarkasti.

- Osaamisen lisäksi mitattiin matematiikkaan liittyviä asenteita ja tunteita sekä kiusaamisen kokemusta.

2.1 Johdannoksi

2.1.1 Arvioinnin taustaa

Vuoden 2021 keväällä kerättiin matematiikan osaamiseen liittyvää tietoa 9. luokan oppilailta. Tiedonkeruu oli suunniteltu vuoden 2020 keväälle, mutta COVID-19-pandemian vuoksi tiedonkeruu siirtyi vuodella. Tulokset heijastelevat siis osittain COVID-19-pandemian vaikutuksia koulutukseen (ks. Metsämuuronen & Nousiainen, 2021; Metsämuuronen & Seppälä, 2021).

Tiedonkeruu oli monella tavalla aiemmista matematiikan arvioinneista poikkeava. Ensiksi taustalla oli vasta valmistunut matematiikan oppimistulosarvioinnin viitekehys (ks. Ukkola & Kivistö, 2023). Toiseksi käytettävissä oli syksyllä 2021 käyttöön tulleiden uusien päättöarvioinnin kriteerien valmistelutyön välituloksia ja myöhemmin julkaistu luonnos (OPH, 2020). Kolmanneksi päädyttiin kahdeksaan aiemmista arvioinneista poikkeavaan ratkaisuun itse arvioinnin tehtäväsarjojen rakenteessa ja toteutuksessa. (1) Kun aiemmin varsinaisia tehtäväsarjoja on ollut yksi, vuoden 2021 tiedonkeruussa käytettiin yhdeksää tehtäväsarjaa. Tämä mahdollisti sen, että (2) arviointi voitiin suorittaa vapaammin yhden viikon aikana, koska (3) samassa koulussa ja luokassa oli käytössä useita erilaisia tehtäväsarjoja. Yksi sarjoista oli käytössä kaikissa kouluissa ja tämä (4) yhteinen tehtäväsarja ja siihen liittyvät tehtävät suunniteltiin poikkeuksellisesti julkaistavaksi raportoinnin yhteydessä vapaaseen käyttöön (ks. Liite 1 raportin lopussa). Kuusi tehtäväsarjoista oli kohdennettu niin, että (5) niissä oli yhteisten tehtävien lisäksi painotettuna jokin matematiikan kuudesta sisältöalueesta; näitä tehtäväsarjoja nimitetään tuonnempana ”erikoisversioiksi” ja kaikissa kouluissa käytettyä tehtäväsarjaa ”yleisversioksi”. Arvioinnin ensimmäisen vaiheen lisäksi (6) osa opiskelijoista osallistui arvioinnin toiseen vaiheeseen, jossa kartoitettiin yhdellä versiolla, minkälaisia matemaattisiin oppimisvaikeuksiin tai numerotaidottomuuteen eli dyskalkuliaan liittyviä piirteitä heikosti menestyneillä oppilailla oli, ja toisella versiolla, mille tasolle matematiikan taidoiltaan parhaat osaajat ovat edenneet. Heikoimmin menestyneet oppilaat tekivät toiminnallisen laskutaidon diagnostisen standarditestin (FUNA, 2019; Räsänen ym., 2021) mahdollisen dyskalkulian arvioimiseksi, ja parhaimmin menestyneet tekivät ylöspäin eriytetyn, vaikean testin. Näihin liittyviä tuloksia raportoidaan tulevassa raportissa. Koko arviointi toteutettiin digitaalisesti, ja (7) kaikista tehtävistä noin 90 % arvioitiin automaattisesti. Ne loput 10 % tehtävistä, joissa oppilaiden piti perustella vastauksensa, olivat opettajien pisteittämiä. Lopuksi (8) opettajille annettiin oppilaiden vastausten perusteella kaksi arvosanaehdotusta, joita opettaja saattoi käyttää oppilasarvioinnin tukena. Näiden muodostamista kuvataan tarkemmin koko raportin liitteessä 2.

Vaikka nyt raportoitava matematiikan oppimistulosarviointi poikkeaa monella tavalla aiemmin tehdyistä, se toteutettiin pääpiirteissään oppimistulosarvioinneissa vakiintuneita käytänteitä noudattaen (ks. Metsämuuronen 2009a). Arvioinnin suunnittelua ja toteutusta on kuvattu aiemmin julkaistussa raportissa (Metsämuuronen & Nousiainen, 2021), mutta asiat kerrataan tässä kokonaisuuden hahmottamiseksi.

2.1.2 Arvioinnin suunnittelu

Hanke käynnistyi vuoden 2019 alussa hankesuunnitelman laadinnalla. Arviointia varten koottiin Karvin ulkopuolisista matematiikan opetuksen asiantuntijoista ja Karvin edustajista koostuva asiantuntijaryhmä, jonka tehtävä oli varmistaa, että arviointi kattaa perusopetuksen opetussuunnitelman matematiikan sisältöalueet ja tavoitteet mahdollisimman laajasti. Ulkopuolisina asiantuntijoina ryhmässä toimivat professori Markku Hannula (Helsingin yliopisto), MAOL ry:n toinen varapuheenjohtaja Tuula Havonen, professori Peter Hästö (Turun yliopisto), apulaisprofessori Mikko-Jussi Laakso (Turun yliopisto), neuropsykologian erikoispsykologi Pekka Räsänen (Turun yliopisto) ja kehitysjohtaja Kaisa Vähähyyppä (MAOL ry.). Näistä Hannula, Havonen ja Vähähyyppä olivat jo aiemmassa vaiheessa osallistuneet matematiikan viitekehyksen laatimiseen (ks. Ukkola & Kivistö, 2023).

Asiantuntijaryhmän lisäksi koottiin kokeneista suomen- ja ruotsinkielisistä opettajista tehtävänlaatijaryhmä, jonka tehtävänä oli tuottaa riittävästi arviontiin sopivia tehtäviä ja niiden arviointiohjeita matematiikan eri tavoitteista ja sisältöalueilta. Tehtävänlaatijoina toimivat asiantuntijaryhmän lisäksi Aleksi Hermonen, Jussi Juurikka, Vuokko Kangas, Jani Kiviharju ja Anna Pomoell. Arviointiin valittujen suomenkielisten tehtävien kääntämisestä ruotsin kielelle vastasi Tora Smeds.

Karvista matematiikan oppimistulosarvioinnin suunnitteluun ja toteutukseen osallistuivat johtava arviointiasiantuntija ja menetelmäasiantuntija Jari Metsämuuronen, toisena menetelmäasiantuntijana Jukka Marjanen, arviointiasiantuntijoina Mika Puukko (1/2019–4/2019), Anne Kivistö (5/2019—7/2019), joka toimi jo aiemmassa vaiheessa viitekehystyön parissa, Pia Koskinen (1/2020–12/2020) ja Saara Nousiainen (8/2019–12/2019, 1/2021 eteenpäin). Ruotsinkielisinä asiantuntijoina toimivat Chris Silverström ja Jan Hellgren. Matematiikan viitekehyksen laadinnassa, asiantuntijaryhmässä ja tehtävien laadinnan ryhmässä Karvista oli mukana myös arviointiasiantuntija Annette Ukkola.

Tämä matematiikan arviointi oli ensimmäinen kokonaisuudessaan digitaalisesti toteutettu 9. luokan matematiikan arviointi. Koska Karvin oma digitaalisen testauksen järjestelmä oli samanaikaisesti kahden muun arviointihankkeen käytössä, arvioinnin teknisestä toteutuksesta vastaamaan valittiin Turun yliopiston oppimisanalytiikan keskus (myöhemmin Oppimisanalytiikan tutkimusinstituutti; ”ViLLE-tiimi”). ViLLE-tiimin kanssa oli toteutettu myös osa vuoden 2015 arvioinnista, jolloin puolet oppilaista suoritti arviointitehtäviä digitaalisesti. ViLLE-tiimissä tehtävien digitalisoinnista vastasi Aleksi Hermonen, joka osallistui myös tehtävien laadintaryhmään. Arvioinnin edellyttämien muutosten tekemisestä itse ViLLE-järjestelmään, vastasi Teemu Rajala. Näitä teknisiä ratkaisuja kuvataan tarkemmin luvussa 3. Matematiikan päässälaskutehtävien äänityksen toteutti Ari Maijanen Jyväskylän yliopiston soveltavan kielentutkimuksen keskuksesta.

Tässä osuudessa kuvataan mittaristot (luku 2.2), vertaistamiseen ja muuttujamuunnoksiin liittyviä erityiskysymyksiä (2.3), aineisto koonti, katoanalyysi ja lopullisen aineiston keskeiset piirteet (2.4) ja käytetyt menetelmät ja erikoistermit (2.5).

2.2 Tehtäväsarjat ja niiden luotettavuus

2.2.1 Matematiikan viitekehys ja määrittelykehys (*specification grid*)

Ennen arvioinnin aloittamista Karvissa käynnistettiin viitekehystyö, jossa Perusopetuksen opetussuunnitelmien perusteisiin (POPS; OPH 2014) kirjatut äidinkielen ja kirjallisuuden ja matematiikan tavoitteet purettiin ja yhdenmukaistettiin vertailukelpoisiksi kaikille luokka-asteille. Äidinkielen ja kirjallisuuden osalta viitekehys julkaistiin jo aiemmin (Karvi, 2019) ja matematiikan viitekehys julkaistaan tämän raportin yhteydessä erillisenä julkaisuna (Ukkola & Kivistö, 2023). Viitekehys toimii perustana eri vuosiluokkien entistä vertailukelpoisemmalle arvioinnille: osaamisen kehittymistä voidaan seurata tarkemmin koulun alusta (ks. mm. Ukkola & Metsämuuronen, 2019) kolmannelle luokalle (Ukkola & Metsämuuronen, 2021, 2023) ja myöhemmin kuudennelle luokalle. Nyt käsillä oleva 9. luokan arviointi perustuu samoille käsitteille kuin alempien vuosiluokkien arviointi, vaikka sisältöalueittaisia eroja eri vuosiluokkien arvioinneissa tietenkin on.

POPS:ssa matematiikan sisältöalueita on vuosiluokille 7–9 määritelty kuusi (S1–S6) ja tavoitteita 20 (T1–T20). Sisältöalueet ovat *Ajattelun taidot ja menetelmät* (S1), *Luvut ja laskutoimitukset* (S2), *Algebra* (S3), *Funktiot* (S4), *Geometria* (S5) sekä *Tietojen käsittely ja tilastot sekä todennäköisyys* (S6). Karvin aiemmista matematiikan arvioinneista poiketen, tässä arvioinnissa kaikkia sisältöalueita varten valmisteltiin oma tehtäväversionsa, jossa oli 10–12 sisältöalueeseen liittyvää tehtävää. Eri versiot linkitettiin toisiinsa tehtävillä, jotka olivat samoja kaikissa tehtäväsarjoissa (ks. tarkemmin luku 2.3).

Matematiikan tavoitteet voidaan jakaa kolmeen ryhmään. Ensimmäisen ryhmän muodostavat matematiikan merkitykseen, arvoihin ja asenteisiin liittyvät tavoitteet: [Oppilas] *uskoo kykyihinsä ja haluaa oppia lisää* (T1) ja *Oma-aloitteisuus ja sosiaalisuus työskentelyssä* (T2). Näitä tavoitteita kartoitettiin taustakyselyillä ja asennemittaristolla. Toisen ryhmän muodostavat työskentelyn taitoihin liittyvät tavoitteet, kuten että [Oppilas] *havaitsee ja ymmärtää yhteyksiä oppimiensa asioiden välillä* (T3), *käyttää (relevantisti) matemaattisia käsitteitä ja merkintätapoja* (T4), ja *ratkaisee loogista ja luovaa ajattelua vaativia tehtäviä ja käyttää erilaisia ratkaisustrategioita* (T5). Näitä tavoitteita varten ei kehitetty erillisiä tehtäviä, vaan näitä arvioitiin välillisesti muiden tehtävien kautta.

Kolmas tavoiteryhmä koostuu matematiikan käsitteellisistä ja tiedonalakohtaisista tavoitteista, kuten että [Oppilas] *osaa tehdä matemaattisia päätelmiä, laskea päässä ja käyttää taitojaan eri konteksteissa* (T10), *osaa peruslaskutoimituksia rationaaliluvuilla* (T11), ja *ymmärtää reaaliluvun käsitteen* (T12). Nämä tavoitteet luokittuvat POPS:ssa sisältöalueille niin, että T10 ja T20 ovat sisältöalueen S1 tavoitteita, T10–T13 ovat sisältöalueen S2 tavoitteita, T14 ja T15 ovat sisältöalueiden S3 ja S4 tavoitteita, T16–T18 ovat sisältöalueen S5 tavoitteita ja T13 ja T19 ovat sisältöalueen S6 tavoitteita. Sisältöalueet ja tavoitteet on tiivistetty taulukkoon 1, johon palataan luvussa 2.2.4 mittareiden luotettavuustarkastelujen yhteydessä.

TAULUKKO 2.1. Sisältöalueet ja tavoitteet vuosiluokilla 7–9 (OPH, 2014)

sisältöalueet		tavoite ja sen yksinkertaistettu kuvaus ("Oppilas…")	
S1	Ajattelun taidot ja menetelmät	T10	laskee päässään, tekee päätelmiä
		T20	ajattelee ja ratkaisee algoritmisesti ja (myös) ohjelmoiden
S2	Luvut ja laskutoimitukset	T10	laskee päässään, tekee päätelmiä
		T11	peruslaskutoimituksia rationaaliluvuilla
		T12	ymmärtää reaaliluvun käsitteen
		T13	laskee prosenttiosuuden, prosenttiluvun osoittaman määrän, muutos- ja vertailuprosentin
S3	Algebra	T14	ratkaisee yhtälöitä
		T15	tulkitsee ja tuottaa funktion
S4	Funktiot	T14	ratkaisee yhtälöitä
		T15	tulkitsee ja tuottaa funktion
S5	Geometria	T16	ymmärtää geometristen käsitteiden yhteyksiä
		T17	hyödyntää suorakulmaiseen kolmioon ja ympyrään liittyviä ominaisuuksia
		T18	laskee pinta-aloja ja tilavuuksia
S6	Tietojen käsittely ja tilastot sekä todennäköisyys	T13	laskee prosenttiosuuden, prosenttiluvun osoittaman määrän, muutos- ja vertailuprosentin
		T19	määrittää tilastollisia tunnuslukuja ja laskee todennäköisyyksiä

Tehtäväsarjat suunniteltiin niin, että tehtäviä oli kattavasti kaikkien tavoitteiden T10–T20 osalta. Lopullisessa mittaristossa Tavoitteen T12 (reaalilukujen käsitteeseen tutustuminen) painoarvo oli kuitenkin niin pieni, ettei eri testiversioita ylittävää summaa muodostettu. Tehtävät kuitenkin laskettiin mukaan kaikista tehtävistä laskettuun kokonaissummaan.

Ennen kuin tehtäviä alettiin laatia, valmisteltiin ns. määrittelykehys (*Specification grid* tai *Table of Specification*). Määrittelykehyksessä kuvataan, kuinka paljon ja kuinka vaikeita tehtäviä lopullisissa tehtäväsarjoissa tarvittiin eri sisältöalueilta ja tavoitteilta. Lisäksi siinä kuvataan, mitä ajattelun tasoa Bloomin ja kollegoiden (1956) tai Andersonin ja Krathwohlin ja kollegoiden viitekehyksessä (2001) tehtävän oikea ratkaisu tyypillisesti edellyttää. Näitä seikkoja käsitellään luvussa 2.2.2. Lopullisten mittareiden luotettavuutta käsitellään luvussa 2.2.3.

Arvioinnissa päädyttiin ratkaisuun, jossa kutakin sisältöaluetta kohden oli erillinen tehtäväsarja, joka muodostui yhteisestä osuudesta ja kohdennetuista tehtävistä sekä yksi yleisversio, jota käytettiin kaikissa kouluissa. Usean toisiinsa linkittyneen tehtäväsarjan yhdistäminen tiedonkeruussa vastaa kansainvälisissä tiedonkeruissa käytettävää menettelyä. Tämä puolestaan edellyttää pistemäärien muuttamista samalle asteikolle. Tätä kuvataan tarkemmin luvussa 2.3.

2.2.2 Tehtävien laadinta ja esitestaus

Käsillä oleva matematiikan oppimistulosarviointi oli ensimmäinen, jossa vuonna 2014 käyttöön otettujen, POPS:ssa kuvattujen tavoitteiden toteutumista oli mielekästä arvioida 9. luokan oppilailla. Tästä syystä arviointi ja tehtäväsarjat suunniteltiin sellaisiksi, että niillä saadaan mahdollisimman laajasti tietoa osaamisesta ja siihen liittyvistä tekijöistä. Taustalla valmisteltu viitekehys, samanaikaisesti valmisteilla ollut päättöarvioinnin kriteerien uudistus ja kokonaan uutena tavoitteena POPS:aan tullut ohjelmointi ja algoritminen ajattelu johtivat siihen, että tehtäväsarjoja päätettiin valmistaa useita: yksi yleisversio, joka päätettiin julkaista, ja useita eri sisältöalueille painottuvia tehtäväsarjoja.

Vertailukelpoisuuden vuoksi huomioitiin Karvin aiemmissa tehtäväsarjoissa käytetyt periaatteet (ks. Metsämuuronen, 2009a): saada tietoa sekä oppilaiden päässälaskutaidoista että perustelu- ja ongelmanratkaisutaidoista, joita on vaikea mitata yksinomaan monivalintatehtävillä. Näin tehtäväsarjoihin suunniteltiin yhtäältä erilliset päässälaskuja mittaavat osat (osa A), toiseksi objektiiviset[1] osat (monivalinnat, yhdistämistehtävät ja lyhytvastaukset; osa B) ja kolmanneksi vaativampaa ongelmanratkaisutaitoa mittaavat osat (osa C). Tavoitteena oli rakentaa sellainen digitaalinen tehtävistö, jossa opettajan pisteitettäväksi jäävien tehtävien osuus oli mahdollisimman pieni. Lopullisissa tehtäväsarjoissa kaikki päässälaskut ja objektiiviset tehtävät ja osa ongelmanratkaisutehtävistä oli automaattisesti pisteitettyjä; vain perustelutehtävissä tarvittiin opettajan työpanosta.

Koska tehtäväsarjoja suunniteltiin useita, tällä oli ymmärrettävästi vaikutusta tarvittavien tehtävien määrään. Yleensä oppimistulosarviointien esitestauksessa on noin kaksinkertainen määrä tehtäviä lopulliseen arviointiin nähden (Metsämuuronen, 2009a). Siinä, missä tehtäviä laativa ryhmä saattoi aiemmin valmistella 100 osioita, nyt käsillä olevaan arviointiin se laati 262 uutta tehtävää kattaen laajasti POPS:n sisältöalueet ja tavoitteet sekä vaikeustasot ja ajattelun tasot.

Ennalta valmistellun määrittelykehyksen perusteella tiedettiin, kuinka monta eri vaikeustason ja ajattelun tason tehtävää tarvittiin kullekin tavoitteelle ja sisältöalueelle. Kunkin **sisältöalueen ja tavoitteen painoarvon** yksikäsitteinen määrittely on hankalaa, sillä toisin kuin monissa muissa maissa, Suomessa ei POPS:n yhteydessä julkaista ns. opetusohjelmaa (*syllabus*) eli opetussuunnitelman laatijoiden käsitystä siitä, kuinka kauan ja miten kutakin yksittäistä asiaa tulee opettaa. Sen sijaan Suomessa luotetaan opettajaan, joka myös vastaa eri opetussisällöissä etenemisestä. Käytännössä etenemiseen vaikuttaa, kuinka paljon käytössä olevissa oppikirjasarjoissa käsitellään kutakin asiaa. Eri sisältöalueiden ja tavoitteiden painoarvot määriteltiin tässä arvioinnissa karkeasti siten, että tutkittiin yleisimmin Suomessa käytössä olleiden yläluokkien opetukseen tarkoitettujen

1 Osaamismittareiden tehtävät jaetaan perinteisesti objektiivisiin ja subjektiivisiin sen perusteella, tarvitaanko pisteityksessä subjektiivinen arvio ratkaisun onnistumisesta vai ei. Kansallisissa arvioinneissa periaatteena on ollut, että ainakin puolet lopullisesta pistemäärästä tulisi tulla suoraan oppilaan vastauksista ilman opettajan tai muun sensorin subjektiivista arviota (Metsämuuronen, 2009a). Karvin tehtäväsarjoissa opettajan pisteittämät tehtävät ovat tyypillisesti olleet "subjektiivisia" ja esimerkiksi monivalinta-, yhdistämis- ja lyhytvastaustehtävät "objektiivisia". Termiä "ongelmanratkaisutehtävä" käytetään tässä historiallisista syistä. Tietenkin kaikkiin tehtäviin liittyy "ongelman" ratkaisu, mutta aiemmissa matematiikan arvioinneissa "ongelmanratkaisutehtävä" terminä on varattu vaativammille, tuottamistyyppisille tehtäville. "Tuottamistehtävät" (*productive items*) voisi olla myös vaihtoehtoinen nimi. Digitaalinen testaus on muuttanut tätä logiikkaa hieman, sillä aiemmin tuottamistehtäviä ei kyetty arvioimaan objektiivisesti, mutta nykyiset järjestelmät mahdollistavat tämän ainakin osittain.

oppikirjojen sivumäärät eri sisältöalueilla ja tavoitteissa.[2] Tavoitteita T10 ja T20 (päässälaskut ja algoritminen ajattelu) ei käsitelty oppikirjoissa, mutta myös näiltä alueita sisällytettiin lopullisiin tehtäväsarjoihin. Näin päädyttiin taulukon 2 mukaisiin suhteellisiin osuuksiin. Huomataan, että päässälaskut sijoittuvat kahdelle sisältöalueelle (S1 ja S2) ja tavoitteet T13, T14 ja T15 luokittuvat useampaan sisältöalueeseen. Tämä johtuu siitä, että yksittäinen tehtävä voi olla ratkaistavissa usealla tavalla ja siksi siinä tarvittava osaaminen saattoi luokittua usealle eri sisältöalueelle ja tavoitteelle. Tällöin tehtävä myös lasketaan mukaan useampaan summaan.

TAULUKKO 2.2. Suuntaa antavat alustavat laskennalliset painotukset eri sisältöalueilla ja tavoitteilla

sisältöalueet		tavoite ja sen yksinkertaistettu kuvaus		tavoitteen suhteellinen painoarvo (%)	sisältöalueen suhteellinen painoarvo (%)
S1	Ajattelun taidot ja menetelmät	T10	päässälasku	8,3	17
		T20	algoritmit ja ohjelmointi	8,3	
S2	Luvut ja laskutoimitukset	T10	päässälasku	8,3	12
		T11	peruslaskutoimitukset	2,6	
		T12	reaaliluku	3,4	
		T13	prosenttilaskut	6,5	
S3	Algebra	T14	yhtälöt	7,2	9
		T15	funktiot	2,3	
S4	Funktiot	T14	yhtälöt	7,2	9
		T15	funktiot	2,3	
S5	Geometria	T16	geometriset käsitteet	11,0	32
		T17	suorakulmainen kolmio ja ympyrä	9,6	
		T18	pinta-ala ja tilavuus	11,8	
S6	Tietojen käsittely ja tilastot sekä todennäköisyys	T13	prosenttilaskut	6,5	19
		T19	tilastot ja todennäköisyys	13,0	

Ajattelun tasot valittiin Bloom–Anderson–Krathwohl-viitekehyksestä (Bloom, 1956; Anderson & Krathwohl, 2001), joka oli POPS:n ajattelun taustalla (ks. Ukkola & Kivistö, 2023). Viitekehyksen mukaisesti tehtävän suorittamisessa vaadittavat ajattelun tason jaettiin neljään luokkaan: muistaminen (H1), ymmärtäminen (H2), soveltaminen (H3) ja korkeammat taidot (H4), joista viimeinen sisältää analysoinnin, syntetisoinnin, arvioinnin ja luomisen taidot. Näiden ajattelun tasojen eroja erityisesti matematiikan tehtävien laadinnassa ja arvioimisessa ovat kuvanneet esimerkiksi Metsämuuronen (2009b, 2017b) sekä Metsämuuronen ja Räsänen (2018). Esimerkiksi tavoitteeseen T17 kuuluvan hypotenuusan osaamisessa *muistamista* edellyttävä tehtävä voisi liittyä hypotenuusan tunnistamiseen suorakulmaisesta kolmiosta; *ymmärtämistä* edellyttävä tehtävä voisi liittyä Pythagoraan lauseen käyttöön hypotenuusaa laskettaessa tilanteessa, jossa

2 Tarkempiakin keinoja olisi ollut käytettävissä, mutta päädyttiin karkeaan ja helppoon ratkaisuun. Oppikirjoissa on esimerkiksi ajankäyttösuosituksia kunkin aukeaman käsittelyyn. Niinpä esimerkiksi geometrian sisältöalueen painotus olisi voinut olla erilainen, sillä tältä sisältöalueelta *sivuja* saattaa tulla enemmän kuin joltain muulta alueelta, koska kuvia on paljon.

kaksi muuta sivua on annettu; *soveltamista* edellyttävässä tehtävässä hypotenuusan ominaisuuksia ja laskemista sovelletaan arkielämän tilanteessa kuten esimerkiksi seinää vasten nojaavien tikapuiden pituuden laskemissa; ja *korkeampaa ajattelua* vaativissa tehtävissä hypotenuusan tai suorakulmaisen kolmion ominaisuuksia käytetään luovasti jokin muun asian selvittämiseksi. Viimeksi mainittu tilanne syntyisi esimerkiksi silloin, kun useita laskukaavoja—mukaan lukien suorakulmaisen kolmion lauseketta—tulee soveltaa ongelman ratkaisemiseksi (ks. esimerkiksi luvussa 2.2.4 kuvion 2 tehtäväesimerkki).

Perinteisesti kansallisissa matematiikan arvioinneissa painopiste on ollut soveltavissa tehtävissä, ja selvästi pienempi osa tehtävistä on joko muistamista tai korkeampaa ajattelua mittaavia (Metsämuuronen, 2009a). Tässä arvioinnissa lopullisiin tehtäväsarjoihin suunniteltiin tulevan muistamista edellyttäviä tehtäviä noin 15 %, ymmärtämistä edellyttäviä tehtäviä noin 30 %, soveltamista edellyttäviä tehtäviä noin 40 % ja korkeampaa ajattelua edellyttäviä tehtäviä noin 15 %. Koska osa tehtävistä jouduttiin poistamaan lopullisesta mittaristosta, tämä ei täysin toteutunut.

Tehtäväsarjojen **vaikeustasot** määriteltiin yleisen käytännön mukaisesti kohderyhmälle keskivaikeaksi valitsemalla tehtäväsarjoihin sopivan vaikeita tehtäviä. Tämä maksimoi mittarin erottelukyvyn eli reliabiliteetin (ks. Lord, 1952; Metsämuuronen, 2009a). Toisaalta tehtäväsarjoihin valitaan perinteisesti tehtäviä siten, että mukana on helppoja, keskivaikeita ja vaikeita tehtäviä. Tämä antaa myös matematiikan taidoiltaan heikommille mahdollisuuden osoittaa osaamistaan, ja myös matematiikan taidoiltaan parhaat oppilaat saavat haastetta. Niinpä tehtäviä laadittaessa muokattiin kustakin tehtävästä helpompi ja vaikeampi versio. Esitestauksen perusteella selvitettiin tehtävän tarkempi vaikeustaso ja erottelukyky lopullista valintaa varten.

Tehtävätyyppien osalta tehtäväsarjoista suunniteltiin hieman aiemmista mittauksista poikkeavia. Tämä tarve tuli yhtäältä digitaalisena toteutetun arvioinnin luonteesta verrattuna paperi-kynämenetelmällä toteutettuun arviointiin, ja toisaalta halusta vähentää opettajille arvioinnista aiheutuvaa työmäärää. Niinpä monet sellaiset tehtävätyypit, joissa opettaja on aiemmin pisteittänyt oppilaiden vastauksia, pyrittiin muuttamaan sellaiseen muotoon, että niiden automaattinen pisteitys olisi mahdollista. Tämä taas johti siihen, että lopullisissa tehtäväsarjoissa objektiivisten osioiden suhteellinen osuus nousi korkeammaksi (66 %) kuin aiemmissa arvioinneissa. Lopullisissa tehtäväsarjoissa opettajan pisteittämiä, subjektiivisia osioita oli 9 % ja kaikkiaan ongelmanratkaisutehtäviä 22 %. Myös päässälaskujen osuus oli pienempi kuin aiemmissa arvioinneissa (13 %). Huomataan kuitenkin, että tehtävien pistemäärät olivat korkeampia ongelmaratkaisutehtävissä. Lopullisiin tehtäväsarjoihin valittujen 127 osion pistemäärästä 8 % muodostui päässälaskuista (osa A), 60 % monivalinta- ja yhdistämistehtävistä (osa B) ja 32 % ongelmanratkaisutehtävistä (osa C).

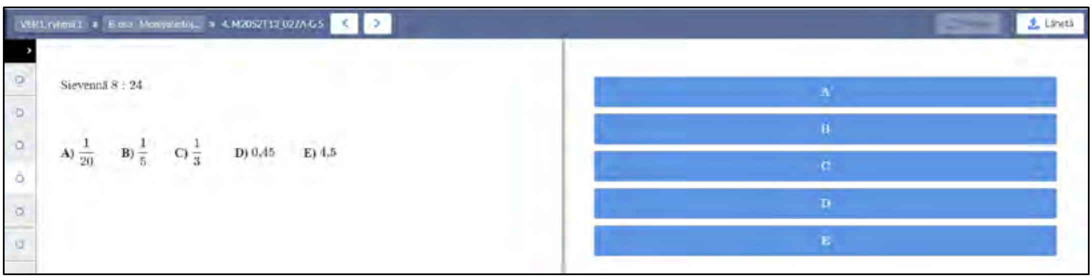
Tehtävien laadinta oli systemaattista: tiedettiin, kuinka monta tehtävää kaikkiaan tarvitaan kultakin sisältöalueelta ja tavoitteelta, mikä on tehtävissä edellytettävä ajattelun taso ja tehtävän vaikeustaso. Esitestausvaihetta varten muodostettiin tehtäviä niin runsaasti, että niitä oli riittävästi valittavissa myös lopullista arviointia varten. Osa tehtävistä suunniteltiin niin, että niiden avulla saatiin tietoa 3. luokan ja 6. luokan oppisisällöistä, ja vastaavasti osa tehtävistä valittiin lyhyen matematiikan oppimäärän ylioppilaskoetehtävistä tuomaan haastetta matematiikan taidoiltaan erittäin hyville oppilaille.

Kaikkiaan 344 osiota esitestattiin syksyllä 2019. Tehtävistä 262 oli uusia tehtäviä, 39 oli aiempiin arviointeihin linkittyviä tehtäviä ja 43 vanhoja ylioppilastehtäviä. Esitestaukseen osallistui 9. luokan oppilaita ja lukion ensimmäisen vuosikurssin opiskelijoita kaikkiaan 642. Esitestauksessa tehtävät oli jaettu kahdeksaan eri tehtäväversioon, jotta ne saatiin pidettyä riittävän lyhyinä ja jotta testin pituus ei vaikuttaisi suuresti osioiden vaikeustasoihin. Esitestatuista tehtävistä asiantuntijaryhmä valitsi lopulliset arviointitehtävät tehtävien vaikeustason ja erottelukyvyn perusteella. Valitsemalla sopivasti vaikeustasoltaan erilaisia tehtäviä, päästään koko mittarissa haluttuun vaikeustasoon, ja erottelukyvyltään parhaimpien tehtävien valitseminen maksimoi koko mittarin erottelukyvyn eli reliabiliteetin. Lopullisten tehtäväsarjojen ominaisuuksia käsitellään tarkemmin luvussa 2.2.4.

2.2.3 Esimerkkejä erityyppisistä tehtävistä

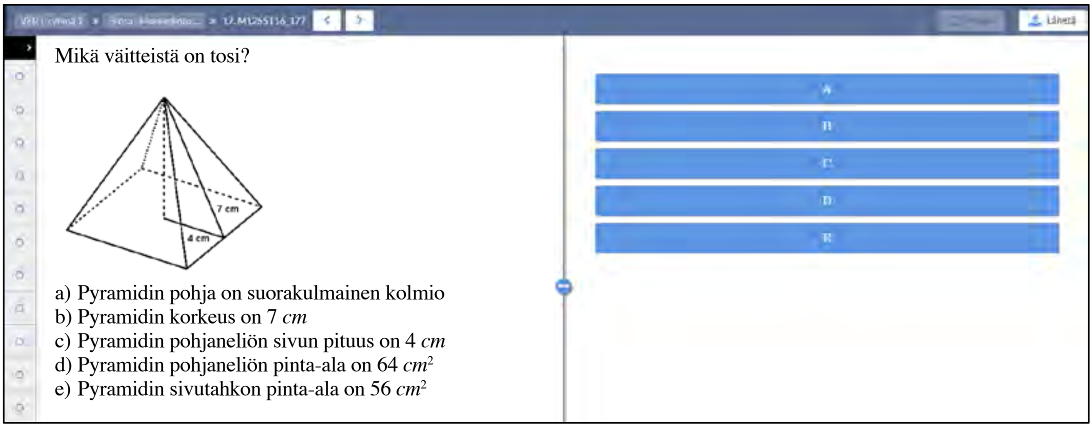
Yksi tehtäväsarjoista osiotietoineen julkaistaan koko raportin liitteenä. Luvussa 5 Metsämuuronen ja Suomilammi (2023) esittelevät yksittäisten osioiden avulla, kuinka osaamisen taso on muuttunut vuosien varrella. Tässä yhteydessä näytetään esimerkinomaisesti, millaisia tehtäviä lopullisiin tehtäväsarjoihin valittiin.

Oppilaiden osaamisskaalan selvittämiseksi arviointitehtäviksi valikoitiin erittäin helppoja, melko helppoja, keskivaikeita, melko vaikeita ja erittäin vaikeita tehtäviä. Kuviossa 1 on esimerkki arvioinnissa olleesta helposta monivalintatehtävästä, jossa oppilaan tuli selvittää tai päätellä, mitä tulee vastaukseksi, kun lausekkeen 8:24 sieventää. Tehtävä luokittui ensisijaisesti sisältöalueelle S2 (luvut ja laskutoimitukset) ja tavoitteille T10 (laskee päässään ja tekee päätelmiä) ja T11 (suorittaa peruslaskutoimituksia rationaaliluvuilla). Ajattelun tasoilla tämä tehtävä sijoittuu tasolle H2 eli ”ymmärtäminen”. Tälle perusteluna on se, että tehtävän oikein suorittaminen ei perustu ensisijaisesti käsitteiden tai asian muistamiseen (H1) eikä tehtävään liity suoraan varsinaista käytännöllistä soveltavaa elementtiä (H3). Tehtävän oikein suorittaminen ei myöskään edellyttänyt ensisijaisesti korkeampaa ajattelua (H4). Lopullisessa arvioinnissa tehtävän ratkaisi oikein 73 % oppilaista.



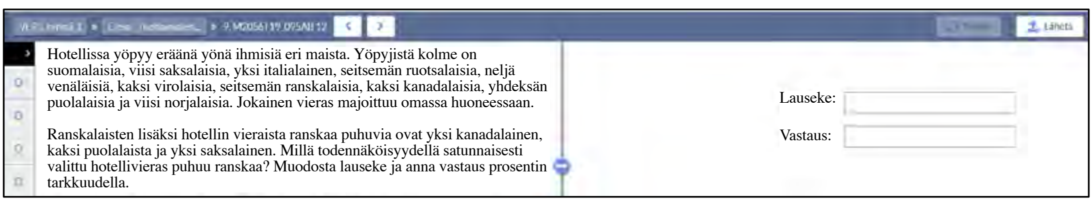
KUVIO 2.1. Esimerkki helposta laskutehtävästä

Kuviossa 2 on esimerkki keskivaikeasta monivalintatehtävästä, jossa oppilas arvioi annettujen väitteiden sopivuutta kuvassa näkyvään pyramidiin. Kyseessä on ensisijaisesti sisältöalueelle S5 (Geometria) ja tavoitteille T16 (ymmärtää geometristen käsitteiden yhteyksiä), T17 (hyödyntää suorakulmaiseen kolmioon ja ympyrään liittyviä ominaisuuksia) ja T18 (laskee pinta-aloja ja tilavuuksia) sijoittuva tehtävä. Ajattelun tasoista tehtävä luokittuu tasolle H4 (”korkeammat taidot”), johon sisältyy muun muassa kyky yhdistää useista yksittäisistä faktoista uusia tietoja sekä arvioida ja tehdä päätelmiä näiden faktojen pohjalta. Lopullisessa arvioinnissa 60 % oppilaista ratkaisi tehtävän oikein.



KUVIO 2.2. Esimerkki keskivaikeasta geometrian tehtävästä

Kuviossa 3 on korkeimmalle vaikeustasolle suunniteltu tuottamistehtävä, joka vaatii oppilaalta ongelmanratkaisutaitoa ja taitojen soveltamista todennäköisyyksien laskemisessa. Tehtävä sijoittuu sisältöalueelle S6 (Tietojen käsittely ja tilastot sekä todennäköisyys), tavoitteelle T19 (määrittää tilastollisia tunnuslukuja ja laskee todennäköisyyksiä) ja ajattelun tasolle H3 (soveltaminen). Osittain tehtävä olisi voinut sijoittua myös ajattelun tasolle H4 (korkeammat taidot). Taso H3 on perusteltu tehtävän käytännöllisen luonteen vuoksi: kykeneekö oppilas soveltamaan todennäköisyyden laskemista kompleksisessa ”arkielämän” tilanteessa—vaikka kukaan tuskin pohtisi juuri kyseistä asiaa omassa arkielämässään. Lopullisessa arvioinnissa tehtävän ratkaisi täysin oikein 3 % oppilaista.



KUVIO 2.3. Esimerkki vaikeasta todennäköisyyslaskusta

2.2.4 Lopulliset tehtäväsarjat ja niiden luotettavuus

Lopulliset tehtäväsarjat

Arvioinnissa käytetyt tehtävät laadittiin pääosin perusopetuksen opetussuunnitelman perusteissa vuosiluokille 7–9 kuvattujen matematiikan tavoitteiden ja sisältöalueiden pohjalta. Lisäksi mukaan valittiin tehtäviä myös vuosiluokkien 3–6 sisällöistä ja sisältöjen puolesta soveltuvista julkaistuista ylioppilaskoetehtävistä. Mittarin laadinnan taustalla oli yhtäältä Karvissa laadittu matematiikan viitekehys, jossa opetussuunnitelman eri vuosiluokkien matematiikan opetusta koskevat tavoitelauseet oli muokattu yhtenäiseksi dokumentiksi ja toisaalta uusi päättöarvioinnin kriteerien luonnos, joka otettiin myöhemmin käyttöön.

Arvioinnin ensimmäisessä vaiheessa oppilaat saivat vastattavakseen yhden seitsemästä toisiinsa linkitetystä tehtäväsarjasta (ks. otannan periaatteista luvussa 2.4). Yksi tehtäväsarjoista (Ver1) oli yhteinen kaikissa kouluissa ja tämä tehtäväsarja suunniteltiin julkaistavaksi hankkeen myöhemmässä vaiheessa. Kuusi muuta, sisältöaluepainotettua tehtäväsarjaa (VerS1–VerS6) jaettiin kouluissa niin, että kaikissa kouluissa oli kaksi erikoisversiota yleisversion lisäksi. Sisältöalueittain painotetuissa tehtäväsarjoissa tehtävistä noin 80 % oli identtisiä sisältäen tehtäviä kaikilta sisältöalueilta, joiden lisäksi 20 % oli tietyn sisältöalueen osaamista mittaavia tehtäviä. Yleisversion tehtävistä 72 % oli käytössä myös erikoisversioissa niin , että valtaosa näistä oli yhteisiä kaikille versioille ja jokaisessa erikoisversiossa oli lisäksi 1–2 sisältöaluekohtaista linkkitehtävää yleisversioon. Tehtäväsarjojen osioiden määrät ja osuudet on koottu taulukkoon 3. Arvioinnin ensimmäisen vaiheen suoriutumisen perusteella osa oppilaista valittiin toiseen vaiheeseen diagnostisiin testeihin, jotka eriytettiin ensimmäisessä vaiheessa heikoimmin ja parhaimmin menestyneille oppilaille.

Tarkkaavainen lukija huomaa taulukosta 3 esimerkiksi yleistehtäväsarjan (Ver1) osalta, että vaikka tehtäväsarjassa oli 60 yksittäistä osiota, eri sisältöalueiden tehtävistä tulee yhteensä 75 osiota ja eri tavoitteiden tehtävistä 89 osiota. Tämä johtuu siitä, että monet tehtävistä voitiin luokitella usealle sisältöalueelle ja vieläkin useammalle tavoitteelle. Toiseksi huomataan, että suunniteltu osioiden suhteellinen painoarvo poikkeaa selvästi lopullisesta. Myös tämä johtuu samojen osioiden käyttämisestä useilla eri sisältöalueilla ja tavoitteissa; vaikka tehtävä saattoi ensisijaisesti kuulua jollekin sisältöalueelle ja tavoitteeseen, sen toissijainen sijoittuminen sai aikaan sen, että suunniteltu suhteellinen osuus ei vastannut tarkasti toteutunutta osuutta. Pääsääntöisesti kuitenkin ne sisältöalueet ja tavoitteet, jotka painottuvat opetuksessa, painottuvat myös tehtäväsarjoissa kokonaisuutena. Kolmanneksi huomataan, että tavoitteen T12 painoarvo on hyvin pieni. Tästä syystä arviointitehtäviä oli liian vähän, jotta sen osalta olisi voitu muodostaa analyysien kannalta uskottavasti toimivaa summamuuttujaa. Tavoite tulee huomioitua kuitenkin sisältöalue- ja kokonaissummassa.

Arviointitehtävät laadittiin sekä suomeksi että ruotsiksi, ja tehtävien kieliversiot vastasivat toisiaan mahdollisimman täsmällisesti. Suomea tai ruotsia toisena kielenä (S2) opiskelevat oppilaat suorittivat saman tehtäväsarjan kuin äidinkielenään suomea tai ruotsia opiskelevat oppilaat koulunsa opetuskielen mukaisesti. Näin voitiin tuottaa tietoa osaamisen eroista myös äidinkielen oppimäärän perusteella.

TAULUKKO 2.3. Eri testiversioiden sisällöllinen kattavuus

	Ver1	VerS1	VerS2	VerS3	VerS4	VerS5	VerS6	osuus tehtävistä (%)	suunniteltu osuus (%)
Osioita yhteensä	60	60	58	59	58	59	58		
Maksimipistemäärä	65	71	67	69	67	72	79		
Oppilaita yhteensä	4 326	1 439	1 445	1 371	1 359	1 287	1 255		
osioiden määrät eri sisältöalueilla									
S1 (Ajattelu)	17	25	15	15	15	16	15	22,5	16,7
S2 (Luvut ja laskutoimitukset.)	14	17	21	11	11	12	11	18,5	12,5
S3 (Algebra)	9	10	8	22	9	10	9	14,7	9,5
S4 (Funktiot)	8	6	6	6	15	6	6	10,1	9,5
S5 (Geometria)	19	17	14	18	15	26	15	23,7	32,4
S6 (Tilasto ja todennäköisyys)	8	7	7	6	6	6	15	10,5	19,5
osioiden määrät eri tavoitteilla									
T10 (ajattelu)	13	21	13	13	13	14	13	16,6	8,3
T11 (peruslaskut)	7	8	12	6	6	6	6	8,5	2,6
T12 (reaaliluku)	2	0	4	0	0	0	0	1,0	3,4
T13 (prosentti)	7	6	10	6	6	7	7	8,2	6,5
T14 (yhtälöt)	10	8	8	21	9	9	8	12,1	7,2
T15 (funktio)	8	6	6	6	14	6	6	8,7	2,3
T16 (geometria)	8	8	7	8	7	16	7	10,1	11,0
T17 (trigonometria)	10	9	9	9	9	11	9	11,0	9,6
T18 (pinta-ala)	13	10	10	12	10	12	10	12,8	11,8
T19 (todennäköisyys)	7	6	6	5	5	5	14	8,0	11,0
T20 (algoritmit ja ohjelmointi)	4	4	2	2	2	2	2	3,0	8,3
osioiden määrät eri ajattelun tasoilla									
H1 (muistaminen)	5	4	4	13	8	10	5	12,0	15
H2 (ymmärtäminen)	23	25	20	19	21	18	24	36,9	30
H3 (soveltaminen)	23	20	25	20	19	23	20	36,9	40
H4 (korkeammat taidot)	7	9	8	9	9	8	8	14,3	15
osioiden määrät tehtävätyypeittäin									
TE1 (päässälasku)	8	10	7	7	7	7	7	13,0	
TE2 (objektiiviset tehtävät)	40	38	38	40	38	39	39	65,8	
TE3 (ongelmanratkaisut)	12	12	13	12	13	13	12	21,2	

Lopullisissa tehtäväsarjoissa oli kaikkiaan 127 erillistä osiota[3], jotka tyypillisesti olivat automaattisesti pisteitettyjä monivalinta-, yhdistämis- tai lyhytvastauksia edellyttäviä ns. objektiivisia tehtäviä (79 %). Lyhytvastauksien osalta järjestelmään oli etukäteen syötetty useita erilaisia pisteen tuottavia hyväksyttäviä vastausmalleja. Hylättyjä vastauksia ei käyty systemaattisesti läpi. Lopullisista tehtävistä 12 (9 %) oli opettajan pisteittämiä—nämä olivat käytännössä kaikki perusteluja vaativia tehtäviä. Jokaisessa tehtäväsarjassa oli 6–9 tuottamistehtävää, joissa oppilaalta vaadittiin erilaisia perusteluja ja matemaattisen ajattelunsa auki kirjoittamista. Näissä tehtävissä automaattista pisteitystä ei voitu hyödyntää, vaan opettajat pisteittivät ne annettujen pisteittämisohjeiden perusteella. Kyseiset tehtävät myös sensoroitiin, eli pisteityksen oikeellisuus tarkistettiin Karvissa kesällä 2021, jolloin myös satunnaiset virheellisesti annetut pisteet korjattiin. Automaattisesti pisteitettyjä vastauksia ei sensoroitu systemaattisesti.

Korkean pistemäärän—tai keskimääräisenkään—saaminen pelkällä arvaamisella on testissä epätodennäköistä.[4] Hyvää suoritusta ei siis arvioinnissa pysty tekemään vain arvaamalla. Jos testissä huijaaminen on mahdotonta, hyvä suoritus on aina aidosti hyvä suoritus. Sen sijaan heikon suorituksen voi tehdä myös osaamiseltaan keskitasoinen tai hyväkin oppilas, mikäli ilmenee teknisiä ongelmia tai motivaatio arviointitehtävien suorittamiseen on vähäinen. Parhaan osaamisen näyttämiseen pyrittiin motivoimaan sillä, että opettaja saattoi käyttää arvioinnin kautta saatavaa tietoa oppilasarvioinnin tukena. Opettajia varten mallinnettiin kansallisen arvosanajakauman ja arvioinnissa osoitetun osaamisen perusteella vertailukelpoinen arvosanaehdotus arviointiin osallistuneille oppilaille. Tämän toivottiin myös toimivan eräänlaisena kalibroinnin työvälineenä niissä kouluissa, joissa arvosanalinjat ovat ehkä olleet näytettyyn osaamiseen nähden muita kouluja tiukemmat tai löysemmät.

Osaamismittareiden validiteetti ja reliabiliteetti

Tehtäväsarjojen ja yksittäisten tehtävien laatimisessa noudatettiin seitsemää perusperiaatetta: (1) tehtävien tulee olla suoraan yhteydessä POPS:n sisältöihin (jotta voidaan taata mittauksen rakennevaliditeetti), (2) tehtäväsarjojen tulee kattaa sisältöalueet niin laajasti kuin mahdollista (sisällön validiteetti), (3) tehtävien tulee kattaa riittävällä laajuudella kaikki ajattelun tasot (ekologinen validiteetti), (4) tehtäväsarjojen tulee olla sopivan vaikeita (*face*- tai ”näennäis”validiteetti), (5) mittareiden tulee olla erottelukyvyltään tarkkoja (reliabiliteetti), (6) tehtäväsarjojen tulee linkittyä toisiinsa sopivasti valituilla linkkitehtävillä (jotta voidaan käyttää IRT-mallinnusta ja vertaistaa tulokset aiempina vuosina tehtyihin arviointiin) ja (7) tuloksia tulee voida tarkastella suhteessa aiempiin 9. luokan mittauksiin, alempien vuosiluokkien tuloksiin sekä ylioppilaskokeen tehtäviin. Näistä tehtäväsarjojen ja kokonaismittarin validiutta ja reliabiliteettia käsitellään tässä yhteydessä ja tehtäväsarjojen linkitystä toisiinsa ja muihin arviointeihin luvussa 2.3.

3 Aiemmin raportoitiin, että tehtäviä oli 133 (Metsämuuronen & Nousiainen, 2021). Teknisesti tämä pitää paikkansa. Ennen lopullista analyysia tehtävistä viisi kuitenkin poistettiin heikon erottelukyvyn vuoksi. Tehtävät osoittautuivat liian vaikeiksi oppilaille.

4 Eri tehtäväsarjoissa saattoi täysin arvaamalla saada keskimäärin 9–11 monivalintatehtävää oikein riippuen monivalintatehtävien määrästä ja vaihtoehtojen määrästä. Täysin arvaamalla oppilas olisi siis teoriassa päätynyt pistemäärään 243–268 eli noin 2,5 hajontayksikköä keskiarvoa huonompaan suoritukseen käytetyllä asteikolla, jossa keskiarvo on 500 ja hajonta 100. Tosiasiallisesti arvauksen todennäköisyys on pienempi, sillä oppilaat harvoin arvaavat vastauksia täysin satunnaisesti, vaan päätyvät tiettyihin vääriin vastausvaihtoehtoihin tietyn tyyppisen laskuvirheen seurauksena.

Mittareiden validiteetti

Rakennevaliditeetin ja sisällön validiteetti viittaavat siihen, että tehtäväsarjojen rakenteet ja sisällöt perustuvat ”teoreettiseen viitekehykseen” eli POPS:aan (ks. Metsämuuronen, 2009b, 2017b). Validiteetin käsitteen perussisältö täyttyy: juuri näitä sisältöalueita ja tavoitteita pitikin arvioida (ks. Taulukko 3).

Ekologinen validiteetti ja *face*-validiteetti viittaavat siihen, että tehtäväsarjat ovat käytännön koulutyön näkökannalta mielekkäitä. Arvioinnin tehtävät kattavat erilaisia ajattelun tasoja painottuen soveltamiseen, ja tehtävien keskimääräiset vaikeustasot osoittavat lisäksi, että keskimäärin noin puolet tehtävistä osattiin oikein (p = 0,50; ks. Taulukko 4). Tämä mahdollistaa sen, että mittareiden erottelukyky eli reliabiliteetti voi tulla korkeaksi (Lord, 1952), sillä mittarin reliabiliteetti on suoraan yhteydessä osioiden ja summamuuttujan välisten korrelaatioiden suuruuteen (ks. Metsämuuronen, 2022a–2022f). Yleensä Karvin tehtäväsarjoissa tehtävät on valittu niin, että ratkaisuosuudet ovat kuitenkin hieman tätä korkeammalla—noin tasolla p = 0,60–0,65 (Metsämuuronen, 2009a), sillä monivalintatehtävissä arvauksen mahdollisuus kohottaa ratkaisuprosenttia (ks. Lord 1952). Tämän arvioinnin tehtäväsarjoissa käytettiin aiempaa useampia haasteellisia tehtäviä, mikä osaltaan laski ratkaisuosuuksia. Osa selityksestä tulee myös digitaalisena toteutetun arvioinnin luonteesta, jota pohditaan luvussa 2.3.2; oppilaat jättivät vastaamatta useisiin tehtäviin. Keskimäärin tehtäväsarjat olivat kuitenkin ”sopivia” tai hieman vaikeahkoja 9. luokan oppilaille.

Se, että keskimääräinen vaikeustaso ratkaisuprosentteina mitattuna näyttää selvästi matalammalta kuin aiemmissa mittauksissa, ei vaikuta suoraan lopullisen vaikeustason määrittymiseen. Tämä johtuu siitä, että ennen analyyseja pistemäärät vertaistetaan yhteismitalliseksi aiempien vuosien aineistojen kanssa. Tätä käsitellään luvussa 2.3.

Yhteenvetona voidaan todeta, että arvioinnin validiteetti eli osuvuus on riittävä—ellei jopa hyvä—uskottavien päätelmien tekemiseen matematiikan osaamisesta 9. luokan lopussa: mittarin rakenteen, sisällön ja sopivuuden (ekologisen ja näennäisvaliditeetin) osalta arviointi on ollut osuva.

Mittareiden reliabiliteetti

Yksittäisten tehtäväsarjojen ja näistä koostetun lopullisen yhdistetyn aineiston summamuuttujien erottelukyky eli reliabiliteetti on korkea kuten tyypillisesti Karvin oppimistulosarvioinneissa yksittäisissä tehtäväsarjoissa. Eri tehtäväsarjoista muodostetun summamuuttujan reliabiliteetin laskemiseksi ei ole yleisessä käytössä olevia reliabiliteetin mittoja. Tällaisia kuitenkin kehitettiin arvioinnin sivutuotoksena. Kertoimiin liittyviä teoreettisia yksityiskohtia kuvaa Metsämuuronen (2022d) ja oleellisia seikkoja on kuvattu myös tämän luvun liitteessä 1.

Yhdistetyn aineiston reliabiliteetin arviointiin liittyy kaksi keskestä näkökulmaa. Yhtäältä yhdistetyn aineiston reliabiliteetti voidaan estimoida eri mittaversioiden reliabiliteettien keskiarvojen avulla ja toisaalta yksittäisten osioiden erottelukyvyn avulla (ks. taulukot 4a ja 4b). Toisaalta perinteisillä reliabiliteettikertoimilla (alfa, theta, omega, maximal reliability) on taipumus antaa aliarvio reliabiliteetista ja tätä korjataan ns. deflaatiokorjattujen reliabiliteettikertoimien avulla (Metsämuuronen, 2022a, 2022b, 2022f; ks. myös Liite 1).

Jos reliabiliteettia arvioi tehtäväsarjojen keskimääräisten reliabiliteettien avulla, kokonaisosaamisen osalta reliabiliteetti on tasolla 0,91–0,92, kun sitä estimoidaan traditionaalisella alfa-kertoimella (ks. kaava 9 Liitteessä 1) ja tasolla 0,95–0,96, jos käytetään deflaatiokorjattua alfaa (ks. kaavat 10 ja 11 Liitteessä 1). Hieman tehokkaammilla kertoimilla (theta- ja omegakertoimilla; ks. kaavat 3 ja 4 Liitteessä 1) deflaatiokorjaus johtaa hieman korkeampiin arvoihin 0,96–0,97. Jos reliabiliteettia arvioi yksittäisten osioiden erottelukyvyn avulla (ks. kaavat 7 ja 8 Liitteessä 1), reliabiliteetit ovat vielä tätäkin korkeampia; kaikki kertoimet antavat yhdenmukaisesti arvioksi 0,98).

TAULUKKO 2.4a. Kokonaisosaamisen mittarin erottelukykyyn liittyviä tunnuslukuja eri versioissa

	Ver1	VerS1	VerS2	VerS3	VerS4	VerS5	VerS6	keskiarvo
Osioiden määrä	60	60	58	59	58	59	58	
Otoskoko	4 451	1 477	1 478	1 408	1 393	1 323	1 299	
vaikeustaso (p) keskiarvo	0,525	0,483	0,494	0,523	0,481	0,514	0,496	0,502
vaikeustaso (p) minimi	0,032	0,019	0,021	0,028	0,027	0,023	0,028	
vaikeustaso (p) maksimi	0,916	0,936	0,916	0,906	0,915	0,904	0,986	
Reliabiliteetti, traditionaaliset kertoimet								
alfa	0,922	0,916	0,912	0,918	0,912	0,906	0,912	0,914
theta1)	0,931	0,926	0,922	0,927	0,923	0,920	0,924	0,925
omega1)	0,927	0,920	0,916	0,921	0,916	0,914	*(2)	0,919
Reliabiliteetti; deflaatiokorjatut kertoimet; korjaus Somersin D:llä								
alfaD	0,960	0,959	0,956	0,958	0,958	0,955	0,959	0,958
thetaD	0,966	0,966	0,963	0,964	0,965	0,964	0,967	0,965
omegaD	0,966	0,966	0,963	0,965	0,965	0,964	0,967	0,965
Reliabiliteetti; deflaatiokorjatut kertoimet; korjaus Goodmanin-Kruskalin G:llä								
alfaG	0,962	0,961	0,959	0,961	0,961	0,957	0,961	0,960
thetaG	0,967	0,968	0,965	0,966	0,967	0,966	0,969	0,967
omegaG	0,967	0,968	0,965	0,967	0,967	0,966	0,969	0,967
Reliabiliteetti; osioista laskettu; deflaatiokorjatut kertoimet; korjaus D:llä ja G:llä								
thetaDH								0,984
omegaDH								0,983
thetaGH								0,985
omegaGH								0,984
Keskimääräinen osio-summa-korrelaatio								
R	0,420	0,408	0,404	0,413	0,406	0,394	0,407	0,407
D	0,561	0,564	0,551	0,558	0,564	0,552	0,574	0,561
G	0,571	0,575	0,562	0,569	0,575	0,563	0,585	0,572

1) Huom. summamuuttuja on oleellisesti eri kuin vertaistettu summamuuttuja
2) estimaattia ei muodostu, koska vaadittava matriisi ei ole ”positiivisesti definiitti” eli laskuoperaatiossa joudutaan tekemään ”nollalla jakaminen”, jota ei ole määritelty. Tämänkaltainen tilanne syntyy, jos yhdenkin osion korrelaation summamuuttujaan on täydellinen (*R* = 1)

TAULUKKO 2.4b. Sisältöalueita koskevien mittareiden erottelukykyyn liittyviä tunnuslukuja

sisältöalueet			traditio-naalinen reliabiliteetti	deflaatiokorjattu reliabiliteetti; korjaus D:llä		deflaatiokorjattu reliabiliteetti; korjaus G:llä	
		kerroin (kaava Liitteessä 1)	tehtävä-sarjojen keskiarvo (kaava 9)	tehtävä-sarjojen keskiarvo (kaavat 3 ja 4)	osioiden perusteella (kaavat 7 ja 8)	tehtävä-sarjojen keskiarvo (kaavat 3 ja 4)	osioiden perusteella (kaavat 7 ja 8)
S1	Ajattelun taidot ja menetelmät	alfa	0,801				
		theta (3, 7)	1)	0,899	0,945	0,920	0,954
		omega (4, 8)	1)	0,914	0,948	0,930	0,956
S2	Luvut ja lasku-toimitukset	alfa	0,771				
		theta	1)	0,920	0,957	0,936	0,965
		omega	1)	0,930	0,959	0,942	0,966
S3	Algebra	alfa	0,789				
		theta	1)	0,913	0,953	0,931	0,960
		omega	1)	0,919	0,955	0,925	0,961
S4	Funktiot	alfa	0,586				
		theta	1)	0,783	0,927	0,868	0,948
		omega	1)	0,843	0,933	0,898	0,951
S5	Geometria	alfa	0,612				
		theta	1)	0,842	0,954	0,890	0,964
		omega	1)	0,879	0,957	0,912	0,967
S6	Tietojen käsittely ja tilastot sekä toden-näköisyys	alfa	0,551				
		theta	1)	0,929	0,954	0,967	0,971
		omega	1)	0,942	0,958	0,972	0,973

1) ei soveltuva kertoimeksi, koska näissä kertoimissa summamuuttuja on oleellisesti eri kuin mittauksessa käytetyt summamuuttujat

Kaikki arviot viittaavat siihen, että osaamismittarit ovat erittäin luotettavia. Sisältöalueiden summien reliabiliteetit ovat hieman matalammalla tasolla sisältöalueesta riippuen ja eri kertoimet antavat hieman suurempaa vaihtelua arvioon kuin kokonaissumman osalta (0,78–0,97). Sisältöalueiden summien reliabiliteetit ovat myös selkeästi deflatoituneita, jos arviona käytetään traditionaalisia kertoimia (alfa vaihtelee 0,55–0,80). Tämä johtuu yhtäältä siitä, että osa mittareista on lyhyitä ja toisaalta siitä, että niissä on vaikeustasoltaan äärimmäisiä osioita. Tällöin perinteisissä kertoimissa sisällä oleva osio–summa korrelaatio voi antaa suurenkin aliarvion todellisesta korrelaatiosta.

Muiden mittareiden ja taustamuuttujien luotettavuustarkasteluja

Osa oppilaita koskevista tiedoista saatiin Koski-tietovarannosta oppilaan oppijanumeron perusteella. Koski-tietovarannosta saatiin tieto muun muassa oppilaan sukupuolesta, arvosanoista, S2-statuksesta ja kolmiportaisen tuen tasosta. Kouluja koskevat tiedot saatiin vastaavasti Opetushallituksen ylläpitämästä Vipunen-tietokannasta. Osa tiedoista on koulujen tallentamia, ja niissä havaittiin jonkin verran puutteita ja suoranaisia virheitä (esim. virheellisiä oppijanumeroita), joiden vaikutus luotettavuuteen jää kuitenkin vähäiseksi. Näiden mahdollisten virheellisten tietojen osuus rekisterissä kuitenkin pienenee ajan myötä.

Arviointitehtävien lisäksi oppilaat vastasivat oppilaskyselyyn, jonka avulla kartoitettiin muun muassa asennoitumista matematiikkaan, huoltajien koulutustaustaa sekä käytäntöjä matematiikan tunneilla. Oppilaskyselyn näkökulmasta keskeisiä tässä raportissa kuvattavia kokonaisuuksia ovat matematiikkaan koskevat asenteet, emootiot matematiikan yhteydessä sekä koulukiusaaminen.

Asennemittarit

Asenteisiin liittyvä kokonaisuus on sama Fenneman ja Shermanin (1976) testin (*Fennema–Sherman attitude scale*, FSAS) pohjalta muokattu 15 osion kokonaisuus, jonka tuloksia on raportoitu samanlaisena vuodesta 2001 lähtien (ks. Metsämuuronen, 2009a; 2012). Asennemittarissa on kolme ulottuvuutta: käsitys itsestä oppiaineen osaajana (OSAA), josta usein käytetään nimitystä ”minäpystyvyys”, oppiaineesta pitäminen (PITÄÄ), ja käsitys oppiaineen hyödyllisyydestä (HYÖTY). Reliabiliteetit ovat tyypillisesti olleet melko korkeita ($\alpha \approx 0{,}90$).

Kutakin dimensiota mitataan viidellä osiolla (Taulukko 8). Alkuperäistä Fenneman ja Shermanin mittaria on lyhennetty, ja sen pituus vastaa kansainvälisissä sovelluksissa käytettyä versiota (ks. esimerkiksi *Programme for International Student Assessment* [PISA], OECD, 2003a; 2006 ja *Trends in International Mathematics and Science Study* [TIMSS] -mittaukset, TIMSS, 2003; 2006; ks. Metsämuuronen, 2009a, 2012). Suomalaisessa versiossa väitteitä on selvästi yksinkertaistettu ja negatiivisia osioita on joko käännetty tai poistettu niin, että kahdella dimensioista (PITÄÄ ja HYÖTY) on vain yksi käänteinen osio ja yhdellä (OSAA) kaksi käänteistä osiota. Lisäksi kansainvälisissä mittauksissa asteikosta on käytetty 4-portaista asteikkoa, mutta Karvin mittauksissa asteikossa on ylemmillä vuosiluokilla käytetty 5-portaista Likertin asteikkoa vaihtoehdoilla *Olen täysin eri mieltä* (1), *Olen jonkin verran eri mieltä* (2), *kantani on epävarma tai minulla ei ole selvää käsitystä* (3), *Olen jonkin verran samaa mieltä* (4) ja *Olen täysin samaa mieltä* (5) (ks. tarkempi vertailu kansainväliseen mittaristoon Metsämuuronen, 2012). Analyyseja varten lopullisten summien asteikko skaalattiin asteikolle, joka vaihteli välillä 0–4. Tällöin arvo 2 vastaa neutraalia, arvot 0–1 negatiivista asennetta ja arvot 3–4 positiivista asennetta.

Kun asenneosioissa vaihtoehtoja on viisi, deflaatiokorjatuissa reliabiliteettikertoimissa käytetyt korrelaatiokertoimet D ja G antavat ilmeisiä aliarvioita osion ja summan yhteydestä (ks. Metsämuuronen, 2020a, 2020b, 2021a). Laskettaessa deflaatiokorjattuja reliabiliteetteja (ks. Metsämuuronen, 2022c, 2022f), on suositeltavampaa käyttää *dimensio*korjattuja kertoimia D_2 ja G_2 (Metsämuuronen, 2020b, 2021a). Kun tiedetään D:n ja G:n arvot ja osiossa olevien kategorioiden määrä, D_2 ja G_2 lasketaan seuraavasti:

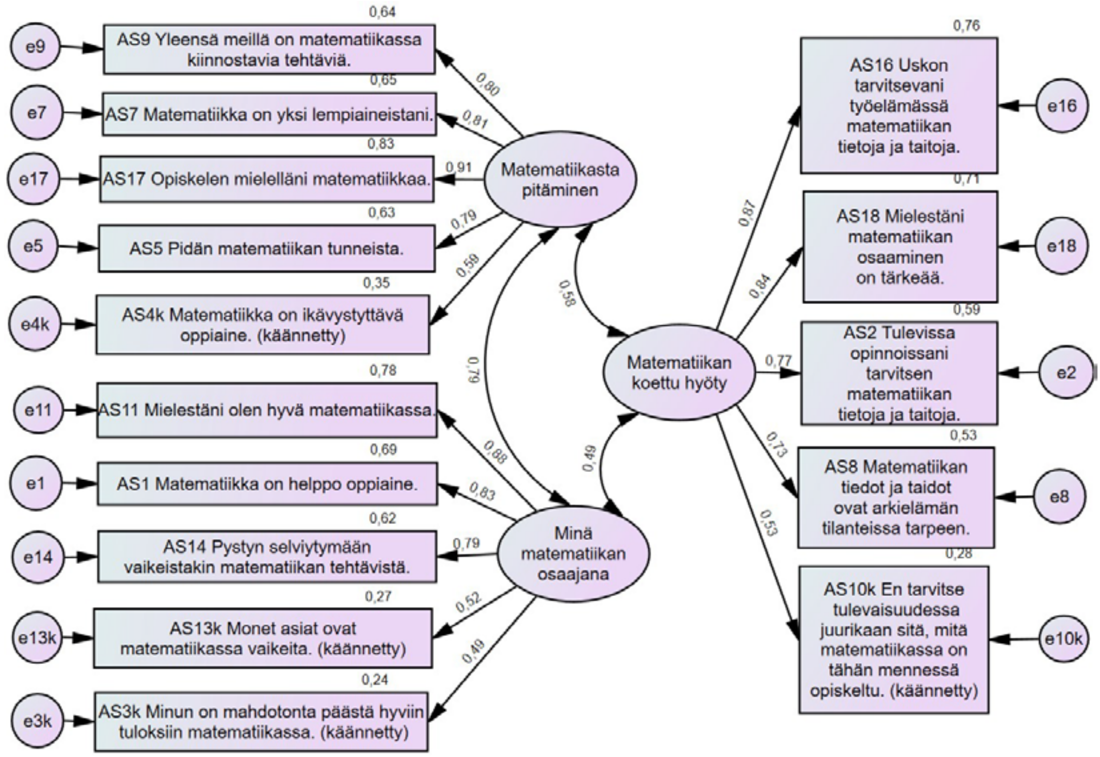
$$D_2 = D \times \left(1+\left(1-abs(D)\right) \times A\right)$$

ja

$$G_2 = G \times \left(1+\left(1-abs(G)\right) \times A\right),$$

missä *abs* viittaa itseisarvoon eli D:n ja G:n arvoon ilman etumerkkiä ja $A = \left[(df(g)-1)/df(g)\right]^3$, missä $df(g)$ on kategorioiden määrä miinus 1. Kun nyt kategorioiden määrä on 5, $df(g) = 4$ ja $A = (3/4)^3 = 0{,}422$. Jos siis D:n arvo olisi 0,80, D_2:n arvo olisi $D_2 = 0{,}80 \times \left(1+\left(1-0{,}80 \times \left(0{,}422\right) = 0{,}867.\right.\right.$

Mittarin rakennetta ja sen yhteyttä alkuperäiseen FSAS-testiin käsittelee esimerkiksi Metsämuuronen (2012). Mittarin rakenne käsiteltävässä aineistossa havainnollistetaan kuviossa 4. Malliin liittyvistä yleisistä suhteellisista sopivuuskertoimista (*fit index*) NFI (*normed fit index*) = 0,915 ja IFI (*incremental fit index*) = 0,916 viittaavat siihen, että aineisto vastaa teoreettista mallia perinteisten nyrkkisääntöjen perusteella riittävän hyvin (NFI:n ja IFI:n tulisi olla yli 0,90), joskin CFI (*comparative fit index*) = 0,916 ja TLI (*Tucker–Lewis index*) eli NNFI (*non-normed fit index*) = 0,884 ovat hieman matalia hyväksyttävälle mallille (perinteisesti rajaa CFI > 0,95 ja TLI, NNFI > 0,90 pidetään alarajana; ks. raja-arvoista esimerkiksi Hu & Bentler, 1999; Schreiber ym. 2012; Schumacker & Lomax, 2004). CFI tosin reagoi suureen otoskokoon negatiivisesti eli se rankaisee suuresta otoskoosta. Vastaavasti TLI on herkkä mallin kompleksisuudelle ja se rankaisee mallin korkeasta parametrien määrästä. Absoluuttisista sopivuusindekseistä *root mean square error of approximation* RMSEA = 0,087 viittaa mallin keskeneräisyyteen; perinteisenä nyrkkisääntönä hyvälle mallille on ollut RMSEA < 0,05 (Browne & Cudeck, 1993) tai < 0,07 (Hooper ym., 2008; Steiger, 2007). Tämän osalta muistetaan, että indeksi on herkkä mallin vapausasteille; monimutkaisempi malli, jossa on enemmän parametreja, saa automaattisesti suuremman indeksin arvon. Mallissa ilmenevä kehittämistarve voi ilmentää ilmiön liiallista yksinkertaistamista. Yleensä havaittuihin muuttujiin liittyvien virheiden eli jäännösvarianssien oletetaan olevan toisiinsa nähden korreloimattomia, mikä oletus ei indeksien perusteella täysin toteudu. Huomataan myös, että käännetyt osiot latautuvat faktoreille järjestelmällisesti matalammin kuin positiivisesti kysytyt osiot.



KUVIO 2.4. Asennetestin mittausmalli

Kuvion 4 malli havainnollistaa, miten kolme faktoria, käsitys itsestä matematiikan osaajana, matematiikasta pitäminen ja matematiikan koettu hyöty korreloivat keskenään selvästi (*r* = 0,49–0,79). Erityisesti matematiikasta pitäminen ja kokemus itsestä matematiikan osaajana korreloivat selvästi (*r* = 0,79).

Asennemittareiden reliabiliteetit ovat riittävän korkeita, jotta oppilaat voidaan erotella toisistaan luotettavasti. Alfa vaihtelee 0,83–0,91 ja deflaatiokorjatut kertoimet 0,86–0,95 osa-alueesta riippuen (Taulukko 5).

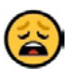
TAULUKKO 2.5. Asennemittareiden osa-alueet

Asenneasteikot	osioiden määrä	reliabiliteetti				
		alfa	thetaD2	omegaD2	thetaG2	omegaG2
Käsitys itsestä matematiikan osaajana (OSAA)	5	0,860	0,897	0,923	0,916	0,937
Matematiikasta pitäminen (PITÄÄ)	5	0,885	0,911	0,933	0,927	0,945
Matematiikan koettu hyödyllisyys (HYÖTY)	5	0,831	0,865	0,902	0,889	0,918
Kokonaisasenne (OSAA + PITÄÄ + HYÖTY)	15	0,910	0,930	0,937	0,933	0,940

Emootiomittarin luotettavuus

Matematiikan opiskeluun ja matematiikkaan itseensä liittyviä emootioita eli ns. akateemisia tunteita tutkittiin uudella mittarilla, jossa yhdistettiin viisi positiivista tunnetilaa (innostunut, kiinnostunut, onnistunut, tyytyväinen ja varma), viisi negatiivista tunnetilaa (vihainen, avuton, ahdistunut, pettynyt, ja epävarma) ja näitä kuvaavat emojit (Kuvio 5). Kysymykseen *Missä määrin alla esitetty tunnetila yhdistyy matematiikan opintoihisi?* vastattiin vaihtoehdoilla *ei lainkaan* (0), *harvoin* (1), *joskus* (2), *usein* (3) ja *lähes aina* (4). Mittari kehiteltiin aiemmassa matematiikan arvioinnissa käytetystä testistä (Metsämuuronen, 2017; Metsämuuronen & Tuohilampi, 2017a; Metsämuuronen & Salonen, 2017), jossa yhdeksän erilaista tunnetilaa (innostus, kiinnostus, tylsyys, pitäminen, turhautuminen, viha, ahdistus, avuttomuus, tyytyväisyys) yhdistettiin matematiikan opiskeluun. Aiemmassa mittarissa oli kuitenkin ongelmana, että apaattista tunnetilaa kuvaava emootio ”tylsyys” ei erottunut itsenäisenä faktorina vaan latautui negatiivisesti positiivisten tunnetilojen faktorille. Jotta faktorirakennetta saataisiin myös tältä osin selkeämmäksi, jatkettiin tässä arvioinnissa mittarin kehittämistä.

Mieti matematiikan opiskeluasi kokonaisuutena. Missä määrin alla esitetty tunnetila yhdistyy matematiikan opiskeluusi?
1 = ei lainkaan, 2 = harvoin, 3 = joskus, 4 = usein, 5 = lähes aina

		ei lainkaan	harvoin	joskus	usein	lähes aina
	innostunut					
	kiinnostunut					
	onnistunut					
	tyytyväinen					
	varma					
	vihainen					
	avuton					
	ahdistunut					
	pettynyt					
	epävarma					

KUVIO 2.5. Emootio ja emojit tunnemittarissa

Uutta mittaria kehiteltäessä pyrittiin löytämään aiempaa selkeämpiä tunnetilan ilmaisuja—myös visuaalisesti—ja välttämään neutraalia tai apaattista emootiota kuvaavia osioita. Samalla sanamuotoja muokattiin henkilökohtaisemmaksi; esimerkiksi ”kiinnostus” muutettiin muotoon ”kiinnostunut” ja ”innostuneisuus” muotoon ”innostunut”. Positiivisista emootioista mittarissa säilytettiin *innostunut*, *kiinnostunut* ja *tyytyväinen*, ja pitämisen tilalle valittiin *onnistunut* ja *varma*. Jälkimmäiset liittyvät selkeämmin standardimittariin sisältyneissä ”käsitys itsestä matematiikan osaajana” tai ”oppiaineesta pitäminen” faktoreissa olleisiin osioihin (ks. Taulukko 6 ja Kuvio 6a). Negatiivisista emootioista mittarissa säilytettiin *vihainen*, *ahdistunut* ja *avuton* ja turhautumisen tilalle valittiin *pettynyt* ja *epävarma*. Emootioista vastinpari *varma—epävarma* toimii mittarissa kontrollimekanismina: loogisesti jos vastaaja olisi varma, hän ei voisi olla epävarma.

Uudistetun mittarin faktorirakenne on selkeä: oppilaat jakautuvat selkeästi positiivisesti ja negatiivisesti matematiikasta ajatteleviin ryhmiin (Taulukko 6). Tosin asia ei ole aivan niin yksinkertainen, mikä havainnollistuu, kun tarkastellaan vastinparia *varma—epävarma* (Taulukko 7). Näiden muuttujien välinen korrelaatio on -0,366, mikä on tilastollisesti merkitsevästi nollasta poikkeava ja voimakkuudeltaankin merkittävän suuri, mutta yllättävän matala ottaen huomioon, että tunnetilat periaatteessa kumoavat toisensa: voisimme jopa odottaa lähes täydellistä negatiivista korrelaatiota. Huomataan kuitenkin, että oppilaista, jotka ilmaisivat olevansa *lähes aina* ”varmoja” matematiikan suhteen, 10 prosenttia ilmaisi olevansa myös ”epävarmoja” lähes aina (Taulukko 7). Vastaavasti oppilaista, jotka ilmaisivat, että *eivät koskaan* olleet ”varmoja”, 21 prosenttia ei koskaan kokenut olevansa myöskään ”epävarma”. On epätodennäköistä, että tuhannet oppilaat olisivat vastanneet emootiokysymyksiin tietoisesti virheellisesti. Todennäköisempää on, että epäloogisuudet ilmentävät emootioihin liittyvää epävakautta; välillä voi olla varma olo ja tunne voi muuttua hetkessä. Emootioihin liittyvää systematiikkaa ja niiden yhteyttä osaamiseen tutkitaan tulevissa raporteissa.

TAULUKKO 2.6. Emootiomittareiden faktorit

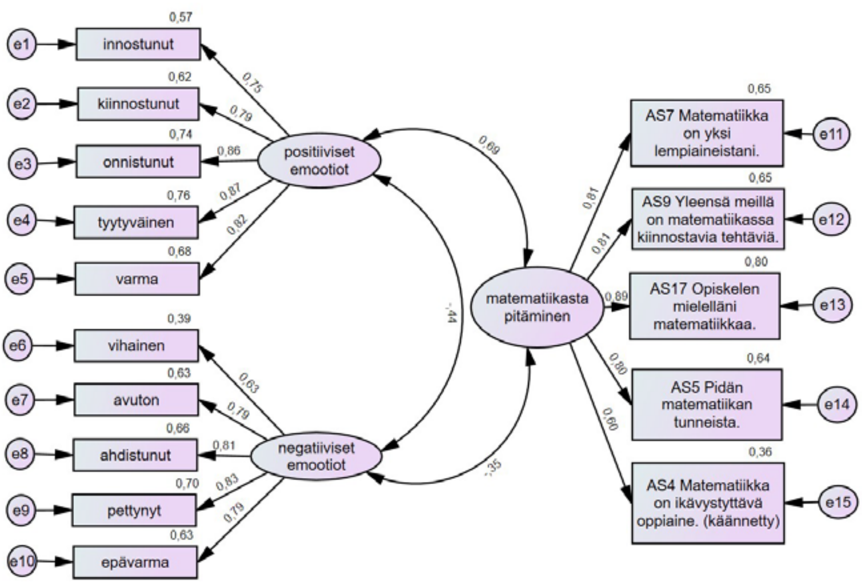
	Faktori[1] ja faktorilataukset	
	1 (positiiviset tunnetilat)	2 (negatiiviset tunnetilat)
innostunut	0,772	0,071
kiinnostunut	0,818	0,089
onnistunut	0,855	-0,036
tyytyväinen	0,861	-0,044
varma	0,771	-0,115
vihainen	-0,084	0,574
avuton	-0,050	0,762
ahdistunut	0,024	0,825
pettynyt	0,045	0,863
epävarma	0,050	0,820

1) Hajotusmenetelmä: Maximum Likelihood.
Rotaatiomenetelmä: Promax with Kaiser Normalization.

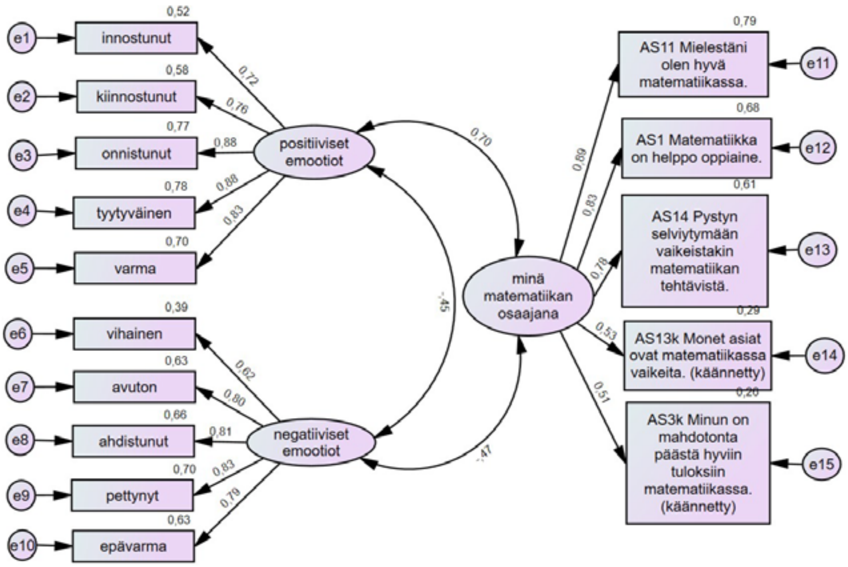
TAULUKKO 2.7. ”Varmaksi” ja ”epävarmaksi” kokeneiden oppilaiden yhteys

		Tunnetila: epävarma (%)					
		ei lainkaan	**harvoin**	**joskus**	**usein**	**lähes aina**	**N**
Tunnetila: varma (%)	**ei lainkaan**	20,9	6,9	11,9	18,5	**41,8**	1158
	harvoin	5,6	20,2	25,8	**33,2**	15,2	2200
	joskus	9,7	22,2	**45,6**	17,9	4,6	3942
	usein	20,8	**37,6**	28,2	11,7	1,7	2739
	lähes aina	**44,2**	28,7	12,7	4,4	10,1	663
	N	1611	2622	3358	1997	1114	10702

Uuden mittarin rakennetta, ominaisuuksia ja kehittämisehdotuksia käsitellään tarkemmin luvussa 5 (Salonen, 2023). Tässä yhteydessä huomataan kuitenkin, että positiivisten emootioiden muodostama faktori korreloi voimakkaan positiivisesti standardimittarin faktoriin ”matematiikasta pitäminen” (*r* = 0,69), ja kielteiset emootiot korreloivat selvästi negatiivisesti (*r* = –0,35; Kuvio 6a). Kuten asennemittarin yhteydessä, mallin yleiset yhteensopivuusindeksit kertovat, että testattu malli todentuu havaitussa aineistossa riittävän hyvin (NFI = IFI = CFI = 0,922), joskin kehittämisvaraa on (RSMEA = 0,088). Mallin perusteella positiivisten emootioiden muodostama faktori korreloi vahvan positiivisesti myös standardimittarin faktoriin ”minä matematiikan osaajana” (*r* = 0,70), kun taas kielteiset emootiot korreloivat siihen vielä negatiivisemmin (*r* = –0,47) kuin matematiikasta pitämisen faktoriin (*r* =–0,35; Kuvio 6b). Myös tämän mallin osalta yleiset yhteensopivuusindeksit kertovat, että testattu malli todentuu havaitussa aineistossa riittävän hyvin (NFI = IFI = CFI = 0,922), joskin kehittämisvaraa on (RSMEA = 0,084).



KUVIO 2.6a. Emootiot ja ”matematiikasta pitäminen” faktori



KUVIO 2.6b. Emootiot ja "minä matematiikan osaajana" faktori

Emootiomittareiden korkeat reliabiliteetit (positiiviset tunnetilat 0,94–0,96 ja negatiiviset tunnetilat 0,90–0,94) heijastavat selkeää faktorirakennetta tai ainakin sitä, että molemmissa faktoreissa oppilaat pystytään erottelemaan toisistaan erittäin tarkasti (Taulukko 8). Huomataan kuitenkin, että negatiivisten tunnetilojen mittarin erottelukyky on jossain määrin matalampi kuin positiivisten tunnetilojen.

TAULUKKO 2.8. Emootiomittareiden reliabiliteetit

Tunnetila matematiikan opiskelussa	osioiden määrä	reliabiliteetti				
		alfa	thetaD2	omegaD2	thetaG2	omegaG2
positiiviset tunnetilat	5	0,910	0,937	0,952	0,953	0,963
negatiiviset tunnetilat	5	0,879	0,902	0,927	0,920	0,940
positiivinen kokonaisuutena	10	0,853	0,874	0,922	0,883	0,930

Huomataan myös, että jos negatiivisen tunnetilan muuttujat käännetään positiivisiksi ja lasketaan yhteen positiivisten muuttujien kanssa—kuten yleensä tehdään asennetyyppisissä testeissä (ks. esimerkiksi edellä Fennema–Sherman-mittari)—näin syntyvän kokonaismittarin reliabiliteetti on matalampi kuin kummankaan osa-mittarin reliabiliteetti (0,87–0,93). Tämä ei ole tyypillistä, ja kertonee siitä, että positiivisten ja negatiivisten emootioiden samanaikainen kokeminen on mahdollista, jonka vuoksi nämä summat on järkevää pitää erillisinä. Osa selitystä voi olla myös Pekrunin (mm. 2017) oppimis- ja opettamiskontekstiin kehitetyssä ns. akateemisten emootioiden viitekehyksessä oleva ajatus siitä, että sekä positiiviset että negatiiviset tunnetilat saattavat olla aktivoivia tai passivoivia. Tätä käsittelee tarkemmin Salonen (2023) luvussa 5.

Kiusaamismittarin luotettavuus

Kiusaamista käsittelevä kuuden osion kokonaisuus on lainattu PISA- ja TIMSS-mittauksista (OECD, 2019). Siinä kiusaamista tarkastellaan järjestelmällisenä psyykkisenä, sosiaalisena ja fyysisenä toimintana oppilasta vastaan. Asteikkona oli neliportainen järjestysasteikko (1) ei lainkaan, (2) harvemmin kuin kerran viikossa, (3) noin kerran viikossa ja (4) useita kertoja viikossa. Faktorianalyysin perusteella ilmiö on yksiulotteinen (Taulukko 9). Kiusaamista koskevan summamuuttujan reliabiliteetti on korkea (0,98–0,99) eli mittari pystyy erottelemaan enemmän kiusatut vähemmän kiusatuista lähes täydellisesti. Erot kiusattujen ja ei-kiusattujen välillä ovat aineistossa selvät: osa oppilaista ei koe kiusaamista lainkaan, osaa kiusataan monella tavalla ja usein.

TAULUKKO 2.9. Kiusaamismittarin osiot ja faktoripisteet

	Faktorilataus[1]
	”kiusaamisen vakavuus”
Oppilaat ovat jättäneet minut tarkoituksella ulkopuoliseksi	*0,829*
Oppilaat ovat pilkanneet minua	*0,825*
Oppilaat ovat uhkailleet minua	*0,946*
Oppilaat ovat ottaneet tai tuhonneet minulle kuuluneita esineitä	*0,931*
Oppilaat ovat lyöneet tai tönineet minua	*0,910*
Oppilaat ovat levittäneet minusta ilkeitä juoruja	*0,874*

1) Hajotusmenetelmä: Maximum Likelihood.

reliabiliteetti	alfa	thetaD$_2$	OmegaD$_2$	thetaG$_2$	OmegaG$_2$
Kiusaamisen vakavuus	0,957	0,977	0,981	0,984	0,987

SES-indikaattoreiden luotettavuus

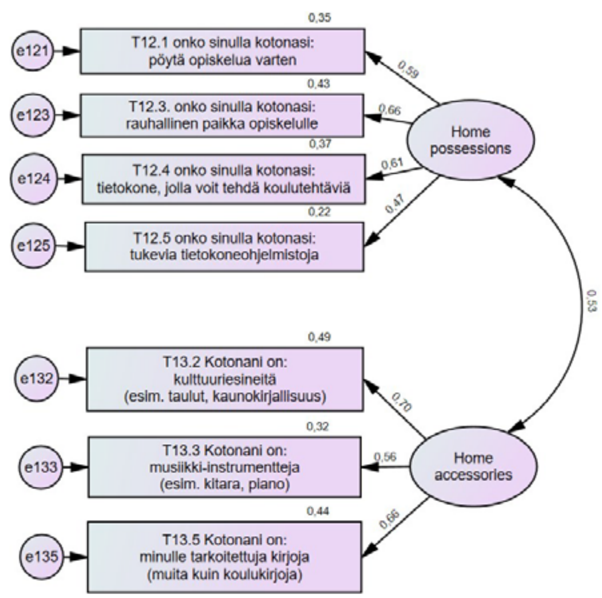
Vanhempien koulutustaustan lisäksi sosioekonomista statusta (SES) kartoittavina kysymyspattereina käytettiin PISA- ja TIMSS-tiedonkeruista lainattuja *Home possessions*– ja *Home accessories*-mittareita (OECD, 2019). Näistä ensin mainittu sisälsi viisi oppilaan käytössä olevaa resurssia kuten pöytä opiskelua varten, rauhallinen paikka opiskeluille ja oma huone. Jälkimmäinen sisälsi kolme kotona olevaa kulttuurista harrastuneisuutta ilmentävää asiaa kuten musiikki-instrumentteja, kirjoja ja taideteoksia (Taulukko 10).

Molemmissa mittareissa perinteinen alfa-kerroin (ja samoin theta ja omega) tuottavat selkeän aliarvion reliabiliteetista (0,614–0,651). Tämä johtuu siitä, että hyvin harvat oppilaat kertoivat kotonaan olevan esimerkiksi e-kirjojen lukulaite. Järjestyskorrelaation näkökannalta (*G* ja *D*) niillä harvoilla oppilailla, joilla esimerkiksi e-kirjojen lukulaite oli kotonaan, saivat korkeampia kokonaispisteitä. Siksi deflaatiokorjatut reliabiliteetit antavat oleellisesti korkeammat arviot reliabiliteetista (0,873–0,986; Taulukko 10). Näistä *Home possessions* -mittari on hieman tarkempi (reliabiliteetti 0,949–0,986) kuin *Home accessories* -mittari (0,873–0,955).

TAULUKKO 2.10. SES-mittareiden erottelukyvyn indikaattoreita

Mittarit			osio-summa korrelaatio		
***Home possessions* -mittari**			**Rit**	**G**	**D**
T12.1 onko sinulla kotonasi: pöytä opiskelua varten			0,632	0,952	0,896
T12.2 onko sinulla kotonasi: oma huone			0,486	0,884	0,797
T12.3. onko sinulla kotonasi: rauhallinen paikka opiskelulle			0,730	0,967	0,901
T12.4 onko sinulla kotonasi: tietokone, jolla voit tehdä koulutehtäviä			0,698	0,961	0,908
T12.5 onko sinulla kotonasi: tukevia tietokoneohjelmistoja			0,699	0,976	0,929
***Home accessories* -mittari**					
T13.1 Kotonani on: E-kirjojen lukulaitteita (esim. Kindle)			0,477	0,812	0,716
T13.2 Kotonani on: kulttuuriesineitä (esim. taulut, kaunokirjallisuus)			0,746	0,937	0,858
T13.3 Kotonani on: musiikki-instrumentteja (esim. kitara, piano)			0,685	0,88	0,785
T13.4 Kotonani on: kaikilla lapsilla oma huone			0,463	0,686	0,592
T13.5 Kotonani on: minulle tarkoitettuja kirjoja (muita kuin koulukirjoja)			0,721	0,921	0,839
reliabiliteetti	**alfa**	**thetaD**	**omegaD**	**thetaG**	**omegaG**
Home possessions	0,651	0,968	0,949	0,986	0,978
Home accessories	0,614	0,920	0,873	0,955	0,929

Rakenneyhtälömallituksessa malleissa heikosti korreloivien osioiden poisjättäminen parantaa niitä oleellisesti. Jäljelle jää 4 ja 3 osiota, joiden muodostama malli on erittäin toimiva (Kuvio 7). Tätä indikoivat sekä yleiset yhteensopivuuskertoimet (NFI = 0,980, IFI, 0,981, CFI = 0,981) että absoluuttinen yhteensopivuuskerroin (RMSEA = 0,041).



KUVIO 2.7. SES-mittareiden mittausmalli

Toisen vaiheen mittareiden luotettavuustarkasteluja

Arvioinnin ensimmäiseen vaiheeseen osallistuivat kaikki otoskoulujen 9. luokkalaiset. Arvioinnin toisessa vaiheessa käytössä oli kaksi tarkentavaa testiä, joihin oppilaat valittiin ensimmäisen vaiheen suoriutumisen ja viimeisimmän matematiikan arvosanan perusteella. Toinen testeistä oli suunniteltu toiminnallisen laskutaidon diagnostiseksi testiksi niille oppilaille, jotka suoriutuivat arvioinnin ensimmäisessä vaiheessa heikosti ja joiden matematiikan kouluarvosana oli 5 tai 6 (nk. FUNA-testi; Funa, 2019; Räsänen ym., 2021). Näille oppilaille valittiin vertailuryhmäksi oppilaat, joiden arvioinnissa osoittama osaaminen oli keskimäärästä ja arvosana oli 8. Tarkoituksena oli ensimmäistä kertaa selvittää, minkälaisia dyskalkuliaan liittyviä haasteita oppilailla mahdollisesti on ja kyetä ehkä tarjoamaan tulosten pohjalta apuvälineitä opettajille näiden haasteiden voittamiseen jo varhaisemmilla luokilla. Aineistossa ei ollut tietoa siitä, oliko oppilaille tehty erillisiä dyskalkuliaan liittyviä diagnooseja. Oppilaat käyttivät toisen vaiheen tehtävien tekemiseen enintään 45 minuuttia. FUNA-testin tehtävät olivat kokonaisuudessaan automaattisesti pisteitettyjä ja Karvin asiantuntijat pisteittivät ylöspäin eriytetyn version tehtävät kesällä 2021. Ylöspäin eriytettyä tehtäväsarjaa käsitellään tarkemmin tässä yhteydessä. Molempien testien tuloksia raportoidaan tulevassa julkaisussa.

Tässä yhteydessä tarkennetaan ylöspäin eriytetyn testin luotettavuutta. Useinkaan yksittäiseen tehtäväsarjaan ei voida sisällyttää kuin korkeitaan yksi erittäin vaativa tehtävä, ja näin parhaiden oppilaiden suoritus voi jäädä aliarvioiduksi. Toinen diagnostisista versioista oli suunnattu ensimmäisessä vaiheessa poikkeuksellisen hyvän suorituksen tehneille oppilaille, jotka ratkaisivat varsinaisen tehtäväsarjan tehtävistä 85 % tai enemmän oikein. Näiden oppilaiden arvosana oli tyypillisesti 10 tai 9, mutta myös joitain arvosanan 8 saaneita oppilaita sijoittui tähän ryhmään. Tämä ylöspäin eriytetty lyhyehkö tehtäväsarja (myöhemmin ”vaikea tehtäväsarja”) oli selvästi vaativa ja tarkoitus oli selvittää, kuinka pitkälle matematiikan taidoiltaan taitavimmat oppilaat ovat edenneet 9. luokan loppuun mennessä.

Vaikeassa tehtäväsarjassa oli 13 vaativahkoa tehtävää, jotka yhtä lukuun ottamatta olivat samoja, joita oppilaat testin eri versioissa olivat jo edellisessä vaiheessa ratkaisseet. Ainakin jokin tehtävistä oli siis tuttu kullekin oppilaalle aiemmassa vaiheessa tehdyn tehtäväsarjan perusteella. Tehtävät edustivat laajasti eri sisältöalueita ja tavoitteita, ja valtaosa tehtävistä oli soveltavia siinä mielessä, että tehtävän aihepiiri haettiin arkielämästä (Taulukko 11a).

TAULUKKO 2.11a. Vaikean tehtäväsarjan osuvuuteen liittyviä tietoja

osio	sisältö	tehtävätyyppi	maksimi-pisteet	Sisältöalue (S1–S6)	Tavoite (T10–T20)	Ajattelun taso (H1–H4)
1	epäyhtälö + välivaiheet	ongelmanratkaisu	3	S3	T14	4
2	tasogeometria + epäyhtälö	monivalinta	1	S3	T14	2
3	yhtälönratkaisu + perustelu	ongelmanratkaisu	2	S3	T14	3
4	suunnikkaan piiri + perustelu	ongelmanratkaisu	3	S5	T16	3
5	tikapuiden pituus + lasku + yksikkömuunnos	ongelmanratkaisu	3	S5, S3	T17	3
6	monikulmion pinta-ala + lauseke	ongelmanratkaisu	3	S5	T18, T17	3
7	geogebra (vakion vaikutus funktion)	lyhytvastaus	1	S4	T15	4
8	minimipituuden määritys	monivalinta	1	S5, S1	T16, T10	3
9	funktion tekeminen + lauseke	lyhytvastaus	1	S4	T15, T14	3
10	funktion tekeminen + lauseke	lyhytvastaus	1	S4	T15, T14	3
11	ohjelmointi	lyhytvastaus	1	S1	T20	2
12	todennäköisyys, lauseke	ongelmanratkaisu	2	S6	T19	3
13	todennäköisyys, lauseke	ongelmanratkaisu	2	S6	T19	3

TAULUKKO 2.11b. Vaikean tehtäväsarjan vaikeustasoon ja erottelukykyyn liittyviä tietoja

osio	Puuttuvia tietoja (%)	vaikeustaso			erottelukyky[1]		
		ratkaisuosuus vaikeaan tehtäväsarjaan osallistuneilla	ratkaisuosuus koko aineistossa	ratkaisuosuuksien ero aineistojen välillä	$D_{i\theta}$	$G_{i\theta}$	$R_{i\theta}$
1	1,8	0,49	0,10	0,39	0,521	0,569	0,430
2	0,0	0,83	0,23	0,60	0,439	0,477	0,293
3	0,3	0,93	0,40	0,53	0,513	0,554	0,283
4	0,9	0,85	0,27	0,58	0,490	0,531	0,396
5	0,9	0,90	-	-	0,508	0,551	0,388
6	0,6	0,62	0,14	0,47	0,417	0,460	0,453
7	1,8	0,76	0,16	0,60	0,569	0,613	0,384
8	0,3	0,35	0,20	0,15	0,287	0,318	0,267
9	0,3	0,40	0,33	0,07	0,577	0,626	0,391
10	0,3	0,39	0,31	0,08	0,536	0,583	0,375
11	0,6	0,47	0,12	0,35	0,363	0,399	0,334
12	0,9	0,13	0,02	0,04	0,342	0,377	0,292
13	0,9	0,20	0,03	0,08	0,356	0,392	0,332
keskiarvo		**0,56**	**0,19**	**0,33**	**0,455**	**0,496**	**0,355**

1) D = Somersin delta (Somers, 1962); G = Goodmanin–Kruskalin gamma (Goodman & Kruskal, 1954);
R = tulomomenttikorrelaatiokerroin (Bravais, 1844; Pearson, 1896)

TAULUKKO 2.11c. Vaikean tehtäväsarjan reliabiliteetin indikaattoreita

Kerroin	painokerroin		
	R	D	G
alfa	0,392	0,628	0,701
theta		0,697	0,757
omega		0,774	0,811

Erottelukyvyltään osiot eivät yllä läheskään koko aineiston tasolle: vaikeassa tehtäväsarjassa esimerkiksi keskimääräinen erottelukyky on D = 0,45 kun se arvioinnin ensimmäisen vaiheen tehtäväsarjoissa oli 0,56 (vrt. edellä Taulukko 4a). Tästä suoraan seuraa se, että vaikean tehtäväsarjan mittarin reliabiliteetit ovat kauttaaltaan matalampia (< 0,82) kuin varsinaisessa arvioinnissa käytetyissä tehtäväsarjoissa (> 0,95). Tämä ei tietenkään ole yllätyksellistä, sillä testiin osallistuneet oli valikoitu niin, ettei heidän välillään ollutkaan juuri eroa, ja testattavia ei ollut tarkoituskaan pystyä erottelemaan toisistaan.

Vaikeaan tehtäväsarjaan osallistuneet eivät juuri jättäneet puuttuvia tietoja, vaan oppilaat yrittivät suoriutua tehtävistä, vaikka eivät olisi osanneetkaan (ks. Taulukko 11b, osiot 12 ja 13). Arvioinnin ensimmäisessä vaiheessa parhaimmin menestyneet oppilaat ovat selkeästi ikätovereitaan parempia tasogeometriaan (osio 2), suunnikkaan piirin (osio 4), funktion muotoon liittyvissä tehtävissä (osio 7) sekä tehtävissä, joissa helpohkoon yhtälönratkaisuun on pitänyt kirjoittaa perusteluja eli lauseke (osio 3). Näiden osalta ero ratkaisuprosenteissa yli 50 prosenttiyksikköä.

Muita arvioinnin luotettavuuteen liittyviä seikkoja

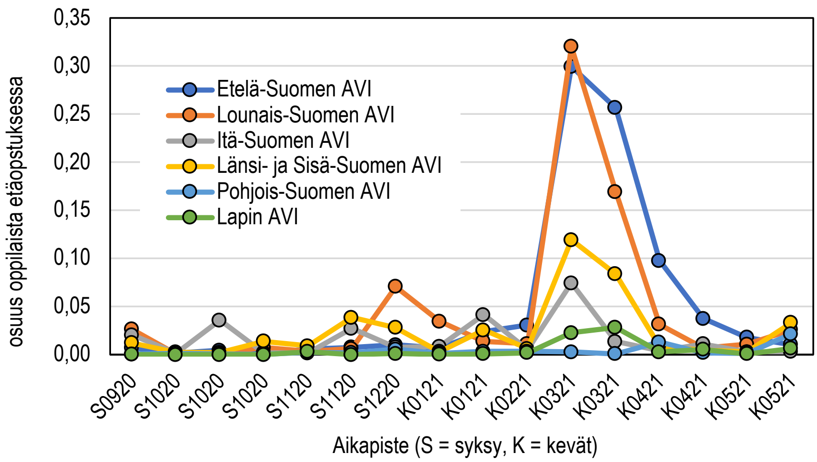
Arvioinnin luotettavuuteen liittyi seikkoja, joilla voi mahdollisesti olla arvioinnin luotettavuutta heikentäviä vaikutuksia. Tarkemmin tässä yhteydessä näistä käsitellään digitaaliseen arviointijärjestelmään, puuttuviin tietoihin ja COVID-19-pandemiaan liittyviä vaikutuksia aineistoon.

Arviointi toteutettiin digitaalisesti ja jotkut koulut raportoivat yhteysongelmista tietoverkoissa. Esimerkiksi Tampereen seudulla tietoliikenteen runkoverkko kaatui juuri arviointiviikolla ja useat koulut ottivat yhteyttä arvioinnin järjestelyihin liittyen. Osa oppilaista pystyi tekemään oman tehtäväsarjansa mobiiliyhteyksillään ja osa kouluista siirsi arvioinnin toteutuksen toiselle päivälle. Yleisesti ottaen arvioinnin suorittamiselle oli varattu viikko, jolloin yksittäisenä päivänä ilmennyt ongelma ei estänyt arvioinnin suorittamista. Joissain kouluissa koulun lähiverkko kuormittui liikaa ja yhteydet olivat hitaita. Tämän kaltaisilla seikoilla oli vaikutuksia muun muassa siihen, kuinka arvioinnissa kyettiin vastaamaan ensimmäisessä osassa olleeseen päässälaskukokonaisuuteen, joka oli aikarajoitettu (15 minuuttia). Jos hitautta tai katkoksia ilmeni juuri päässälaskuvaiheessa, ei tilannetta voinut paikata myöhemmin arviointitehtävien tekemisen aikana, kuten muiden osa-alueiden kohdalla. Joidenkin oppilaiden osalta tilanne johti siihen, että suuri (tai tietty) osa arviointitehtävistä jäi tekemättä ja että aineistossa oli systemaattisesti puuttuvia tietoja tietyissä kouluissa.

Digitaalisen testauksen erityisvaikutuksia tarkennetaan luvussa 2.3 (vaikutuksia vertaistamiseen) ja luvussa 2.4 (puuttuvien tietojen vaikutusta katoon).

Kevään 2021 arviointi tehtiin ”COVID-19-pandemian varjossa” (Metsämuuronen & Nousiainen, 2021). Keväästä 2020 alkaen koettiin koko maailman mittakaavassa poikkeuksellinen COVID-19-pandemia, joka vaikutti koko arviointia edeltävän vuoden opintoihin ja joka oli edelleen ajankohtainen vuonna 2021. Pandemian seurauksena monissa kouluissa jouduttiin siirtymään etäopetukseen. Kansainvälisesti arvioiden siirtymä onnistui Suomessa verraten sujuvasti, ja nopeasti kehittyneiden digivalmiuksien ansiosta koulut ja oppilaitokset olivat etäopetuksen ja siihen liittyvien käytänteiden osalta nopeasti toimivia (Goman ym. 2021, s. 112). Keväällä 2020 lähes kaikki oppilaat olivat etäopetuksessa lukuun ottamatta pienimpiä alakoululaisia ja erityistä tukea saavia oppilaita. Syksyn 2020 ja kevään 2021 aikana oppilaat olivat vaihtelevia jaksoja alue-, koulu- ja yksilökohtaisesti osin etäopetuksessa ja osin lähiopetuksessa.

Aluehallintoviraston (AVI) COVID-19-pandemian vaikutuksia perusopetukseen käsittelevän aineiston mukaan (AVI, 2022[5]) syksyllä 2020 etäopetuksessa olleiden oppilaiden osuus oli varsin pieni eikä osuus vaihdellut systemaattisesti eri puolilla Suomea. Sen sijaan keväällä 2021 etätyöskentely oli selvästi yleisempää pahimmilla COVID-19-leviämisalueilla Etelä- ja Lounais-Suomessa kuin esimerkiksi Pohjois-Suomessa, jossa etäopetusta ei aineiston mukaan ollut juuri lainkaan (Kuvio 8). Huomionarvoista on, että niillä AVI-alueilla, joilla etätyöskentely oli yleisempää (Etelä-, Lounais- ja Länsi- ja Sisä-Suomen AVI-alueilla), myös osaamisen jakaumat ovat selvästi kulmihuippuisia (ks. jakaumamuoto luvussa 4) toisin kuin Pohjois-Suomen AVI-alueella, jossa jakauma on lähes normaalinen.



KUVIO 2.8. Etäopetuksessa olleiden oppilaiden osuudet eri AVI-alueilla syksyllä 2020 ja keväällä 2021

5 Karvi sai Aluehallintovirastolta käyttöönsä etäopetuksessa olleiden oppilaiden osuuksia koskevan kuntakohtaisen aineiston. Aineisto ei sisällä tietoja kaikista kunnista; tiedonkeruuseen osallistuminen oli vapaaehtoista.

COVID-19-pandemian seuraukset ovat tätä kirjoitettaessa vielä epäselviä, mutta on mahdollista, että poikkeusolojen seurauksena on syntynyt niin oppimis-, osaamis-, turvallisuus- kuin hyvinvointivajettakin (ks. tarkemmin Metsämuuronen & Nousiainen, 2021; Metsämuuronen & Seppälä, 2021). Näistä oppimis- ja osaamisvaje voi näkyä heikentyneinä suorituksina ja osaamisen eriytymisenä (ks. Metsämuuronen & Nousiainen, 2021), mutta tämä ei sinänsä vaikuta tutkimuksen luotettavuuteen. Sen sijaan on mahdollista, että oppilaiden kokema turvallisuus- ja hyvinvointivaje on saattanut vaikuttaa oppilaiden testikäyttäytymiseen esimerkiksi kohonneen testiahdistuksen tai helpommin periksi antamisen muodossa. Näistä asioista ei tällä hetkellä kuitenkaan ole tutkittua tietoa saatavilla, jolloin niiden vaikutus jää lähinnä arvailujen asteelle.

Osa oppilaista (6 %) teki testin kotona etätestauksena. Tähän liittyy monia epävarmuustekijöitä, sillä luokkakokeesta poiketen ei tehtäväsarjaan vastattaessa ollut kattavia kontrollimekanismeja esimerkiksi lunttaamisen kontrolloimiseksi. Kotoa käsin arviointiin osallistuneet oppilaat menestyivät arvioinnissa tilastollisesti merkitsevästi paremmin ($p < 0{,}003$) kuin koulussa kontrolloiduissa olosuhteissa osallistuneet oppilaat. Ero oli 13 pisteen luokkaa, mikä ei kuitenkaan ole merkittävän suuri (Cohenin $f = 0{,}03$). Siten näiden oppilaiden suoritukset eivät merkittävästi vääristä kansallista tulosta, joten heidät pidettiin mukana aineistossa.

Yhteenvetoa keskeisten mittareiden luotettavuudesta

Vaikka mittaukseen liittyy testin luonteesta, puuttuvista tiedoista ja COVID-19-pandemiasta johtuvaa epävarmuutta, varsinaisen tiedonkeruun mittarit osoittautuivat kaiken kaikkiaan sekä osuviksi eli valideiksi että erottelukyvyltään riittävän tarkoiksi eli reliaabeleiksi. Mittauksen tarkkuutta voi arvioida mittausvirheen avulla.

Perinteinen mittausvirheen kaava on $S.E.m = \sigma_X \sqrt{1 - REL}$, missä σ_X on keskihajonta ja REL viittaa mittauksen reliabiliteettiin. Kun tiedetään esimerkiksi, että summamuuttujan keskihajonta on $\sigma_X = 113{,}35$ ja reliabiliteetti $REL = 0{,}98$, voidaan arvioida, että yksittäisen oppilaan osaamista arvioidessa kokonaissummassa on keskimäärin (ainakin) ±16 pisteen verran virhettä, kun asteikon keskiarvo on 500 (Taulukko 12). Tämä vastaa alkuperäisessä pistesummassa noin yhden pisteen poikkeamaa suuntaansa. Esimerkiksi jos oppilas sai keskitasoa osoittavan pistemäärän, yleisversion kokonaissummassa 35 pistettä, mittausvirhe huomioiden ”todellinen” pistemäärä olisi ollut 34–36 pistettä.

Reliabiliteetti on periaatteessa yhteydessä myös kansallisen keskiarvon keskivirheeseen:

$S.E.(\bar{X}) = \frac{\sigma_X}{\sqrt{REL \times n}}$.[6] Koska mittauksen reliabiliteetti ei koskaan ole täydellinen $(REL < 1)$, on sillä taipumus lisätä keskiarvon keskivirhettä. Koska reliabiliteetti mittareissa on korkea, ei sillä ole käytännössä vaikutusta aineiston keskivirheeseen; keskiarvossa on ±1 pisteen verran harhaa ilman reliabiliteetin huomioimista tai sen kanssa.

6 Kaavan on johtanut Johtava arviointiasiantuntija Jukka Marjanen (Karvi).

TAULUKKO 2.12. Keskeisten osaamismittareiden reliabiliteetit ja keskivirheet

summa	N	Minimi	Maksimi	keskiarvo	keskihajonta	REL	*S.E.m*	*S.E.*($\bar{X}$)
kokonais-summa	12482	-195	1043	452	113,35	0,98	±16	±1,0
S1	12482	-58	893	444	139,66	0,95	±31	±1,3
S2	12482	-53	941	448	131,97	0,96	±26	±1,2
S3	12482	-23	1007	454	125,65	0,96	±25	±1,1
S4	12482	-25	815	452	138,98	0,95	±31	±1,3
S5	12482	-92	824	456	129,94	0,96	±26	±1,2
S6	12482	-164	1004	458	167,73	0,97	±29	±1,5

2.3 Vertaistamiseen liittyviä erityiskysymyksiä

2.3.1 Vertaistamisen periaatteet

Eri vuosien summapisteiden ja saman mittauskerran eri tehtäväsarjojen pistemäärät on kyettävä yhteismitallistamaan, jotta aineistoja on mielekästä yhdistellä ja vertailla. Tätä kutsutaan vertaistamiseksi. Englanniksi samaan asiaan viitataan termeillä *equating* (mm. Béguin, 2000; Dorans, Moses, & Eignor, 2010; Kolen & Brennan, 2004), *scaling* (nn. AERA, APA, NCME, 2014) tai *linking* (mm. Linn, 1993; Mislevy, 1992). Näistä *linking* on yleistermi, ja *equating* on voimakkain linkitys eri mittareiden välillä. Ajatuksena on luoda linkitys kahden summapistemäärän välille niin, että yksilö olisi voinut tehdä minkä hyvänsä testiversioista (Dorans ym., 2010).

Sopivasti valittujen, eri tehtäväsarjoille yhteisten linkkitehtävien avulla voidaan sarjojen pistemäärät saattaa vertailukelpoisiksi. Lopullisiin tehtäväsarjoihin valittiin 20 linkkitehtävää (16 %) aiempien vuosien tehtäväsarjoista, mikä on riittävä määrä vahvaan vertaistamiseen. Tehtäväsarjojen välillä linkitys oli vieläkin voimakkaampi: erikoisversiot linkittyivät toisiinsa yhteisellä 48 osioisella perustestillä (80–82 % linkkitehtäviä tehtäväsarjasta riippuen), ja yleisversio linkittyi erikoisversioihin 49 tehtävällä (72 %) niin, että valtaosa näistä oli erikoisversioita yhdistävässä perussarjassa ja lisäksi jokaisessa erikoisversiossa oli 1–2 sisältöaluekohtaista linkkitehtävää yleisversioon.

Vertaistamisessa käytettiin *Item Response Theory* (IRT) -mallinnus ja tarkemmin (yksiparametrista) Rasch-mallinnusta (Rasch, 1962; Lord, Novick & Birnbaum, 1968), jossa vertaistaminen tapahtuu osioiden vaikeustasojen perusteella ilman muita parametreja. IRT-mallinnus tuottaa kussakin aineistossa ja kullekin osamittarin pistemäärälle standardipisteen. Näin siis keskimääräinen pistemäärä koko aineistossa saa standardipisteeksi $\theta = 0$ ja tätä matalampi pistemäärä johtaa negatiivisiin standardipisteisiin ($\theta < 0$) ja tätä korkeampi pistemäärä positiivisiin standardipisteisiin ($\theta > 0$). Tässä arvioinnissa vertailu tehtiin vuoden 1998 arviointiin nähden, jolloin eri vuosien tuloksia voidaan verrata toisiinsa vuoden 1998 tasoon suhteutettuna.

Eri vuosien jakaumakeskiarvot ja -hajonnat vaihtelevat hieman. Vertaistamisen välivaiheessa vuoden 1998 aineiston keskiarvoksi muodostui 0,259 ja keskihajonnaksi 0,980. Näitä tietoja käytettiin hyödyksi, kun eri vuosien arvot standardoitiin asteikolle, jonka keskiarvo on 0 ja hajonta 1. Näin siis keskitason suorituksen vuonna 1998 tehnyt oppilas saa pistemääräkseen 0,00 ja jakauman keskihajonta on 1,00 ja muiden vuosien jakaumat suhteutuvat ensimmäisen mittauskerran keskiarvoon ja hajontaan. Tämä pistemäärä muunnetaan ns. 10T-muunnoksella (100 x muuttuja + 500) niin, että keskitasoinen oppilas vuonna 1998 saa arvokseen 500 ja jakauman hajonnaksi tulee 100. Muiden vuosien aineistot (2000, 2002, 2004, 2011, 2012, 2015, 2021) vertaistettiin näihin arvoihin.

Teknisesti vertaistaminen tehtiin niin, että ensin analysoitiin vuoden 1998 aineisto, ja tämän aineiston osioiden vaikeustasoparametriarvot kiinnitettiin. Tämän jälkeen vuosien 2000–2015 tulleet uudet osiot ja niiden vaikeustasoparametrit annettiin olla vapaasti estimoituvia ilman vuoden 2021 aineistoa. Ennen vuoden 2021 analyysia nämäkin parametrit kuitenkin kiinnitettiin, niin, että kaikki linkkitehtävät tuli vertaistetuksi vuoden 1998 tasoon. Lopuksi kaikki vuoden 2021 osioparametrit kiinnitettiin niin, että eri osa-alueiden vertaistetut pistemäärät olivat toisiinsa verrannollisia. Tämä poikkeaa joistain aiemmista analyyseista (ks. esimerkiksi 1. ja 3. luokan aineistot; Ukkola & Metsämuuronen, 2019, 2021, 2023) siinä, että keskiarvoksi ei automaattisesti tule identtistä arvoa (aiemmin 500), vaan tässä arvioinnissa voidaan tarkastella jossain määrin myös sitä, kuinka eri osa-alueilla osaamisen tasot poikkeavat toisistaan. Tällöin vertaaminen tapahtuu suhteessa teoreettiseen ”yleiseen matematiikan osaamiseen”, jota kokonaisjakauma edustaa.

Kaikkien osamittareiden (S1–S6; T10–T20; TE1–TE3; H1–H4) jakaumat vertaistettiin vuoden 1998 tasoon nähden, mikä tuo mahdollisuuden arvioida myös sellaisia summamuuttujia, joiden vertailu ei ehkä ole edes mielekästä. Esimerkiksi kokonaan uutena tavoitteena vuonna 2021 mukaan tullutta ohjelmointia ja algoritmista ajattelua ei mitattu lainkaan vuonna 1998. Tämän osa-alueen pistemäärä kertoo kuitenkin teoriassa, kuinka osaaminen olisi muuttunut, jos myös vuonna 1998 oppilaat olisivat suorittaneet vastaavanlaisia tehtäviä, sikäli kun osa-alueen tehtävät heijastavat myös ”yleistä matematiikan osaamista”.

2.3.2 Puuttuvat tiedot vertaistamisen haasteena

Edellä kuvattiin digitaaliseen testaukseen liittyviä erityiskysymyksiä, joista yksi oli, että puuttuvia tietoja oli enemmän kuin aiempina vuosina. Osittain tätä saattoi selittää digitaalisen testin luonne. Erityinen haaste testaukseen tuli siitä, että *monet oppilaista eivät ilmeisestikään halunneet vastata perustelutyyppisiin tehtäviin*. Jos tehtävässä piti laskea jotain, jossa piti antaa sekä vastaus (1 piste) että sen perustelu eli yleensä varsinainen laskutoimitus (1 piste), monilta oppilailta jäi perustelu antamatta. Tätä asiaa ja siihen liittyviä syitä tarkastellaan lähemmin luvussa 3.

Tässä yhteydessä otetaan esiin asiaan liittyvä ilmeinen haaste: perustelujen puuttumisella oli suora vaikutus siihen, kuinka hyvin tehtäväsarjoihin valitut linkkitehtävät toimivat. Mikäli oppilaiden vastaustapa jossain tehtävässä muuttuu kokeiden välillä oleellisesti, on vaikea perustella, miten osio voisi toimia linkkitehtävänä.

Linkkiosioista neljä osoittautui haasteellisiksi. Näistä kolmessa oppilas sai yhden pisteen oikeasta vastauksesta ja toisen pisteen oikeaan vastaukseen johtaneesta perustelusta. Päätellen siitä, että oikean vastauksen antaneiden (1 pistettä oikein) määrät olivat kasvaneet 6–21 prosenttiyksikköä, mutta kahden pisteen saaneiden osuuksissa oli tapahtunut 27–43 prosenttiyksikön lasku aiempaan nähden, monet oppilaista osasivat kyllä itse laskun, mutteivat osanneet tai vaivautuneet kirjoittamaan testialustalle sen perusteluja (Taulukko 13).

TAULUKKO 2.13. Esimerkki perustelujen puuttumisesta linkkitehtävissä (osuudet vastanneista)

	aiempi arviointi (paperi-kynä-testi)				2021 (digitaalinen testi)				ero		
pistesumma osiossa	0	1	2	3	0	1	2	3	1 pistettä	2 pistettä	3 pistettä
linkkitehtävä1	0,122	0,072	0,806	-	0,328	0,133	0,538	-	-0,061	0,268	-
linkkitehtävä 2	0,265	0,072	0,664	-	0,487	0,225	0,288	-	-0,153	0,376	-
linkkitehtävä 3	0,123	0,163	0,714	-	0,339	0,375	0,286	-	-0,212	0,428	-
linkkitehtävä 4	0,391	0,056	0,060	0,492	0,568	0,222	0,037	0,173	-0,166	0,024	0,319

Nämä neljä osiota poistettiin linkkitehtävistä, mutta ne jätettiin tehtäväsarjoihin yksittäisinä osioina. Vertaistamisen näkökannalta tarkastellen näiden osioiden osioparametreja ei kiinnitetty vuoden 2021 aineistossa, vaan ne otettiin mukaan ikään kuin uusina osioina. Mikäli nämä tehtävät olisi jätetty linkkitehtäviksi, kansallinen osaamisen taso olisi ollut raportoituihin tuloksiin nähden oleellisesti matalampi.

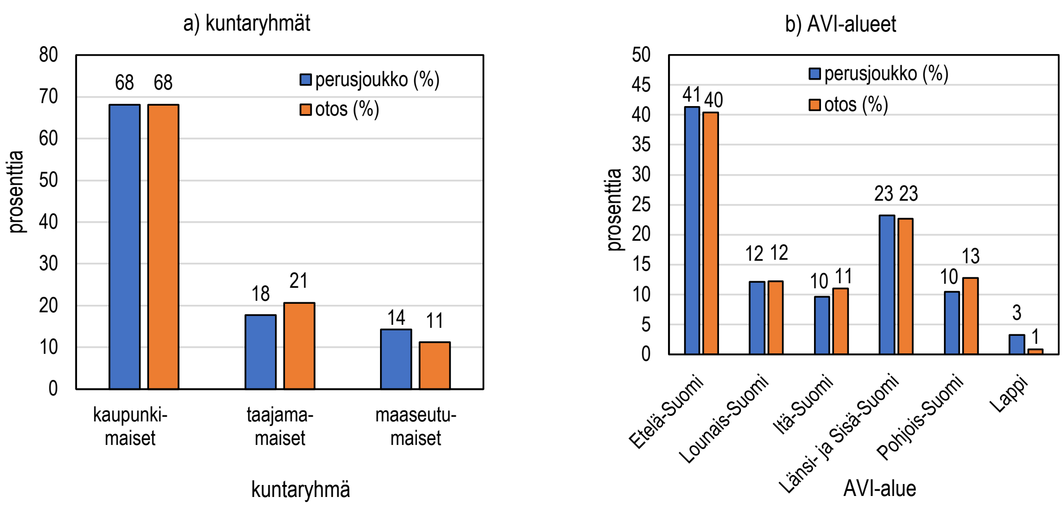
2.4 Otos, kato ja lopullinen aineisto

2.4.1 Otostamisen periaatteet ja otoksen osuvuus

Arviointia varten suunniteltiin kaksi erilaista otosta tai saman otoksen eri osia. Yhtäältä koulut valittiin edustamaan satunnaisesti Suomen kouluja, ja toisaalta eri tehtäväsarjoja valittiin satunnaisesti eri kouluihin eli tehtäväsarjoja vaihdeltiin koulukohtaisesti. Koulujen valinta tehtiin ositetusti kahdessa vaiheessa. Ensin koulutuksen järjestäjistä valittiin edustava otos eri kieliryhmistä (suomenkieliset/ruotsinkieliset koulut; kansainväliset koulut jätettiin otannan ulkopuolelle), AVI-alueilta ja kuntaryhmistä (kaupunkimaiset/taajamamaiset/maaseutumaiset koulut). Toisessa vaiheessa kultakin koulutuksen järjestäjältä valittiin edustava otos kouluja niin, että jos kouluja oli 3 tai vähemmän, järjestäjän kaikki koulut tulivat mukaan, jos kouluja oli 4–9, joka toinen otettiin mukaan, jos kouluja oli 10–20, otokseen valittiin joka kolmas koulu ja jos kouluja on yli 20, mukaan otettiin joka neljäs systemaattisesti ja satunnaisesti poimien koulun koon mukaan järjestetystä listasta. Tällä menettelyllä saatiin mukaan automaattisesti eri kokoisia kouluja kultakin suuremmalta koulutuksen järjestäjältä. Otokseen valikoituneista kouluista kaikki oppilaat, jotka opiskelivat 9. luokalla keväällä 2021, osallistuivat arviointiin. Kaikkiaan koulujen ja oppilaiden määrä oli suurempi kuin oppimistulosarvioinneissa on yleensä.

Tehtäväsarjojen koulukohtaisessa valinnassa käytettiin niin sanottua matriisiotantaa, jota sovellettiin seuraavasti: Tehtäväsarjoista yksi (Ver1) oli yhteinen kaikissa kouluissa ja kuusi muuta, sisältöaluepainotettua tehtäväsarjaa (VerS1–VerS6) annettiin vastattavaksi niin, että kaikissa kouluissa oli *kaksi* erikoisversiota yleisversion lisäksi. Matriisiotannan mukaisesti tehtäväsarjat kuitenkin jaettiin niin, että saman koulutuksen järjestäjän kouluista saatiin tietoa kaikkien tehtäväsarjojen osalta. Jos tehtäväsarjoja ei voitu jakaa järjestäjän koulujen kesken, kuten esimerkiksi erittäin pienten järjestäjien tapauksissa, ne jaettiin samaan ositteeseen kuuluvien toisen (pienen) järjestäjän koulujen kanssa. Näin kaikkia sisältöalueita koskevaa tietoa tuli kaikkialta Suomesta erityyppisistä kouluista.

Lopulliseen otokseen kuului yhteensä 12 482 oppilasta (23 % ikäluokasta) yhteensä 167 koulusta. Aineistossa on lievä yliedustus taajamamaisista kunnista tulleita oppilaita (21 % otoksessa ja 18 % perusjoukossa) ja vastaavasti maaseutumaiset kunnat ovat hieman aliedustettuna (14 % perusjoukossa ja 11 % otoksessa; Kuvio 9a). Maantieteellisesti aineistossa on pientä aliedustusta Lapin aluehallintoviraston (AVI) alueelta ja yliedustusta Pohjois-Suomen AVI-alueelta (Kuvio 9b). Aliedustus Lapin AVI-alueella johtui siitä, että vaikka koulujen määrä otoksessa oli sopiva, koulut osoittautuivat hyvin pieniksi ja näin oppilasmäärä jäi odotettua pienemmäksi. Nämä yli- ja aliedustukset korjataan analyyseissa painokertoimien avulla.



KUVIO 2.9. Otoksen osuvuus

2.4.2 Kato, puuttuvat tiedot ja niiden vaikutus tuloksiin

Ensimmäisen vaiheen tiedonkeruuseen aineistoa saatiin 12 803 oppilaalta. Teoriassa kaikkien koulun yhdeksäsluokkalaisten oppilaiden, mukaan lukien erityisen tuen oppilaiden, oli tarkoitus osallistua arviointiin, mutta rehtoreita oli ohjeistettu jättämään testauksen ulkopuolelle ne, joilla oli esimerkiksi merkittävästi osallistumiseen ja arvioinnissa suoriutumiseen vaikuttava kehitysvamma tai muu vastaava selkeä peruste jäädä arvioinnin ulkopuolelle. Myös COVID-19-pandemia saattoi vaikuttaa siihen, että joitain yksittäisiä oppilaita ei ilmaantunut kouluun arviointipäivänä tai joita ei tavoitettu kotoa etätestausta varten. Näillä yksittäisillä ja satunnaisilla poisjäänneillä ei ole merkittävää vaikutusta kansallisiin tuloksiin.

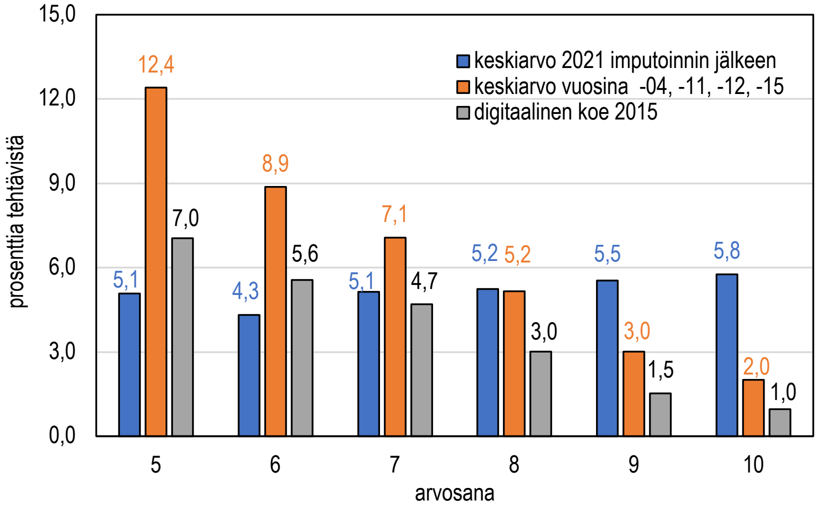
Aineiston alustavan tarkastelun yhteydessä huomattiin, että aineistossa oli oppilastasolla poikkeuksellisen paljon puuttuvia tietoja. Koska tarkoituksena on arvioida, mikä on kansallinen osaamisen taso, mekaanisilla virheillä kuten vastaamatta jättämisellä ja tietoteknisillä häiriöillä on taipumusta tuottaa todellista heikompia tuloksia. Tätä pyrittiin korjaamaan mallintamalla kunkin oppilaan todellinen osaamistaso. Jos mallintaminen ei ollut mahdollista puuttuvien tietojen suuren määrän vuoksi, aineistosta poistettiin ne oppilaat, joilla oli aiemmin näytettyyn osaamiseen—käytännössä viimeisimpään matematiikan arvosanaan—nähden huomattavasti odotettua enemmän puuttuvia tietoja.

Ensimmäisen vaiheen tiedonkeruussa päässälaskutehtäviin vastaaminen oli rajoitettu korkeintaan 15 minuuttia kestäväksi. Kun testaus oli aloitettu, päässälaskuihin käytettävää aikaa ei voinut pysäyttää. Osa oppilaista ei saanut vastattua päässälaskutehtäviin johtuen monesta oppilaasta ja koulusta riippumattomasta tekijästä. Jos oppilas ei ollut vastannut yhteenkään päässälaskutehtävään, mutta tietoja oli muualla tehtäväsarjassa, oletettiin, että syy puuttuviin tietoihin oli mekaaninen. Tällöin systemaattiset puuttuvat tiedot päässälaskuosuudessa korvattiin käyttäen verrokkimenetelmää. Yhtäältä oppilaan osaamisen taso tiedettiin muissa kuin päässälaskutehtävissä ja toisaalta tiedettiin oppilaan viimeisin matematiikan arvosana. Näiden tietojen perusteella puuttuvia tietoja antaneelle oppilaalle etsittiin verrokkioppilas, jolla oli sama pistemäärä ja jonka kouluarvosana oli sama. Puuttuva päässälaskutieto korjattiin noin kahdelta prosentilta oppilaista. Tämän imputoinnin jälkeen tarkasteltiin mahdollisia muissa arvioinnin tehtäväosissa ilmenneitä puuttuvia tietoja.

Kun tutkittiin puuttuvien tietojen määriä, todettiin, että erityisesti oppilailla, joiden viimeinen vahvistettu kouluarvosana oli 9 ja 10, nyt toteutetussa arvioinnissa puuttuvia vastauksia oli oleellisesti enemmän kuin aiemmissa arvioinneissa (Kuvio 10).

Yleisesti ottaen hyvin menestyvät oppilaat eivät juuri jätä vastaamatta tehtäviin, vaan yrittävät loppuun asti löytää ratkaisun tai yrittävät arvata. Kun aiempina vuosina arvosanan 10 saaneet oppilaat jättivät vastaamatta 1–2 prosenttiin tehtävistä, tässä mittauksessa määrä oli lähes kolminkertainen (6 %). Samoin arvosanan 9 saaneilla oppilailla tyhjiä vastauksia oli lähes kaksinkertainen määrä aiempaan nähden. Tämän tulkittiin johtuneen nimenomaan teknisistä ongelmista. Yksittäisten oppilaiden osalta puuttuvia tietoja tietenkin on ollut enemmän kuin keskiarvo kertoo. Jos aivan yksittäiset tapaukset jätetään huomiotta ja tarkastellaan esimerkiksi arvosanana 10 saaneita oppilaita, vuosina 2011–2015 puuttuvia tietoja oli korkeimmillaan 15–20 % tehtävistä. Vuoden

2021 aineistossa luku oli 45 % eli oleellisesti korkeampi. Pääteltiin, että osa näistä on seurausta epäonnistuneista arviointijärjestelyistä. Tästä syystä aineistosta poistettiin oppilaita, joiden todellisesta osaamisesta olisi jäänyt merkittävä epäily. Muutoin puuttuvat tiedot korvattiin tai korvautuivat summamuuttujia muodostettaessa nollalla: oppilas ei ollut kyennyt osoittamaan, että hän olisi osannut suorittaa tehtävän hyväksytysti.



KUVIO 2.10. Puuttuvien tietojen määrät eri vuosien aineistoissa

Aineiston muokkaamisessa päädyttiin seuraaviin periaatteisiin:

- aineistosta poistettiin oppilaat, joiden tulos oli erityisen heikko (0–3 pistettä), jos puuttuvia tietoja oli yli 80 %, mutta oppilaan viimeinen arvosana oli 7–10.
- arvosanan 10 saaneista oppilaista poistettiin ne, joilla oli yli 30 % puuttuvia tietoja.
- arvosanan 9 saaneista poistettiin ne oppilaat, joilla oli yli 40 % puuttuvia tietoja
- arvosanan 8 saaneista poistettiin ne oppilaat, joilla oli yli 50 % puuttuvia tietoja.

Näiden lisäksi aineistosta poistettiin normaaliin tapaan myös ne oppilaat, joilta ei tullut vastauksia lainkaan (puuttuvia tietoja 100 %) ja ne, jotka koulut ilmoittivat poistetuiksi itse arviointitilanteesta esimerkiksi häiriköinnin vuoksi. Kaikkiaan 12 857 Karviin palautuneesta oppilasvastauksesta 128 (1 %) poistettiin em. periaatteiden mukaisesti. Poistamisella ei ole oleellista vaikutusta kansalliseen keskiarvoon, joskin vaikutus on systemaattisesti keskiarvoa kohottava.

2.4.3 Lopullinen aineisto ja sen ominaispiirteitä

Lopullisen aineiston ominaispiirteitä on koottu taulukkoon 14. Otoskouluista 153 oli suomenkielisiä ja 14 ruotsinkielisiä. Vastaavasti oppilaista 11 507 oli suomenkielisiä (92 %) ja 975 ruotsinkielisiä (8 %). Tyypillisesti ruotsinkielistä aineistosta otetaan suurempi otos kuin suhteellinen osuus edellyttäisi, jotta myös ruotsinkielisestä koulutuksesta voitaisiin tehdä tarkempaa analyysia. Lisäksi kokonaisaineistoon kuuluu myös ryhmä oppilaita (n = 347, 3 % kokonaisaineistosta), joilla oli matematiikan osalta yksilöllistetty oppimäärä (HOJKS). Tätä erityisryhmää käsiteltiin analyyseissa erikseen, koska heidän ei edellytetä hallitsevan opetussuunnitelman perusteissa edellytettäviä asioita. Otoksen oppilaista 6 415 (50 %) oli poikia ja 6 407 oli tyttöjä. Seitsemän oppilaan sukupuolitietoa ei saatu selville edes nimen perusteella. Valtaosaan oppilaista saatiin yhdistettyä kansallisesta Koski-tietovarannosta heihin liittyviä keskeisiä taustatietoja kuten erityisopetukseen tai S2-opetukseen osallistumistieto. Vain 320 oppilaan tiedot (3 %) jäivät yhdistymättä Koski-rekisterin tietoihin.

TAULUKKO 2.14. Valittuja aineistoa kuvaavia muuttujia ja niiden jakaumatietoja

Muuttuja		Koko aineisto *n* = 12 482 (*n*)	Koko aineisto *n* = 12 482 (%)	Erityisaineisto (HOJKS) *n* = 347 (%)
Sukupuoli	poika	6195	49,6	63,4
	tyttö	6280	50,3	36,6
Oppilaitoksen kieli	suomi	11507	92,2	92,8
	ruotsi	975	7,8	7,2
Oppimäärä	suomi	10601	85,5	85,6
	ruotsi	922	7,4	7,2
	suomi/ruotsi toisena kielenä	833 + 48	6,7 + 0,4	7,2 + 0
Suuralue	Helsinki-Uusimaa	3467	27,8	17,6
	Etelä-Suomi	2296	18,4	19,6
	Länsi-Suomi	3574	28,6	28
	Pohjois- ja Itä-Suomi	3145	25,2	34,9
AVI (aluehallintovirasto)	Etelä-Suomen AVI	5046	40,4	32,3
	Lounais-Suomen AVI	1529	12,2	15,9
	Itä-Suomen AVI	1379	11	14,7
	Länsi- ja Sisä-Suomen AVI	2831	22,7	18,2
	Pohjois-Suomen AVI	1593	12,8	16,7
Kuntaryhmä	kaupunki	8504	68,1	50,1
	taajama	2582	20,7	29,4
	maaseutu	1396	11,2	20,5

Muuttuja		Koko aineisto *n* = 12 482 (*n*)	Koko aineisto *n* = 12 482 (%)	Erityisaineisto (HOJKS) *n* = 347 (%)
Kotikieli	suomi	9577	76,7	74
	ruotsi	396	3,2	4,3
	jokin muu	368	2,9	2
	suomi ja ruotsi	472	3,8	2,6
	suomi ja muu	892	7,1	7,8
	ruotsi ja muu	62	0,5	0,6
	suomi, ruotsi ja muu	113	0,9	74
	ei tietoa	576	4,6	8,7
S2/SV2-status	Ei	11532	92,4	92,8
	Kyllä	881	7,1	7,2
Kolmiportainen tuki	Ei tuen päätöstä (yleinen tuki)	10493	84,1	6,9
	Tehostettu tuki	1228	9,8	0,9
	Erityinen tuki	499	4	92,2
	HOJKS	0	0	100
Missä suoritti testin	koulussa	11755	94,2	98,8
	etäopetuksessa kotona	727	5,8	1,2
Opiskeleeko erikoisluokalla	ei erikoisluokalla	10193	81,7	82,3
	musiikki	346	2,8	0
	liikunta	380	3,0	0,7
	STEM, mat., luonn., tekn. tai yrittäjyys	209	1,7	6
	ilmaisutaito, viestintä tai taide	224	1,8	0,7
	kieli tai kielikylpy	186	1,5	0
	jopo, pienryhmä tai erityisluokka	117	0,9	10,3
Äidin korkein tutkinto	peruskoulu	216	1,7	3,2
	ammatillinen koulutus	2062	16,5	27
	ylioppilastutkinto	1010	8,1	12,2
	ammattikorkeakoulu	2305	18,5	17,7
	yliopisto	3326	26,6	7,7
	jokin muu	238	1,9	2,6
	ei tiedä	2263	18,1	29,6
Isän korkein tutkinto	peruskoulu	373	3,0	5,5
	ammatillinen koulutus	2866	23,0	29,8
	ylioppilastutkinto	610	4,9	3,2
	ammattikorkeakoulu	2484	19,9	12,3
	yliopisto	2239	17,9	7,8
	jokin muu	257	2,1	2,9
	ei tiedä	2561	20,5	38,5

Arviointiin osallistuivat kaikki valittujen koulujen 9. luokan oppilaat—myös ne, joilla oli erityisen ja tehostetun tuen tarve. Yksittäisiä oppilaita jäi pois rehtorin päätöksellä, mikäli heidän katsottiin olevan kykenemättömiä osallistumaan arviointiin esimerkiksi kehitysvamman vuoksi. Tieto kolmiportaisen tuen tasosta saatiin lähes kaikkien oppilaiden osalta yhdistämällä tietoa Koski-tietovarannosta ja oppilaan taustakyselystä. Tehostettua tukea sai oppilaista 1228 (10 %), erityistä tukea 499 oppilasta (4 %) ja 255 oppilaan osalta tietoa ei ollut saatavilla (2 %). Näiden lisäksi kokonaisaineistoon kuului edellä mainitut 347 oppilasta, joilla oli matematiikan yksilöllistetty oppimäärä (HOJKS).

Koski-tietovarannosta saatujen tietojen perusteella otokseen kuuluvista oppilaista 881 (7 %) opiskeli suomea tai ruotsia toisena kielenä ainakin yhtenä kolmesta viimeisimmästä vuodesta. Toista kieltä eli S2- tai SV2-oppimäärää opiskelevia oppilaita käsitellään tässä raportissa pääosin yhtenä ryhmänä (myöhemmin lyhyemmin S2), koska SV2-oppilaita oli arvioinnissa mukana vain vähän. Kaikki oppilaat tekivät tehtävät koulunsa opetuskielellä eli suomeksi tai ruotsiksi kielitaustastaan riippumatta.

Erikoisluokille osallistuminen oli aineiston perusteella melko harvinaista: 82 % oppilaista ei ollut erikoisluokalla. Matematiikkaan ja luonnontieteisiin, teknologiaan tai yrittäjyyteen liittyvillä erikoisluokilla oli 1,7 % oppilaista, kun liikuntapainotteisilla luokilla oli 3,0 % ja musiikkipainotteisilla luokilla 2,8 % oppilaista.

Tulevissa raporteissa käsitellään perheen sosioekonomista taustaa (SES). Yhtenä indikaattorina on vanhempien koulutustausta. Viidennes aineiston oppilaista ei tiennyt, mikä on oman äidin (18 %) tai isän koulutus (20 %). Koko aineistossa äideistä 45 % ja isistä 38 % on oppilaiden mukaan saanut joko ammattikorkeakoulu- tai yliopistokoulutuksen.

2.5 Analyysimenetelmät ja erityistermit

Lukijalle, jolle tilastolliseen käsittelyyn liittyvät termit eivät ole tuttuja, kuvataan seuraavissa luvuissa lyhyesti analyyseissä käytetyt muuttujat, tilastolliseen päättelyyn liittyvät termit ja käytetyt menetelmät.

2.5.1 Käytetyt muuttujat

Sekä aiemmassa raportissa (Metsämuuronen & Nousiainen, 2021) että tulevissa raporteissa oppimistuloksia, asenteita, emootioita ja kiusaamisen kokemista kuvataan taulukkoon 15 kootuilla osa-alueilla.

TAULUKKO 2.15. Osaamisen ja asenteiden osa-alueet arvioinnissa

Oppimistulokset		Asenteet ja Emootio	Kiusaaminen
Kokonaisosaaminen		**Kokonaisasenne**	**Kokonaiskiusaaminen**
S1	Ajattelun taidot ja menetelmät	AS1. Käsitys itsestä oppiaineen osaajana	
S2	Luvut ja laskutoimitukset	AS2. Oppiaineesta pitäminen	
S3	Algebra	AS3. Oppiaineen koettu hyödyllisyys	
S4	Funktiot		
S5	Geometria	EM1. Positiiviset emootiot	
S6	Tietojen käsittely ja tilastot sekä todennäköisyys	EM2. Negatiiviset emootiot	
T10	päässälasku ja päätelmät		
T11	peruslaskutoimitukset rationaaliluvuilla		
T13	prosenttilaskut		
T14	yhtälön ratkaisu		
T15	funktion tuottaminen ja tulkitseminen		
T16	geometria		
T17	suorakulmainen kolmio ja ympyrä		
T18	pinta-ala ja tilavuudet		
T19	tilastolliset tunnusluvut ja todennäköisyys		
T20	algoritminen ajattelu ja ohjelmointi		

Osaaminen ilmaistaan pääsääntöisesti pistemääränä asteikolla, jossa 9. luokan keskiarvo vuonna 1998 oli 500 ja hajonta 100 (ks. edellä luku 2.3). Jos oppilaan osaaminen vuoden 2021 aineistossa oli 500, hän oli samalla tasolla kuin keskitasoinen oppilas vuonna 1998. Asenteita kuvataan asteikolla 0–4, jossa 0 tarkoittaa selkeää negatiivista asennetta, 2 neutraalia asennetta ja 4 selkeää positiivista asennetta. Kiusaamiseen liittyvä mittari on skaalattu välille 0–4, missä alkuperäisen asteikon mukaisesti 0 tarkoittaa ”ei koskaan” kaikkien kuuden muuttujan suhteen, 1 tarkoittaa keskimäärin ”erittäin harvoin” ja vain jonkin yksittäisen osatekijän muodossa, 2 tarkoittaa keskimäärin ”harvemmin kuin kerran viikossa” usean tai kaikkien osatekijöiden suhteen, 3 tarkoittaa keskimäärin ”noin kerran viikossa” usean tai kaikkien osatekijöiden suhteen ja pistemäärä 4 keskimäärin ”useita kertoja viikossa” kaikkien kuuden osatekijän suhteen.

2.5.2 Käytetyt termit

Raporteissa käytetään tilastolliseen testaukseen liittyvää termiä *tilastollisesti merkitsevä* kuvaamaan sitä, kuinka luotettavasti ryhmien välillä on eroa muissakin kuin otoskouluissa. Otoksessa ryhmien välillä voi olla pieniä eroja, vaikka koko oppilasjoukossa eroa ei olisi. Tämä on siis otoksesta johtuvaa satunnaista vaihtelua. Tilastollisessa testauksessa (kuten esimerkiksi varianssianalyysissa) keskiarvojen eroja testataan sen suhteen, poikkeavatko erot nollasta. Kun tekstissä kerrotaan eron kahden tai useamman ryhmän välillä olevan tilastollisesti merkitsevä, se tarkoittaa, että ero tulisi näkyviin riippumatta siitä, mitkä koulut ja oppilaat sattuivat tulemaan otokseen. Virhepäätelmän riski on hyvin pieni (esimerkiksi korkeintaan 5 prosenttia). Tämän indikaattorina tekstissä käytetään merkintää $p = 0,05$, joka viittaa suoraan 5 prosentin riskiin tehdä virhepäätelmä otoksen perusteella. Vastaavasti tietenkin esimerkiksi merkintä $p = 0,002$ tarkoittaa 0,2 prosentin riskiä tehdä virhepäätelmä ja merkintä $p < 0,001$ sitä, että virhepäätelmän riski jää pienemmäksi kuin 0,1 prosenttia. Joissain kuvissa merkitsevyyttä havainnollistetaan asteriskilla (*): * viittaa arvoon $p < 0,05$ ja *** arvoon $p < 0,001$.

Toinen tilastolliseen testaukseen liittyvä termi on *efektikoko*. Ero ryhmien välillä voi olla tilastollisesti merkitsevä—eli eroa ryhmien välillä on varmasti otoksesta riippumatta—mutta ero ei käytännössä ole merkittävän suurta. *Efektikoko kertoo sen, kuinka suurta ryhmien välinen ero on*. Kun esimerkiksi tyttöjen ja poikien keskiarvot ovat samat, efektikoko on nolla. Toisaalta perinteiset efektikoon mitat kuten Cohenin d ja f lähestyvät nollaa riippumatta keskiarvojen erosta, kun ryhmien väliset otoskoot ovat hyvin erilaisia (ks. Metsämuuronen, 2022g, 2022h).[7] Jos siis esimerkiksi erityistä tai tehostettua tukea saavien oppilaiden osaaminen olisi yleisen tuen piirissä olevien oppilaiden tulosta niin paljon matalampi, että 80 % heistä sijoittuu yleisen tuen oppilaiden keskiarvon alapuolelle, efektikoko on suuri. Tällaisissa tapauksissa perinteiset efektikoon indikaattorit ovat kuitenkin aliarvioita, sillä tehostettua ja erityistä tukea saavien oppilaiden määrät ovat hyvin pieniä yleisen tuen piirissä oleviin oppilaisiin nähden.

Efektikoon mittana raporteissa käytetään pääsääntöisesti Cohenin f-mittaa (Cohen, 1988), koska arvot ovat helposti vertailtavissa eri aineistoissa ja koska niille on olemassa karkeita rajoja kuvaamaan efektikoon pienuutta tai suuruutta. Efektikoko on ”pieni”, kun $f \approx 0,10$ tai pienempi, ”keskisuuri” kun $f \approx 0,20–0,30$ ja ”suuri”, kun $f \approx 0,40$ tai suurempi. Jos efektikoko on keskisuuri tai suuri, tekstissä käytetään termiä ”merkittävä” ja vastaavasti, jos efektikoko on pieni tai olematon, tekstissä saatetaan sanoa: ”ero ei ole merkittävä”. Ero siis voi olla merkitsevä ($p < 0,001$) eli otoksesta riippumatta todellista, mutta ei merkittävän suurta ($f = 0,10$).

Korrelaatioiden ja regressiomallien yhteydessä käytetään edellisten lisäksi termiä selitysaste tai selitysosuus, joka kertoo, kuinka monta prosenttia muuttujat selittävät toistensa vaihtelusta (korrelaatio) tai selitettävän muuttujan vaihtelusta (regressioanalyysi). Kun kaksi muuttujaa on täydellisessä yhteydessä toisiinsa (kuten esimerkiksi osaaminen raakapisteinä ja prosentteina maksimipistemäärästä), korrelaatio muuttujien välillä on $r = 1$. Tällöin toinen muuttuja selittää toisen täydellisesti ja selitysosuus on 100 %. Jos korrelaatio puolestaan on esimerkiksi $r = 0,20$, selitysosuus on $r^2 = 0,2 \times 0,2 = 0,04$ eli muuttujat selittävät toisistaan vain 4 % ja selittymättä jää 96 %. Cohenin d osoittaa erittäin merkittävää yhteyttä muuttujien välillä, kun korrelaatio on noin $r = 0,30$. Useissa todellisen elämän kasvatustieteellisissä yhteyksissä selitysosuudet jäävät melko pieniksi. Tosin Metsämuuronen (2022g) on osoittanut, että tämä voi johtua siitä, että selitysaste on samalla tavalla deflatoitunut kuin edellä kuvattiin korrelaatiokertoimen olevan deflatoitunut. Tällöin todellinen selitysaste voi olla oleellisesti korkeampi kuin perinteisesti on arvioitu.

7 Tämä johtuu siitä, että Cohenin d ja f voidaan esittää korrelaation avulla. Sekä tulomomenttikorrelaatiokerroin Cohenin d:n taustalla että eta-kerroin Cohenin f:n taustalla ovat erittäin herkkä ryhmien otoskoon eroille. Niiden antama estimaatti lähestyy nollaa, kun otoskoot poikkeavat selvästi.

Varianssianalyysin yhteydessä selitysosuuden laskemisessa käytetään yleensä eetan neliötä (*eta-squared*, η^2; Cohen, 1969) tai osittaista eetan neliötä (*partial eta-squared*, η_p^2; Cohen, 1973), kun kyseessä on useita selittäviä tekijöitä (ks. muita vaihtoehtoja esim. Metsämuuronen, 2022g). Raporteissa tämä muutetaan kuitenkin yleensä suoraan efektikooksi ($f = \sqrt{\eta^2/(1-\eta^2)}$; ks. edellä). Perinteisen selitysasteen rinnalla saatetaan antaa vertailutiedoksi myös attenuaatiokorjattu selitysaste (*attenuation-corrected eta-squared*, η_{AC}^2; Metsämuuronen, 2022g). Tämä lasketaan siten, että ensin lasketaan attenuaatiokorjattu eta (η_{AC}): $\eta_{AC} = \frac{\eta_{g|X}^{Obs}}{\eta_{g|X}^{Max}}$, missä $\eta_{g|X}^{Obs}$ on perinteinen havaittu eta ja $\eta_{g|X}^{Max}$ on kyseisessä aineistossa kyseisten muuttujien välille saatavissa oleva maksimaalinen eta. Tämä puolestaan lasketaan niin, että eta lasketaan muuttujien välille niin, että molemmat muuttujat järjestetään suuruusjärjestykseen toisistaan riippumatta. Attenuaatiokorjattu etan neliö on tämä korotettuna toiseen potenssiin eli $\eta_{AC}^2 = \left(\frac{\eta_{g|X}^{Obs}}{\eta_{g|X}^{Max}}\right)^2$.

Regressiomallien yhteydessä kuvataan selitysosuutena yhteiskorrelaatiokertoimen neliö R^2. Tämäkin luku on deflatoitunut (ks. Metsämuuronen, 2022h) ja tällekin saatetaan antaa attenuaatiokorjattu vaihtoehto (R_{AC}^2; Metsämuuronen, 2022b, 2022h), joka lasketaan samalla periaatteella kuin deflaatiokorjattu etan neliö: $R_{AC}^2 = (R_{AC})^2 = \left(\frac{R^{Obs}}{R^{Max}}\right)^2$, missä R^{Obs} on havaittu korrelaatio selitettävän muuttujan ja ennustemuuttujan välillä ja R^{Max} on maksimaalinen korrelaatio, joka saadaan laskemalla toisistaan riippumatta suuruusjärjestyksen laitettujen muuttujien välinen korrelaatio.

2.5.3 Käytetyt menetelmät

Yleensä Karvin raporteissa ilmiöitä kuvataan yksinkertaisilla perustunnusluvuilla kuten keskiarvoilla, keskihajonnoilla ja prosenteilla. Kun ryhmien välisiä eroja kuvataan, käytetään yleensä varianssianalyysin erilaisia muotoja (*general linear model*; GLM). Kullekin oppilaalle on laskettu otannasta johtuva painokerroin. Kun tuloksia yleistetään koskemaan koko perusjoukkoa, yksi oppilas voi aineistossa vastata neljää oppilasta ja toinen oppilas ehkä vain kahta. Painotukset lasketaan otosoppilaiden osuuksina koko oppilasjoukosta. Niinpä oppilas, joka tulee ositteesta, johon kuuluun vain vähän oppilaita (kuten esimerkiksi Lapin AVI-alueen maaseutumaisen kunnan koulusta tullut oppilas) hän edustaa vain pientä joukkoa oppilaita—ehkä itseään. Vastaavasti esimerkiksi Etelä-Suomen AVI-alueen kaupunkimaisen kunnan koulusta tullut oppilas saattaa edustaa viittä tai kuutta oppilasta. Yleensä isoista osoitteista on jo otantavaiheessa tullut valittua useita koulujakin, joten painot vaihtelevat otoksessa välillä 1,7–8,4. Pienet painot liittyvät otoksessa tyypillisesti ruotsinkielisiin kouluihin, joita on otantavaiheessa yliotostettu, jotta näistä oppilasta voidaan sanoa mitään tarkempaa.

Tulokset kuvataan raportissa painotettuina: keskiarvo tai prosentti olisi kuvatun kaltainen koko perusjoukossa eli kaikkien samanikäisten kohortissa eikä vain otoksessa. Jakaumatiedot kuviin on tuotettu SPSS-ohjelmiston *Complex samples* -toiminnolla, joka suoraan tuottaa populaatiofrekvenssit kussakin ryhmässä. Kuvissa suhteelliset osuudet on laskettu tämän jakauman perusteella; näin laskien jakaumatiedot poikkeavat jonkin verran otoksen perusteella laskettavista.

Edellä huomautettiin, että aineiston painotus vaikuttaa keskeisiin tunnuslukuihin. Esimerkiksi kokonaisosaaminen ja -keskihajonta otoksessa ilman painotusta ovat 452 (ka.) ja 113,35 (kh.). Painotuksen jälkeen keskiarvo on 451 ja keskihajonta 227,19. Keskihajonnan poikkeama vaikuttaa merkitsevyystestaukseen oleellisesti. Kun painottamattomana GLM tuottaa tyttöjen ja poikien keskiarvojen eron testauksessa F-arvoksi $F = 1{,}895$ ($p = 0{,}169$), painotettuna F-arvo on $F = 0{,}471$ ($p = 0{,}493$). Kummassakaan tapauksessa ero poikien ja tyttöjen välillä ei ole tilastollisesti merkitsevä, mutta painotuksen kautta tullut keskihajonta saa aikaan sen, että merkitsevyystestauksessa merkitsevyys jaa matalammaksi kuin otoksen kautta laskettuna. Aiemmissa 9. luokan matematiikan arvioinneissa on käytetty yksinomaan otoksen kautta tulleita painottamattomia lukuja. Yleisesti ottaen kuitenkin suuret otoskoot pitävät merkitsevyyttä indikoivat p-arvot tilastollisessa mielessä matalina kummallakin tavalla laskettuna. Vastaavasti mikäli eroa kahden tai useamman ryhmän välillä ei ole, painotetut populaatioarvot antavat ilmiöstä oleellisesti konservatiivisemman arvion, sillä se käyttää laskuissa estimoitujen populaatiokeskiarvojen rinnalla myös estimoituja populaatiovariansseja, jotka voivat olla otosvariansseihin nähden monikertaisia. Nämä eivät ole vertailukelpoisia aiempien tulosten kanssa. Siksi tilastollinen päättely tehdään kuitenkin painottamattomien, otoskeskiarvojen ja -varianssien avulla. Tämä vastaa tilastollisen päättelyn perusajatusta: olisiko ero todellista populaatiossa, jos havaitsisimme nämä keskiarvot ja varianssit otoksessa.

Osassa analyyseja käytetään päätöksentekopuu-analyysia (*Decision Tree Analysis*, DTA). DTA on joukko menetelmiä, joiden avulla analysoidaan laajoja aineistoja ja luokitellaan selittäviä muuttujia (*Independent variables*) kiinnostavan kohdemuuttujan (*Dependent Variable*), kuten esimerkiksi osaamisen kiusaamisen intensiteetin suhteen (ks. tarkemmin Metsämuuronen, 2009b, 2017b). Kyseessä on ns. numeronmurskaustyökalu, joka on erittäin tehokas tilanteissa, joissa ei välttämättä ole olemassa olevaa teoriaa kertomaan, miten selittävät muuttujat pitäisi ryhmitellä, jotta kohdemuuttuja voitaisiin selittää mahdollisimman hyvin. DTA tekee kaikki mahdolliset yksittäisten muuttujien väliset ryhmittelyt ja valitsee niistä tilastollisin perustein parhaan mahdollisen. Muuttaja, jossa tämä tilastollisesti merkitsevin ero syntyy, toimii ns. äitinoodina ja analyysia jatketaan löydetyn muuttujan eri luokissa. Tuloksena on hierarkkinen puu, joka on kyennyt segmentoimaan oppilaat erilaisiin ryhmiin. Joitain esimerkkejä tuloksista esitellään esimerkiksi luvussa 4.

DTA on herkkä muuttujien valinnalle: yhdenkin muuttujan lisääminen tai poistaminen mallista voi muuttaa tulosta oleellisesti. Mallituksella on myös taipumusta yhdistää pienempiä selittävän muuttujan ryhmiä suuremmiksi. Tämä mekanismi periytyy tilastollisen päättelyn ilmeisestä ongelmasta olla herkkä otoskoolle. Tämän raportin analyyseissa käytetään ensisijaisesti DTA:n CHAID-algoritmia (Kass, 1980).

Periaatteessa koulukohtainen aineisto on aina ryvästynyt eli koulun sisäinen vaihtelu on pienempää kuin silloin, jos saman verran oppilaita olisi valittu otokseen täysin satunnaisesti. Tätä korjataan yleisesti monitasomallituksella (ks. esimerkiksi Metsämuuronen, 2009b, 2017b). Suuri otoskoko ja se, että kaikki oppilaat on valittu mukaan otantaan tulleesta koulusta, saa kuitenkin aikaan sen, että erot ryhmien välillä ovat yleisesti ottaen joka tapauksessa tilastollisesti erittäin merkitseviä, eikä monitasomallitus tuo välttämättä aineiston analysoinnissa lisäarvoa oppilastason tarkasteluissa. Monitasomallitusta käytetään ensisijaisesti koulun selitysosuuden määrittämisessä. Tulevassa raportissa esimerkiksi oppilaiden asenteiden tutkimuksessa monitasomallitus on keskeisessä roolissa.

Raportissa keskiarvot ja keskihajonnat kuvataan painotettuina ns. populaatioarvoina. Tämä poikkeaa useimpien aiempien raporttien menettelystä, koska aiemmin aineistoja ei painotettu. Tällä on käytännöllistä merkitystä esimerkiksi keskihajontoihin ja merkitsevyystestaukseen, jota käsiteltiin edellä.

2.6 Yhteenveto

Uudenlainen arviointiformaatti täysin digitaalisena testinä ja tähän liittyvä aiempaa suurempi määrä puuttuvia tietoja ja COVID-19-pandemian aiheuttamat haasteet ovat epäilemättä vaikuttaneet tulosten luotettavuuteen. Näitä kuitenkin on pyritty paikkaamaan aineiston pienellä manipuloinnilla oikeampaan suuntaan: 1 % oppilaista, joilla oli selvästi puutteellisia tietoja, poistettiin aineistosta, ja 2 %:lle oppilaista osa systemaattisesti puuttuvista päässälaskutiedoista korvattiin verrokkimenetelmällä. Osa linkkitehtävistä poistettiin vertailukelpoisemman tuloksen saamiseksi aiempiin mittauksiin nähden, koska vastaustapa oli oleellisesti muuttunut aiempiin mittauksiin nähden. Tämä koski tehtäviä, joissa oppilaan piti antaa omaan laskuunsa perusteluja; monet oppilaat jättivät perusteluja antamatta.

Yhteenvetona voidaan todeta, että kansallisen oppimistulosarvioinnin tarpeisiin kehitetyt mittarit vuoden 2021 matematiikan arvioinnissa ovat yhtäältä *osuvia* kuvaamaan matematiikan osaamista, asenteita, emootioita ja kiusaamisen kokemuksen intensiteettiä ja toisaalta riittävän *tarkkoja* erottelemaan oppilaat toisistaan ja antamaan siis tarkkoja arvioita matematiikan osaamiseen liittyvistä ilmiöistä. Aiempiin matematiikan arviointeihin nähden mittaristo on laajempi ja perusteellisempi. Kun siis osaamista, asenteita, emootioita ja kiusaamista kuvataan raporteissa, tulokset ovat tarkempia kuin ennen. Samoin otoskoko on oleellisesti suurempi kuin aiemmissa mittauksissa—noin kolminkertainen varhaisiin mittauksiin nähden—mikä mahdollistaa sen, että aineistossa on riittävästi myös erikoispopulaatioita kuten erittäin heikosti ja erittäin hyvin menestyviä oppilaita tai maahanmuuttotaustaisia oppilaita. Otoskoon ollessa suurempi näistä erikoispopulaatioista voidaan sanoa jotain tarkempaa kuin aiemmissa arvioinneissa.

2.7 Lähteet

AERA, APA, & NCME (2014). *Standards for Educational and Psychological Testing*. American Educational Research Association, American Psychological Association, & National Council of Measurement in Education.

Anderson, L. W., & Krathwohl, D. R. (toim.) (2000). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Complete edition. Pearson.

Armor, D. (1973). Theta reliability and factor scaling. *Sociological Methodology, 5*, 17–50. https://doi.org/10.2307/270831

AVI (2022). *Perusopetuksen ja varhaiskasvatuksen tilannekuva*. Aluehallintovirasto. https://avi.fi/tietoa-meista/toimintamme/tuotamme-tietoa/perusopetuksen-ja-varhaiskasvatuksen-tilannekuva

Bentler, P. M. (2009). Alpha, dimension-free, and model-based internal consistency reliability. *Psychometrika, 74*(1), 137–143. https://doi.org/10.1007/s11336-008-9100-1

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain*. David McKay Company.

Bravais, A. (1844). Analyse Mathematique. Sur les probabilités des erreurs de situation d'un point. (Mathematical analysis. Of the probabilities of the point errors). *Mémoires présentés par divers savants à l'Académie Royale des Siences de l'Institut de France* (*Memoirs presented by various scholars to the Royal Academy of Sciences of the Institute of France*), *9*, 255–332. https://books.google.fi/books?id=7g_hAQAACAAJ&redir_esc=y

Brown, A. & Croudace, T. (2015). Scoring and estimating score precision using multidimensional IRT. Teoksessa S. P. Reise, & D. A. Revicki (toim.). *Handbook of Item Response Theory Modeling: Applications to Typical Performance Assessment* (ss. 307–333). Routledge/Taylor & Francis Group.

Cheng, Y., Yuan, K.-H., & Liu, C. (2012). Comparison of reliability measures under factor analysis and item response theory. *Educational and Psychological Measurement, 72*(1), 52–67. https://doi.org/10.1177/0013164411407315

Cohen, J. (1969). *Statistical power analysis for the behavioral sciences*. 1. laitos. Academic press.

Cohen, J. (1973). Eta-squared and partial eta-squared in fixed factor ANOVA designs. *Educational and Psychological Measurement, 33*(1), 107–112. https://doi.org/10.1177/001316447303300111

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. 2. laitos. Erlbaum.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297–334. https://doi.org/10.1007/BF02310555

Dorans, N. J., Moses T. P., & Eignor, D. R. (2010). *Principles and practices of test score equating*. ETS RR-10-29. ETS, Princeton, New Jersey. https://onlinelibrary.wiley.com/doi/pdf/10.1002/j.2333-8504.2010.tb02236.x

Falk, C. F., & Savalei, V. (2011). The relationship between unstandardized and standardized alpha, true reliability, and the underlying measurement model. *Journal of Personality Assessment 93*(5), 445–53. https://doi.org/10.1080/00223891.2011.594129

Fennema, E., & Sherman, J. A. (1976). Fennema-Sherman mathematics attitude scales: Instruments designed to measure attitudes toward the learning of mathematics by females and males. *Journal for Research in Mathematics Education, 7*(5), 324–326. https://doi.org/10.2307/748467

Flora, D. B. (2020). Your coefficient alpha is probably wrong, but which coefficient omega is right? A tutorial on using R to obtain better reliability estimates. *Advances in Methods and Practices in Psychological Science, 3*(4), 484–501. https://doi.org/10.1177/2515245920951747

FUNA (2019). *FUNA—Toiminnallisen laskutaidon arviointi*. http://oppimisanalytiikka.fi/funa

Gadermann A. M., Guhn, M., & Zumbo, B. D. (2012) Estimating ordinal reliability for Likert-type and ordinal item response data: A conceptual, empirical, and practical guide. *Practical Assessment, Research, and Evaluation, 17*(3), 1–13. https://doi.org/10.7275/n560-j767

Goodman, L. A., & Kruskal, W. H. (1954). Measures of association for cross classifications. *Journal of the American Statistical Association, 49*(268), 732–764. https://doi.org/10.1080/01621459.1954.10501231

Goman, J., Huusko, M., Isoaho, K., Lehikko, A., Metsämuuronen, J., Rumpu N., Seppälä, H., Venäläinen, S., & Åkerlund, C. (2021). *Poikkeuksellisten opetusjärjestelyjen vaikutukset tasa-arvon ja yhden vertaisuuden toteutumiseen eri koulutusaloilla. Osa III: Kansallisen arvioinnin yhteenveto ja suositukset*. Julkaisut 8:2021. Kansallinen koulutuksen arviointikeskus. https://karvi.fi/wp-content/uploads/2021/04/KARVI_0821.pdf

Heise, D., & Bohrnstedt, G. (1970). Validity, invalidity, and reliability. *Sociological Methodology, 2*, 104–129. https://doi.org/10.2307/270785

Hooper, D., Coughlan, J., Mullen, M. R. (2008). Structural equation modelling. Guidelines for determining model fit. *Electronic Journal of Business Research Methods, 6*(1),53–60. https://academic-publishing.org/index.php/ejbrm/article/download/1224/1187

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

Kaiser, H. F., & Caffrey, J. (1965). Alpha factor analysis. *Psychometrika*, *30*, 1–14. https://doi.org/10.1007/BF02289743

Karvi. (2020). *Koulutuksen arviointisuunnitelma 2020–2023*. Kansallinen koulutuksen arviointikeskus. https://karvi.fi/app/uploads/2020/04/Koulutuksen_arviointisuunnitelma_2020-2023.pdf

Kass, G. (1980). An exploratory technique for investigating large quantities of categorical data. *Applied Statistics, 29*(2), 119–127. https://doi.org/10.2307/2986296

Kim, S., & Feldt, L. S. (2010). The estimation of the IRT reliability coefficient and its lower and upper bounds, with comparisons to CTT reliability statistics. *Asia Pacific Education Review, 11*, 179–188. https://doi.org/10.1007/s12564-009-9062-8

Kolen, M. J., & Brennan, R. L. (2004). *Test equating, linking, and scaling: Methods and practices*. 2. laitos. Springer-Verlag.

Linn, R. L. (1993). Linking results of distinct assessment. *Applied Measurement in Education*, *6*(1), 83–102. http://dx.doi.org/10.1207/s15324818ame0601_5

Livingston, S. A., & Dorans, N. J. (2004). *A graphical approach to item analysis*. (Research Report No. RR-04-10). Educational Testing Service. https://doi.org/10.1002/j.2333-8504.2004.tb01937.x

Lord, F. M. (1952). The relationship of the reliability of multiple-choice test to the distribution of item difficulties. *Psychometrika, 17*(2), 181–194. http://dx.doi.org/10.1007/BF02288781

Lord, F. M. (1958). Some relations between Guttman's principal component scale analysis and other psychometric theory. *Psychometrika*, *23*(4), 291–296. http://dx.doi.org/10.1002/j.2333-8504.1957.tb00073.x

Lord, F. M., Novick, M. R., & Birnbaum, A. (1968). *Statistical theories of mental test scores*. Addison-Wesley.

McDonald, R. P. (1970). Theoretical canonical foundations of principal factor analysis, canonical factor analysis, and alpha factor analysis. *British Journal of Mathematical and Statistical Psychology, 23*, 1–21. http://dx.doi.org/10.1111/j.2044-8317.1970.tb00432.x

McDonald, R. P. (1999). *Test Theory: A Unified Treatment*. Lawrence Erlbaum Associates.

Metsämuuronen, J. (2009a). *Metodit arvioinnin apuna. Perusopetuksen oppimistulosarviointien ja -seurantojen menetelmäratkaisut Opetushallituksessa*. Oppimistulosten arviointi 1/2009. Opetushallitus.

Metsämuuronen, J. (2009b). *Tutkimuksen tekemisen perusteet ihmistieteissä*. 4. laitos. International Methelp Oy.

Metsämuuronen, J. (2012). Challenges of the Fennema–Sherman test in the international comparisons. *International Journal of Psychological Studies, 4*(3), 1–22. http://www.ccsenet.org/journal/index.php/ijps/article/view/16904/12480

Metsämuuronen, J. (2017a). *Oppia ikä kaikki. Matemaattinen osaaminen toisen asteen koulutuksen lopussa 2015.* Julkaisut 1:2017. Kansallinen koulutuksen arviointikeskus. https://karvi.fi/wp-content/uploads/2017/03/KARVI_0117-1.pdf

Metsämuuronen, J. (2017b). *Essentials of Research Methods in Human Sciences*. SAGE Publications, Inc.

Metsämuuronen J (2020a). Somers' *D* as an alternative for the item–test and item–rest correlation coefficients in the educational measurement settings. *International Journal of Educational Methodology, 6*(1), 207–221. https://doi.org/10.12973/ijem.6.1.207

Metsämuuronen, J. (2020b). Dimension-corrected Somers' *D* for the item analysis settings. *International Journal of Educational Methodology, 6*(2), 297–317. https://doi.org/10.12973/ijem.6.2.297

Metsämuuronen J. (2021a). Goodman–Kruskal gamma and dimension-corrected gamma in educational measurement settings. *International Journal of Educational Methodology, 7*(1), 95–118. https://doi.org/10.12973/ijem.7.1.95

Metsämuuronen, J. (2021b). Directional nature of Goodman-Kruskal gamma and some consequences. Identity of Goodman-Kruskal gamma and Somers delta, and their connection to Jonckheere-Terpstra test statistic. *Behaviormetrika, 48*(2), 283–307. http://dx.doi.org/10.1007/s41237-021-00138-8

Metsämuuronen, J. (2022a). Deflation-corrected estimators of reliability. *Frontiers in Psychology, 12*:748672, https://doi.org/10.3389/fpsyg.2021.748672

Metsämuuronen, J. (2022b). Attenuation-corrected reliability and some other MEC-corrected estimators of reliability. *Applied Psychological Measurement, 46*(8). https://doi.org/10.1177/01466216221108131

Metsämuuronen, J. (2022c). Effect of various simultaneous sources of mechanical error in the estimators of correlation causing deflation in reliability. Seeking the best options of correlation for deflation-corrected reliability. *Behaviormetrika, 49*(1), 91–130 https://doi.org/10.1007/s41237-022-00158-y

Metsämuuronen, J. (2022d). Reliability for a score compiled from multiple booklets with equated scores. Preprint. https://www.researchgate.net/publication/358849481_Reliability_for_a_score_compiled_from_multiple_booklets_with_equated_scores

Metsämuuronen, J. (2022e). How to obtain the most error-free estimate of reliability? Eight sources of underestimation of reliability. *Practical Assessment, Research, and Evaluation, PARE, 27*(1), Art. 10. https://doi.org/10.7275/7nkb-j673

Metsämuuronen, J. (2022f). Typology of deflation-corrected estimators of reliability. *Frontiers in Psychology, 13*:891959. http://dx.doi.org/10.3389/fpsyg.2022.891959

Metsämuuronen, J. (2022g). Artificial systematic attenuation in eta squared and some related consequences. Attenuation-corrected eta and eta squared, negative values of eta, and their relation to Pearson correlation. *Behaviormetrika,* https://doi.org/10.1007/s41237-022-00162-2

Metsämuuronen, J. (2022h). Directional nature of the product-moment correlation coefficient and some consequences. *Frontiers in Psychology*, *13*:988660. https://doi.org/10.3389/fpsyg.2022.988660

Metsämuuronen, J., Hermonen, A., Nousiainen, S. & Laakso, M.-J. (2023). Digitaaliseen testaamiseen liittyviä erityiskysymyksiä kansallisen oppimistulosarvioinnin näkökannalta. Teoksessa J. Metsämuuronen & S. Nousiainen (toim.), *Matematiikkaa COVID-19-pandemian varjossa II. Menetelmälliset ratkaisut matematiikan 9. luokan arvioinnissa keväällä 2021* (ss. 83–126). Julkaisut 5:2023. Kansallinen koulutuksen arviointikeskus.

Metsämuuronen, J. & Nousiainen, S. (2021). *Matematiikkaa COVID-19-pandemian varjossa. Matematiikan osaaminen 9. luokan lopussa keväällä 2021.* Julkaisut 27:2021. Kansallinen koulutuksen arviointikeskus. https://karvi.fi/wp-content/uploads/2021/12/KARVI_2721.pdf

Metsämuuronen, J. & Räsänen, P. (2018). Cognitive–linguistic and constructivist mnemonic triggers in teaching based on Jerome Bruner's thinking. *Frontiers in Psychology, 9*:2543. https://doi.org/10.3389/fpsyg.2018.02543

Metsämuuronen, J. & Salonen, V. (2017). *Matematiikan osaamisen piirteitä ammatillisessa koulutuksessa 2015 ja pitkän ajan muutoksia*. Julkaisut 2:2017. Kansallinen koulutuksen arviointikeskus. https://karvi.fi/wp-content/uploads/2017/03/KARVI_0217.pdf

Metsämuuronen, J. & Seppälä, H. (2021). *COVID-19-pandemia, osaamisvaje ja osaamisen eriytyminen*. Policy Brief 1:2021. Kansallinen koulutuksen arviointikeskus. https://karvi.fi/wp-content/uploads/2021/12/KARVI_Policy_brief_0121.pdf

Metsämuuronen, J., & Suomilammi, M. (2023). Kolmen kansallisen populaation keskeiset erottelevat piirteet sekä heikkojen ja parempien oppilaiden osaamisen rajapintatarkastelua. Teoksessa J. Metsämuuronen & S. Nousiainen (toim.), *Matematiikkaa COVID-19-pandemian varjossa II. Menetelmälliset ratkaisut matematiikan 9. luokan arvioinnissa keväällä 2021* (ss. 127–172). Julkaisut 5:2023. Kansallinen koulutuksen arviointikeskus.

Metsämuuronen, J. & Tuohilampi, L. (2017). *Matematiikan osaamisen piirteitä lukiokoulutuksen lopussa 2015*. Julkaisut 3:2017. Kansallinen koulutuksen arviointikeskus. https://karvi.fi/wp-content/uploads/2017/03/KARVI_0317.pdf

Metsämuuronen, J. & Ukkola, A. (2019). *Alkumittauksen menetelmällisiä ratkaisuja*. Julkaisut 18:2019. Kansallinen koulutuksen arviointikeskus. https://karvi.fi/app/uploads/2019/08/KARVI_1819.pdf

Milanzi, E., Molenberghs, G., Alonso, A., Verbeke, G., & De Boeck, P. (2015). Reliability measures in item response theory: Manifest versus latent correlation functions. *British Journal of Mathematical & Statistical Psychology, 68*(1), 43–64. http://dx.doi.org/10.1111/bmsp.12033

Mislevy, R. J. (1992). *Linking educational assessments: Concepts, issues, methods, and prospects*. ETS Policy Information Center.

Moses, T. (2017). A review of developments and applications in item analysis. Teoksessa R. Bennett & M. von Davier (toim.), *Advancing human assessment. The methodological, psychological and policy contributions of ETS* (ss. 19–46). Educational Testing Service. Springer Open. https://doi.org/10.1007/978-3-319-58689-2_2

OECD (2019), *PISA 2018 Results (Volume III): What School Life Means for Students' Lives*. PISA, OECD Publishing. https://doi.org/10.1787/acd78851-en

OPH (2014). *Perusopetuksen opetussuunnitelman perusteet 2014*. Määräykset ja ohjeet 2014:96. Opetushallitus. https://www.oph.fi/sites/default/files/documents/perusopetuksen_opetussuunnitelman_perusteet_2014.pdf

OPH (2020). *Perusopetuksen päättöarvioinnin kriteerit*. Opetushallituksen määräys OPH-5042-2020. Opetushallitus. https://www.oph.fi/sites/default/files/documents/Perusopetuksen%20p%C3%A4%C3%A4tt%C3%B6arvioinnin%20kriteerit%2031.12.2020.pdf

Pearson, K. (1896). VII. Mathematical contributions to the theory of evolution. III. Regression, heredity and panmixia. Philosophical *Transactions of the Royal Society of London, 187*, 253–318. https://doi.org/10.1098/rsta.1896.0007

Pearson, K. (1900). I. Mathematical contributions to the theory of evolution. VII. On the correlation of characters not quantitatively measurable. *Philosophical Transactions of the Royal Society A. Mathematical, Physical and Engineering Sciences, 195*(262–273), 1–47. https://doi.org/10.1098/rsta.1900.0022

Pearson, K. (1913). On the measurement of the influence of "broad categories" on correlation. *Biometrika*, 9(1–2), 116–139. https://doi.org/10.1093/biomet/9.1-2.116

Pekrun, R. (2017). Emotion and Achievement During Adolescence. *Child Development Perspectives, 11*(3) 215–221. https://doi.org/10.1111/cdep.12237

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Studies in Mathematic Psychology I. Danmarks Pædagogishe Institut. Nielsen & Lydiche.

Räsänen, P., Aunio, P., Laine, A., Hakkarainen, A., Väisänen, E., Finell, J., Rajala, T., Laakso, M.-J., & Korhonen, J. (2021). Effects of gender on basic numerical and arithmetic skills: Pilot data from 3rd to 9th grade for a large-scale online dyscalculia screener. *Frontiers in Education*, 6:683672. https://doi.org/10.3389/feduc.2021.683672

Raykov, T., & Marcoulides, G. A. (2017). Thanks coefficient alpha, we still need you! *Educational and Psychological Measurement, 79*(1), 200–210. http://dx.doi.org/10.1177/0013164417725127

Raykov, T., West, B. T., & Traynor, A. (2015). Evaluation of coefficient alpha for multiple component measuring instruments in complex sample designs. *Structural Equation Modeling, 22*(3), 429–438. http://dx.doi.org/10.1080/10705511.2014.936081

Salonen, V. (2023). Tunteiden mittaaminen matematiikan arvioinnissa—Tunnemittari, uskomukset ja kontrolli–arvo-teoria. Teoksessa J. Metsämuuronen & S. Nousiainen (toim.), *Matematiikkaa COVID-19-pandemian varjossa II. Menetelmälliset ratkaisut matematiikan 9. luokan arvioinnissa keväällä 2021* (ss. 173–189). Julkaisut 5:2023. Kansallinen koulutuksen arviointikeskus.

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research*, *99*(6), 323–338. http://dx.doi.org/10.3200/JOER.99.6.323-3

Schumacker, R. E., & Lomax, R. G. (2004). *Beginner's guide to structural equation modeling*. 2. laitos. Lawrence Erlbaum Associates.

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika, 74*(1), 107–120. https://doi.org/10.1007/s11336-008-9101-0

Sijtsma, K., & Pfadt, J.M. (2021). Part II: On the Use, the Misuse, and the Very Limited Usefulness of Cronbach's Alpha: Discussing Lower Bounds and Correlated Errors. *Psychometrika 86*, 843–860. https://doi.org/10.1007/s11336-021-09789-8

Somers, R. H. (1962). A new asymmetric measure of correlation for ordinal variables. *American Sociological Review, 27*(6), 799–811. http://dx.doi.org/10.2307/2090408

Steiger, J. H. (2007). Understanding the limitations of global fit assessment in structural equation modeling. *Personality and Individual differences, 42*(5), 893–898. https://doi.org/10.1016/j.paid.2006.09.017

Ukkola, A. & Kivistö, A. (toim.) (2023). *Matematiikan viitekehys kansallisen koulutuksen arviointikeskuksen oppimistulosarviointiin vuosiluokilla 1–9*. Julkaisut 6:2023. Kansallinen koulutuksen arviointikeskus.

Ukkola, A. & Metsämuuronen, J. (2019). *Alkumittaus—Matematiikan ja äidinkielen ja kirjallisuuden osaaminen ensimmäisen luokan alussa*. Julkaisut 17:2019. Kansallinen koulutuksen arviointikeskus. https://karvi.fi/wp-content/uploads/2019/07/KARVI_1719.pdf.

Ukkola, A., & Metsämuuronen, J. (2021). *Matematiikan ja äidinkielen ja kirjallisuuden osaaminen kolmannen luokan alussa*. Julkaisut 20:2021. Kansallinen koulutuksen arviointikeskus. https://karvi.fi/wp-content/uploads/2021/08/KARVI_2021.pdf

Ukkola, A., & Metsämuuronen J. (2023). *Matematiikan ja äidinkielen taidot alkuopetuksen aikana. Perusopetuksen oppimistulosten pitkittäisarviointi 2018–2020*. Julkaisut 1:2023. Kansallinen koulutuksen arviointikeskus. https://karvi.fi/wp-content/uploads/2023/01/KARVI_0123.pdf

Zumbo, B. D., Gadermann, A. M., & Zeisser, C. (2007). Ordinal versions of coefficients alpha and theta for Likert rating scales. *Journal of Modern Applied Statistical Methods, 6*(1), 21–29. http://dx.doi.org/10.22237/jmasm/1177992180

LIITE 1: Reliabiliteetin laskemiseen liittyviä teoreettisia yksityiskohtia

Metsämuuronen (2022d) keskustelee reliabiliteettikertoimista, jotka mahdollisimman hyvin heijastaisivat todellista reliabiliteettia tilanteessa, jossa summa muodostetaan usesta mittariversiosta. Artikkelista tähän otetaan keskeisiä seikkoja. Teoreettisempi keskustelu on perusteltu kahdesta syystä. Ensinnäkin tiedemaailmassa ei ole julkaistu sellaista yleisesti hyväksyttyä reliabiliteettikerrointa, jonka avulla voisi laskea estimaatin usean mittariversion yhdistelmänä syntyneen summamuuttujan reliabiliteetille. Kertoimia kyllä on sekä moniulotteisen faktorirakenteen kautta muodostetuille summamuuttujille kuten moniulotteinen omega (ks. esimerkiksi Flora, 2020) että moniulotteisen IRT-mallituksen yhteydessä (mm. Brown & Croudace, 2015; Cheng, Yuan & Liu, 2012; Kim & Feldt, 2010; McDonald,1999; Milanzi et al., 2015). Näistä yksikään ei kuitenkaan ole yleisessä käytössä tai edes yleisesti tunnettu tai ne eivät suoraan sovellu tilanteeseen, jossa testattavat vastaavat vain pieneen osaan koko testipatterin tehtävistä. Toiseksi viimeaikaiset tutkimukset traditionaalisten reliabiliteettikertoimien antamista huomattavista aliarvioista ovat huomion arvoista (ks. esimerkiksi Metsämuuronen, 2022a, 2022b, 2022c, 2022e, 2022f). Metsämuuronen (2022d) esittelee kaksi varteen otettavaa vaihtoehtoa yhdistelmämuuttujan reliabiliteetin laskemiseksi, ja niiden taustaa esitellään tässä lyhyesti.

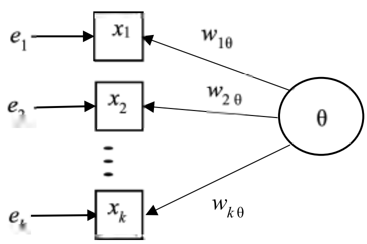
Mittausmalli

Kaikissa psykometrisissa testeissä ja koulukokeissa—niin myös Karvin arviointimittareissa—on aina mittausvirhettä ja tämän virheen suuruuden arviointi on oleellinen osa mittausta. Jos yksilöön liittyvä mittausvirhe (*S.E.m*) on suuri, mittari on epätarkka ja tällöin myös arviointituloksiin liittyvä virhe on suuri. Vastaavasti jos mittausvirhettä on vain vähän, mittari on tarkempi, ja arviointituloksiin liittyvä virhe on pienempää. Epätarkalla mittarilla esimerkiksi pistemäärän 50 saanut oppilas olisi voinut eri mittauskerralla saada 45 pistettä tai 55 pistettä. Mittarin epätarkkuudesta seuraa, että mittari ei pysty erottelemaan toisistaan niitä oppilaita, jotka saivat 45 tai 55 pistettä ja niin osaamisen kokonaisarviointi on epävarmaa. Vastaavasti tarkemmalla mittarilla tämä ns. ”tosi” pistemäärä voi vaihdella välillä 49–51 tai jopa vähemmän, jolloin kansallista osaamista koskea arviointi on selvästi uskottavampaa ja luotettavampaa.[8] Tämän kaikille mittauksille tyypillisen mittausvirheen suuruuden arvioinnissa käytetään ns. reliabiliteettikertoimia, joista tunnetuin ja käytetyin on Cronbachin alfa tai alfa-kerroin (Cronbach, 1951), lyhyemmin alfa. Alfa-kertoimessa on kuitenkin laskentatavasta johtuva rakenteellinen ongelma, jonka vuoksi se antaa käytännössä aina aliarvion mittarin reliabiliteetista; mittari on aina tarkempi kuin mitä alfa indikoi (ks. keskustelua ja kirjallisuutta esimerkiksi Metsämuuronen, 2022a).[9]

8 Muistetaan kuitenkin, että keskiarvoja ja niiden luotettavuutta tarkasteltaessa suuri otoskoko pitää huolen siitä, että keskiarvon keskivirhe jää pieneksi. Keskiarvon keskivirhe ei siis ole riippuvainen yksittäisen oppilaan virheestä.

9 On esitetty, että kun osioihin liittyvät mittausvirheet korreloivat, tällä on se vaikutus, että alfan arvot ovat korkeammat kuin ”tosi” reliabiliteetti (näin mm. Sijtsma, 2009, Sijstma & Pfadt, 2021). Tämä yliarvio on osin teoreettista, käytännössä pientä ja se koskee vain mallinnusvirhettä eikä sen vaikutus ole merkittävä reliabiliteetin estimoinnissa. Tätä suurempi ja radikaalimpi virhe vastakkaiseen suuntaan syntyy osion ja korrelaation välisen korrelaation estimoinnista ja siihen liittyvästä mekaanisesta virheestä, attenuaatiosta tai deflaatiosta, joka alentaa alfan arvoja radikaalisti riippumatta siitä, ovat muuttujat korreloituneita vai eivät. Tästä keskustelevat ja havainnollistavat mm. Gadermann, Guhn, & Zumbo (2012), Metsämuuronen (mm. 2022a, 2022b, 2022f) ja Zumbo, Gadermann, & Zeisser (2007).

Reliabiliteetin suuruus ei ole koskaan varma, vaan se riippuu monesta tekijästä—Metsämuuronen (2022e) kuvaa kahdeksan seikkaa, jotka vaikuttavat reliabiliteetin arvioituun tasoon.[10] Reliabiliteetin tasoa estimoidaan[11] erilaisilla keinoilla. Yleisesti ottaen reliabiliteetin laskemisen keinoja on kuitenkin kehitetty vain tilanteisiin, joissa ilmiötä mitataan yhden tehtäväsarjan avulla (joka voi tosin olla moniulotteinen). Perinteiset reliabiliteettikertoimet alfa, theta, omega ja rho eivät toimi tilanteessa, jossa aineisto koostuu eri tehtäväsarjoista ja ne linkitetty toisiinsa yksittäisillä osioilla. Yksittäisen tehtäväsarjan osalta ajattelemme asiaa visuaalisesti kuvion 1 muodossa (ks. esimerkiksi Metsämuuronen 2022a, 2022b).



KUVIO 1. Yleinen yhden faktorin mittausmalli, jossa virheet eivät korreloi

Kuviossa k viittaa osioiden määrään mittarissa ja e_i viittaa satunnaisvirheeseen yksittäisessä osiossa (ja joita yhteen laskemalla saadaan koko testiin liittyvä satunnaisvirhe $E = \sum_{i=1}^{k} e_i$). Vastaavasti x_i viittaa osioiden havaittuihin pistemääriin (ja joita yhteen laskemalla saadaan osioiden pistemäärien summa eli summamuuttuja $X = \sum_{i=1}^{k} x_i$). Latentti ominaisuus θ, joka voi erilaisissa

10 Simulaation mukaan reliabiliteetin aliarviota ennustavat ainakin seuraavat seikat: (1) mittarin kannalta väärin valittu mittausmalli, (2) (oikein) valitun mittausmalliperheen sisällä tehottoman reliabiliteettiestimaattorin valinta, (3) summamuuttujan (X) muodostamisen tehottomuus latentin ominaisuuden θ ilmentymäksi, (4) osioiden epä-optimaaliset ominaisuudet suhteessa estimaattoriin eli käytännössä yksittäisten osioiden äärimmäinen vaikeustaso, (5) mittausmallissa käytetty tehoton painokerroin, eli deflaatiolle herkkä korrelaatiokerroin (wi), joka yhdistää summamuuttujan osioiden havaittuihin arvoihin, (6) pieni otoskoko, (7) testin äärimmäinen vaikeustaso eli kaikkien osioiden äärimmäinen vaikeustaso ja (8) vähäinen vaihtelu summamuuttujassa. Deflaatiokorjatuissa estimaattoreissa nämä kaikki saatetaan saada huomioiduiksi (ks. Metsämuuronen, 2022e).

11 Reliabiliteetin yhteydessä sen suuruutta ”estimoidaan” otoksen perusteella eli arvioidaan, mikä sen arvo olisi populaatiossa. Estimointi tehdään ”estimaattorilla”, joka antaa ”estimaatin”. Esimerkiksi laajasti käytetty alfa-kerroin (ks. kaava 9) on yksi estimaattoreista ja kertoimen tuottama arvo—esimerkiksi α = 0.90—on yksi mahdollisista estimaateista reliabiliteetin arvoksi. Alfan tuottama estimaatti on aina aliarvio, ja tätä tehokkaammat estimaattorit kuten theta, omega tai rho antavat arvoja, joilla on taipumusta olla lähempänä populaatioarvoa kuin alfa. Siksi monet tutkijat ovat vakavasti ehdottaneet, että alfaa ei tulisi käyttää (ks. kirjallisuutta esimerkiksi Metsämuuronen, 2022a, 2022e). Keskustelu ei kuitenkaan ole lainkaan ohi, sillä monet tutkijat ajattelevat, että kun alfan oletukset täyttyvät, se sopii hyvin yhdeksi arvioksi reliabiliteetin alarajasta (näin mm. Bentler, 2009; Falk & Savalei, 2011; Metsämuuronen, 2017b; Raykov & Marcoulides, 2017; Raykov, West, & Traynor, 2014). Tässä artikkelissa käytetään pääsääntöisesti thetan ja omegan kaavoja niiden paremman käyttäytymisen vuoksi ja perinteistä alfaa käytetään vain vertailutietona sen ilmeisen deflaatiotaipumuksen vuoksi.

mittauksissa ja malleissa konkretisoitua monissa erilaisissa muodoissa kuten faktoripisteenä, yksinkertaisena raakasummana tai—kuten tässä arvioinnissa—IRT-mallituksen kautta syntyneenä vertaistettuna summamuuttujana, viittaa taustalla olevaan ominaisuuteen (tässä arvioinnissa ”matematiikan osaaminen”). Painokerroin w_i on käytännössä korrelaatiokerroin osion ja summan välillä joko faktori- tai pääkomponenttipisteen muodossa tai suoraan korrelaatiokertoimena (esimerkiksi R, D ja G). Latentti ominaisuus θ linkittyy siis havaittuun osion pistemäärään x_i painokertoimen w_i avulla. Kaavan muodossa kuvion 1 mittausmalli esitetään seuraavasti: $\sum_{i=1}^{k} x_i = \sum_{i=1}^{k} w_{i\theta}\theta + \sum_{i=1}^{k} e_i$, mikä vastaa perinteistä ajatusta, että havaittu pistemäärä (X) on ”tosi”pistemäärän (T) ja virheen (E) summa: $X = T + E$.

Malli tuottaa väistämättä aina virhettä—emme pysty ennustamaan oppilaan ”tosi” pistemäärää täydellisesti—ja tämän virheen suuruutta kuvaa em. kaavassa symboli $\sum_{i=1}^{k} e_i$ joka viittaa virhekomponenttien summaan. Olettaen, että sekä osiot että latenttimuuttujat ovat standardoituja muuttujia, tämän virhe-elementin varianssi on $VAR(\sum_{i=1}^{k} e_i) = \sum_{i=1}^{k}\left(1 - w_{i\theta}^2\right)$, ja sitä tarvitaan reliabiliteetin laskuissa. Kaava kertoo suoraan, että mitä korkeampia korrelaatioita osioiden ja summamuuttujan välillä on, sitä pienempää on mallin tuottama virhe(varianssi).

Perinteisiä reliabiliteetin estimaattoreita: Theta ja Omega

Eräs tunnetuista reliabiliteettikertoimista, joka perustuu suoraan kuvion 1 malliin, on thetakerroin (Kaiser & Caffrey, 1965; Lord, 1958, ks. myös Armor 1972):

$$\rho_{TH} = \frac{k}{k-1}\left(1 - \frac{1}{\sum_{i=1}^{k} w_{i\theta}^2}\right), \qquad (1)$$

jossa $w_{i\theta}^2$ viittaa perinteisesti pääkomponenttianalyysin yhteydessä tulleisiin pääkomponenttilatauksiin eli korrelaatioon osion ja pääkomponenttipisteen välillä. Toinen tähän malliin perustuva, tunnettu reliabiliteettikerroin on omega (Heise & Bohrnstedt, 1970; McDonald, 1970), joka tunnetaan myös nimellä McDonaldin omega:

$$\rho_{\omega} = \frac{\left(\sum_{i=1}^{k} w_{i\theta}\right)^2}{\left(\sum_{i=1}^{k} w_{i\theta}\right)^2 + \sum_{i=1}^{k}\left(1 - w_{i\theta}^2\right)}, \qquad (2)$$

jossa $w_{i\theta}^2$ perinteisesti viittaa faktorianalyysin kautta tulleisiin faktorilatauksiin eli korrelaatioon osion ja faktoripisteen välillä. Näitä kaavoja käytetään myös deflaatiokorjatuissa kertoimissa, joissa w vaihdetaan toiseksi kertoimeksi.

Reliabiliteetin estimoinnin haasteita kun summa koostetaan monesta testiversiosta

Edellä todettiin, että yleensä reliabiliteetin estimoinnin menetelmät eivät sovellu yhdistelmämuuttujan reliabiliteetin arviointiin. Metsämuuronen (2022d) esittelee tämän tyyppisten summamuuttujien tapaukseen kaksi perusvaihtoehtoa: yhtäältä mittariversioiden painotuksen kautta saatava estimaatti ja toisaalta hypoteettisen summan kautta saatava reliabiliteetin estimaatti. Nämä esitellään tässä lyhyesti ja niitä käytettiin edellä mittareiden reliabiliteettien estimoinnissa.

Jos vertaistettujen pisteiden kautta syntyvän yhdistelmämuuttujan reliabiliteettia arvioidaan mittariversioiden reliabiliteettien keskiarvojen kautta, tällöin oletetaan, että yhdistelmämuuttujaan liittyvä mittausvirhe syntyy yksittäisissä mittariversioissa ja että mittarit mittaavat samaa latenttia piirrettä riittävän samankaltaisesti. Tämähän ei tietenkään pidä paikkaansa tilanteessa, jossa eri tehtäväsarjoissa mitataan hyvin erilaisia asioita.[12] Oletetaan, että meillä olisi m kappaletta erillisiä mittariversioita samasta latentista piirteestä—tässä arvioinnissa $m = 7$. Jokaisessa yksittäisessä mittauksessa syntyy hieman virhettä, ja kun pistemäärät vertaistetaan, voidaan ajatella, että yhdistelmämuuttujan virheen suuruus on tämän virheen keskiarvo. Tämäkään ei välttämättä pidä paikkansa, sillä eri testisarjoilla on summan suhteen erilainen, mutta tuntematon painoarvo. Kun edellä kuvattu yksinkertainen malli yleistetään usean tehtäväsarjan tapaukseen, malli näyttää kuvion 2 kaltaiselta (Metsämuuronen, 2022d). Nyt, sen sijaan, että meillä olisi vain *yksi* taustalla oleva latentti muuttuja θ (”matematiikan osaaminen”), näitä onkin *useita* ($\theta_1 - \theta_m$, ”matematiikan osaaminen tehtäväsarjassa 1”, ”… sarjassa 2”, …), yksi kutakin tehtäväsarjaa kohden.

Tähän malliin voidaan liittää tapa ajatella yhteismuuttujan reliabiliteettia siten, että yhdistelmämuuttujan reliabiliteetti syntyy osamittareiden reliabiliteettien keskiarvoina, ja tähän malliin liittyvä theta-kerroin saa muodon

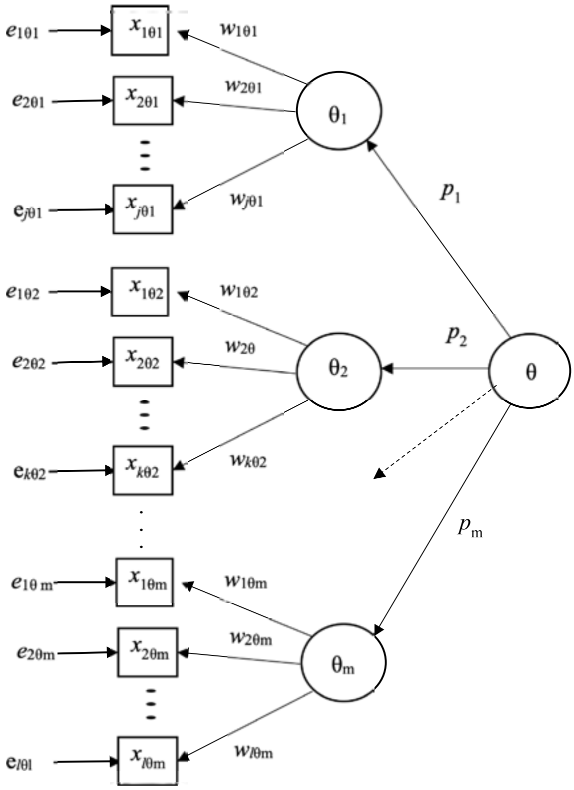
$$\rho_{TH\_wi\theta j} = \frac{1}{m}\sum_{j=1}^{m}\left(\frac{k}{k-1}\left(1-\frac{1}{\sum_{i=1}^{k} w_{i\theta j}^{2}}\right)\right) \tag{3}$$

ja omega-kerroin muodon

$$\rho_{\omega\_wi\theta j} = \frac{1}{m}\sum_{j=1}^{m}\left(\frac{\left(\sum_{i=1}^{k} w_{i\theta j}\right)^{2}}{\left(\sum_{i=1}^{k} w_{i\theta j}\right)^{2} + \sum_{i=1}^{k}\left(1-w_{i\theta j}^{2}\right)}\right), \tag{4}$$

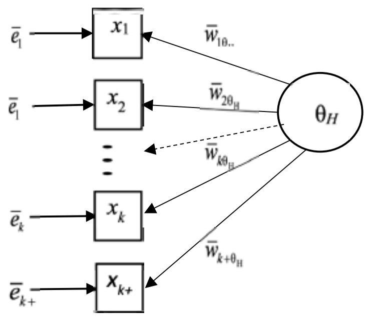
missä symboli $w_{i\theta j}$ viittaa siihen, että saatu arvo riippuu valitusta korrelaatiokertoimesta (w), yksittäisistä osioista (i), valitusta summamuuttujan muodosta (θ) ja että laskussa käytetään useita mittariversioita (j). Painokerrointa p_i pohditaan myöhemmin.

12 Toisaalta näistä eri testiversioista joka tapauksessa muodostetaan yksi yhteinen summamuuttuja, jonka ajatellaan mittaavan riittävällä tarkkuudella jotain yhteistä, yli testiversioiden menevää osaamista. Muutoinhan testattavien vertailu ei ole lähtökohtaisestikaan perusteltua.



KUVIO 2. Yleinen yhden latentin muuttujan mittausmalli, kun käytössä on useita mittariversioita ja summapisteet on vertaistettu

Toinen tapa ajatella yhdistetyn summamuuttujan reliabiliteettia on, että mittausvirhe syntyy yksittäisten osioiden virheen kautta. Tällöin malli visualisoituu kuvion 3 mukaiseksi:



KUVIO 3. Yhden faktorin mittausmalli, jossa kaikki osiot kaikista tehtäväversioista muodostavat hypoteettisen summamuuttujan

Tässä mallissa ajatellaan, että eri tehtäväsarjoissa olleet yksittäiset osiot (joita on $k+$ kappaletta eli enemmän kuin mitä pisimmässä mittarissa on) muodostavat hypoteettisen summamuuttujan (θ_H) ja että yksittäisiin osioihin liittyvä virhe lasketaan linkkitehtävien osalta osiokohtaisen virheen keskiarvoina yli kaikkien tehtäväsarjojen ($\bar{e}_i$). Tällaisten mallien ongelmana on erilaisia asioita sisältävien mittarisarjojen myötä (jatkuvasti) lisääntyvä harhaisuus. Tämän keskivirheen laskemisessa käytetään linkkitehtävien osalta osio-summa-korrelaatioiden keskiarvoa ($\bar{w}_i$). Mallin muodossa asia näyttää seuraavalta:

$$\sum_{i=1}^{k+} x_i = \sum_{i=1}^{k+} \bar{w}_i \theta_H + \sum_{i=1}^{k+} \bar{e}_i \quad (5)$$

Kuten edellä, myös tässä kokonaismittausvirhe lasketaan yksittäisten painokerrointen avulla:

$$Var\left(\sum_{i=1}^{k+} \bar{e}_i\right) = \sum_{i=1}^{k+} \left(1 - \bar{w}_i^2\right) \quad (6)$$

Tässä tapauksessa theta-kerroin saa muodon

$$\rho_{TH\_H} = \frac{k^+}{k^+ - 1}\left(1 - \frac{1}{\sum_{i=1}^{k} \bar{w}_i^2}\right) \quad (7)$$

ja omega-kerroin muodon

$$\rho_{\omega H} = \frac{\left(\sum_{i=1}^{k+} \bar{w}_{i\theta}\right)^2}{\left(\sum_{i=1}^{k+} \bar{w}_{i\theta}\right)^2 + \sum_{i=1}^{k+}\left(1 - \bar{w}_{i\theta}^2\right)}, \quad (8)$$

missä $k+$ viittaa kaikkein yksittäisten, eri tehtäväsarjoissa käytettyjen osioiden määrään. Tämän mallin etu on se, että se vastaa hyvin vertaistamisen perusideaa (ks. luku 2.3): jos oppilaat olisivat tehneet yhden yhteisen tehtäväsarjan seitsemän sijaan, se olisi ollut juuri tämänkaltainen, juuri näistä osioista muodostunut testi. Mallin puutteena on se, että lopullinen yhdistelmämuuttuja ei kuitenkaan synny yksittäisten osioiden kautta. Siksi reliabiliteettikerroin kuvaa hypoteettisen, teoreettisen summamuuttujan reliabiliteettia.

Lisähaaste: perinteisen estimaattoreiden antama deflatoitunut estimaatti

Lisähaaste yhdistelmämuuttujan reliabiliteetin suuruuden arviointiin tulee siitä, että oppimistulosarviointien yhteydessä perinteiset reliabiliteettikertoimet kuten alfa-, theta-, omega- ja rho-kerroin ovat yleensä antavat selkeitä aliarvioita reliabiliteetin suuruudesta (ks. esimerkkejä Gadermann ym., 2012; Metsämuuronen, 2022a, 2022b, 2022e, 2022f). Toisin sanoen estimaatit ovat deflatoituneita. Deflatoituminen on suurinta erityisesti silloin, kun tehtäväsarjassa on mukana helppoja ja vaikeita tehtäviä kuten osaamismittareissa tyypillisesti on.

Aliarvio on mekaaninen ja johtuu siitä, että kertoimiin sisäänrakennettu osio-summa-korrelaatio (*Rit*)—joka teknisesti on tulomomenttikorrelaatiokerroin (R; Pearson, 1896) ja jota joskus nimitetään Pearsonin korrelaatioksi—aliarvioi yhteyden suuruutta aina, kun muuttujien asteikot poikkeavat toisistaan (kuten osion ja summamuuttujan asteikot aina ovat) ja erityisesti, kun osion vaikeustaso on äärimmäinen (helpoilla ja vaikeilla osioilla). Koska *R* on deflatoitunut (ks. simulaatio esimerkiksi Metsämuuronen, 2020a, 2020b, 2021a, 2022c) myös reliabiliteetin estimaatti on deflatoitunut, sillä sama R:n laskentalogiikka sisältyy (kaikkiin) perinteisiin reliabiliteettikertoimiin (ks. Metsämuuronen, 2022a, 2022b, 2022e, 2022f).

Simulaatioiden perusteella Metsämuuronen (2022c) suositteleekin käyttämään estimaattoreissa *R*:n sijaan jotain muuta, deflaatiolle vähemmän altista kerrointa kuten polykorinen korrelaatio (R_{PC}; Pearson, 1900, 1913), r-bireg ja r-polyreg korrelaatio (R_{REG}; Livinstone & Dorans, 2004, Moses, 2017), Somersin delta (*D*; Somers, 1962), Goodman–Kruskal gamma (*G*; Goodman & Kruskal, 1954), dimensiokorjattu *D* (D_2; Metsämuuronen, 2020b, 2021a), dimensiokorjattu *G* (G_2; Metsämuuronen, 2021a), attenuaatiokorjattu *R* (R_{AC}; Metsämuuronen, 2022d, 2022c) tai attenuaatiokorjattu eta (E_{AC}; Metsämuuronen, 2022d).

Deflaatiokorjattuja reliabiliteetin estimaattoreita

Perinteisistä kertoimista alfa antaa aina konservatiivisen arvion reliabiliteetin suuruudesta, theta ja omega aina tätä korkeamman arvion ja rho (maximal reliability) kaikkein korkeimman eli liberaaleimman arvion. Tämä on tendenssinä myös deflaatiokorjatuissa estimaattoreissa: deflaatiokorjattu alfa antaa matalamman arvion kuin deflaatiokorjattu theta, omega tai rho. Deflaatiokorjaukseen vaikuttaa myös se, käytetäänkö korjauksessa konservatiivista vai liberaalimpaa korrelaatiokerrointa. Kokonaismittarin reliabiliteetin arvioinnissa käytettiin konservatiivisempia reliabiliteetin estimaattoreita (alfa, theta, omega) ja sekä konservatiivista että liberaalia korrelaatiokerrointa (ks. taulukot 4a ja 4b).

Deflaatiokorjaus tehdään korvaamalla kertoimiin sisältyvä ja deflaatiota aiheuttava Pearsonin tulomomenttikorrelaatiokerroin (*R*) deflaation näkökannalta paremmin toimivilla korrelaatiokertoimilla. Tässä yhteydessä näistä paremmin toimivista korrelaatiokertoimista käytetään Somersin deltaa (*D*; Somers, 1962) ja Goodmanin–Kruskalin gammaa (*G*; Goodman & Kruskal, 1954). Näistä *D* antaa korrelaatiosta lähes aina konservatiivisempi arvion kuin *G* (ks. poikkeuksia Metsämuuronen, 2021b), mikä johtaa siihen, että *D*:n avulla tehty deflaatiokorjaus tuottaa (lähes) aina matalamman estimaatin reliabiliteetista kuin *G*:n avulla tehty korjaus. Molemmat kuitenkin esitellään, jotta estimaateille saadaan uskottavampi vaihteluväli. Taulukossa 4a esimerkiksi merkintä ”*alfaG*” viittaa deflaatiokorjattuun alfa-kertoimeen, jossa *R* on korvattu *G*:llä. Vastaavalla tavalla ”*thetaD*” viittaa deflaatiokorjattuun theta-kertoimeen, jossa pääkomponenttilataus on korvattu *D*:llä.

Tässä yhteydessä deflaatiokorjatuissa kertoimissa käytössä ovat *D* ja *G*; molempia käytetään sen estimoimiseksi, millä todennäköisyydellä havainnot summamuuttujassa ja osiossa ovat samassa järjestyksessä populaatiossa. Näistä *D* antaa yhteydestä aina konservatiivisemman arvion kuin *G*, koska *D*:n antamaan todennäköisyyteen vaikuttavat myös ne tapaukset, joissa pareissa olevien samojen arvojen vuoksi niiden välinen järjestys ei ole tiedossa (ks. tarkemmin Metsämuuronen, 2021b). *G* jättää nämä tapaukset pois laskuoperaatiosta. Kertoimien tuottamien estimaattien välinen ero ei kuitenkaan ole merkittävä tässä aineistossa; se saattaa näkyä vasta kolmannessa desimaalissa. Koska perinteinen alfa lasketaan aina raakapisteille, mutta raportissa käsitellään IRT-mallituksella saatuja raakapisteiden logistisia muunnoksia (EAP-pisteet; ks. luku 2.3.4), sekä traditionaalinen alfa että deflaatiokorjatut reliabiliteetit on laskettu taulukkolaskentaohjelmistolla käyttäen seuraavia kaavoja. Alfakertoimeen perustuvat kertoimet ovat

$$\rho_{\alpha\_R} = \frac{k}{k-1}\left(1 - \sum_{i=1}^{k} \sigma_i^2 \Big/ \left(\sum_{i=1}^{k} \sigma_i R\right)^2\right), \tag{9}$$

$$\rho_{\alpha\_D} = \frac{k}{k-1}\left(1 - \sum_{i=1}^{k} \sigma_i^2 \Big/ \left(\sum_{i=1}^{k} \sigma_i D\right)^2\right), \tag{10}$$

ja

$$\rho_{\alpha\_G} = \frac{k}{k-1}\left(1 - \sum_{i=1}^{k} \sigma_i^2 \Big/ \left(\sum_{i=1}^{k} \sigma_i G\right)^2\right), \tag{11}$$

(Metsämuuronen, 2021b, 2021c), missä *k* on osioiden määrä, σ_i^2 viittaa osioiden variansseihin ja *R*, *D*, ja *G* viittaavat korrelaatioon osion *i* ja summamuuttujan (θ) välillä, joka on konkretisoitunut IRT-mallituksen kautta syntyneenä theta-arvona. Kaava (9) tuottaa traditionaalisen alfan ja kaavat (10) ja (11) deflaatiokorjatun alfan. Merkintä ρ_{α_G} (= ”*alfaG*”) viittaa siis siihen, että kyseessä on deflaatiokorjattu reliabiliteettikerroin (ρ), jossa pohjana on käytetty alfakerrointa (α) ja jossa deflaatiota aiheuttava korrelaatiokerroin *R* on korvattu vähemmän deflatoituneilla kertoimella *G*.

Kaikki deflaatiokorjatut estimaattorit ovat laskentaperusteeltaan tavanomaisesta (lähestymistavasta) poikkeavia, sillä perinteinen alfa lasketaan korrelaationa *raakapisteiden* ja osioiden välillä (eikä IRT-summan ja osioiden välillä) ja ”*alfaD*” ja ”*alfaG*” ovat alfan uusia muotoja, jotka periaatteessa myös tulisi laskea raakasummien perusteella. Deflaatiokorjattujen reliabiliteetin estimaattoreiden paradigmassa yleisen mittausmallin näkökannalta tämä ei kuitenkaan ole oleellista, sillä muutoinkin estimaattorit alfa, theta, omega ja rho on irrotettu niiden alkuperäisestä yhteydestä ja ajatellaan, että ne edustavat *itsenäisinä kaavoina* erilaisia tapoja estimoida mittarin reliabiliteetti. Tämä ei kuitenkaan ole ongelmatonta ja asiasta keskustelee mm. Metsämuuronen (2022a, 2022b).

Vastaavasti theta-kertoimeen perustuvat deflaatiokorjatut reliabiliteetit voidaan laskea seuraavilla kaavoilla (”*thetaD*”, ”*thetaG*”, ”*omegaD*” ja ”*omegaG*”):

$$\rho_{TH\_D} = \frac{k}{k-1}\left(1 - \frac{1}{\sum_{i=1}^{k} D^2}\right) \tag{12}$$

ja

$$\rho_{TH\_G} = \frac{k}{k-1}\left(1 - \frac{1}{\sum_{i=1}^{k} G^2}\right) \tag{13}$$

ja omega-kertoimeen perustuvat reliabiliteetit kaavoilla

$$\rho_{\omega\_D} = \frac{\left(\sum_{i=1}^{k} D\right)^2}{\left(\sum_{i=1}^{k} D\right)^2 + \sum_{i=1}^{k}\left(1 - D^2\right)}$$

ja (14)

$$\rho_{\omega\_G} = \frac{\left(\sum_{i=1}^{k} G\right)^2}{\left(\sum_{i=1}^{k} G\right)^2 + \sum_{i=1}^{k}\left(1 - G^2\right)} \quad (15)$$

(Metsämuuronen, 2022a), jotka kaikki ovat deflaatiokorjattuja kertoimia. Perinteisiä theta- ja omegakertoimia ei aineistossa ole mielekästä laskea, koska ne estimoivat reliabiliteetin hyvin erilaiselle summalle (pääkomponenttipiste- ja faktoripistemuuttujalle) kuin mikä aineistossa on käytössä (IRT-analyysin tuottama theta).

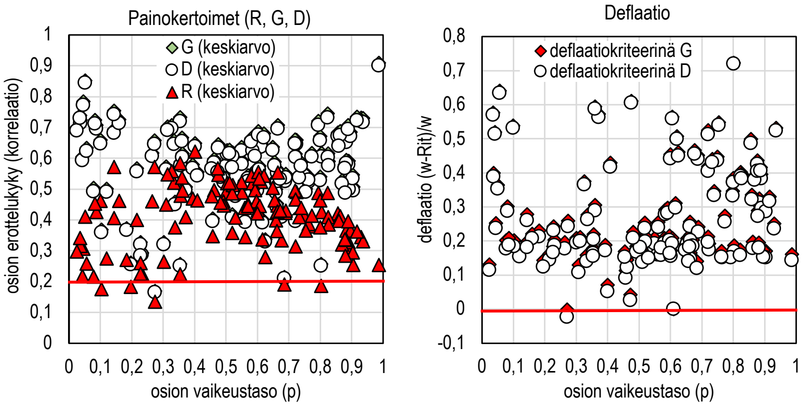
Kaikkiaan voidaan todeta, että tehtäväsarjojen ja lopullisen yhdistetyn aineiston summamuuttujien erottelukyky eli reliabiliteetti on korkea: eri mittaversioiden reliabiliteettien keskiarvojen avulla kokonaisosaamisen reliabiliteetti on traditionaalisella alfakertoimella (kaava 9) arvioituna tasolla 0,91–0,92 ja deflaatiokorjatulla alfalla (kaavat 10 ja 11) arvioituna tasolla 0,95–0,96. Hieman tehokkaammilla kertoimilla (theta ja omega; kaavat 4 ja 5) deflaatiokorjaus johtaa hieman korkeampiin arvoihin 0,96–0,97. Jos reliabiliteettia arvioidaan hypoteettisen summan kautta (kaavat 7 ja 8), reliabiliteetit näyttävät olevan hieman korkeampia: kaikki kertoimet antavat yhdenmukaisesti estimaatin 0,98.

Metodisena yksityiskohtana huomattakoon, että aineistossa deflaation korjaustarve ei ole kovin suurta verrattuna esimerkiksi koulunsa aloittavien oppilaiden aineistossa esiintyneeseen tilanteeseen (ks. Metsämuuronen & Ukkola, 2019; ks. myös uudelleen analysoituna Metsämuuronen, 2022a, 2022b, 2022d), jossa erittäin helpon osatestin reliabiliteetti oli deflatoitunut 0,60 yksikköä eli 70 %.[13]

Toinen metodinen yksityiskohta liittyy erottelukyvyn deflaatioon yksittäisissä osioissa. Yleisesti ottaen *R* aliarvioi yhteyden suuruutta, jos muuttujien asteikot eivät ole identtiset. Aliarvio voi olla merkittävä, jos osion vaikeustaso on äärimmäinen eli sekä erittäin helpoilla että erittäin vaikeilla osioilla. Näiden seikkojen suhteen *D* ja *G* ovat kertoimina vakaampia (ks. simulaatiot Metsämuuronen, 2021a, 2022c). Tämä konkretisoituu aineistossa siinä, että valtaosa niistä osioista, joissa ilmenee yli 40 prosentin suuruinen deflaatio[14] (aineistossa 25 % osioista), liittyi tapauksiin, joissa keskiosaaminen oli joko alle $p = 0,2$ eli vaikeilla osioilla tai joissa se oli yli $p = 0,70$ eli helpoilla osioilla (ks. kuvio 4).

13 Koulutulokkaiden lähtötason selvittämisen yhteydessä (Ukkola & Metsämuuronen, 2019) eräs summamuuttujista sisälsi erittäin helppoja harjoitustehtäviä (ks. Metsämuuronen & Ukkola, 2019). Ko. testissä käytettävän kielen hallintaa kuvaavassa harjoitustehtävässä traditionaalinen alfakerroin antoi reliabiliteetille arvoksi 0,26, mikä viittaisi erittäin matalaan erottelukykyyn. Koska oli ilmeistä, että testissä heikoimmin menestyneet oppilaat olivat heikoimpia myös kokonaispisteiden suhteen, reliabiliteetille laskettiin myös deflaatiokorjattu arvo käyttäen Somersin delta -kerrointa. Tämän korjatun alfakertoimen arvo oli 0.86. Reliabiliteetti oli siis deflatoitunut noin 70 % (=(0.86–0.26)/0.86).

14 Deflaation suuruus esimerkiksi deltaan nähden lasketaan kaavalla: (D – R)/D×100%. Jos siis esimerkiksi D = 0,6 ja R = 0,2, deflaation suuruus on (0,6–0,2)/0,6 = 0,667 eli 66,7 %.



KUVIO 4. Deflaatio yksittäisissä osioissa eri versioissa kokonaisuutena (estimaattien keskiarvo)

Kolmas metodinen yksityiskohta havainnollistuu myös kuviossa 4. Tiedetään, että *D* ja *G* arvioivat useimmissa tapauksissa osio-summa-korrelaation olevan korkeampi kuin mitä *R* sen arvioi olevan. Huomataan kuitenkin, että tämä arvio ei aina ole *korkea*; myös *D* ja *G* tunnistavat *R*:n tapaan heikommin erottelevat osiot. Samoin *ero* ei aina ole suuri *R*:n ja *D*:n ja *G*:n antamien arvojen välillä; myös osa erittäin helpoista ja vaikeista osioista osoittautuu erottelukyvyltään heikommiksi.

3

Arvioinnin tekniseen toteutukseen liittyviä erityiskysymyksiä

Jari Metsämuuronen (Karvi)

Aleksi Hermonen (Turun yliopisto, Oppimisanalytiikan tutkimusinstituutti)

Saara Nousiainen (Karvi)

Mikko-Jussi Laakso (Turun yliopisto, Oppimisanalytiikan tutkimusinstituutti)

3

- Vuoden 2021 arviointi oli ensimmäinen kokonaan digitaalisesti toteutettu 9. luokan matematiikan arviointi.
- Oppilaiden testikäyttäytyminen poikkesi joissain tehtävissä selvästi paperi-kynä-testeistä. Siksi osa tehtävistä, jotka oli tarkoitettu linkittämään testejä toisiinsa, muutettiin ei-linkkitehtäviksi.
- Pieni osa oppilaista teki testin etätestauksena ja heillä tulokset olivat hieman parempia kuin kouluolosuhteissa testatuilla oppilailla. Etätestauksessa ei kyetty kontrolloimaan sitä, vastasiko oppilas rehellisesti itsenäisesti testitehtäviin. Tällä ei kuitenkaan ole merkitystä kansallisen tuloksen kannalta.
- Osassa kouluja tietoverkkoyhteydet olivat haasteellisia. Osalle oppilaista kuultuna testattavat päässälaskut jouduttiin mallittamaan, koska yhteys varsinaiseen testiin ei toiminut.

3.1 Digitaalisen testaamisen lyhyt historia kansallisella tasolla

Jo reilun 10 vuoden aikana on osaamisen testaamisessa siirrytty enenevässä määrin digitaalisesti suoritettavaan testaamiseen. Vaikka eräät kansainvälisesti kuuluisat kielitestit, kuten TOEFL (*Test of English as a Foreign Language*) on tehty digitaalisesti jo vuodesta 2006 lähtien (ks. TOEFL, 2007) ja IELST (*International English Language Test System*) pian tämän jälkeen (ks. IELTS, 2007), ensimmäinen laajamittainen kansainvälinen tutkimus, jossa testejä hallinnoitiin täysin tietokonepohjaisesti oli aikuisten lukutaidon kansainvälisen arvioinnin ohjelma (*Programme for International Assessment of Adult Competencies*, PIAAC; ks. OECD, 2013; 2016) vuonna 2011. Tätä ennen Tanskassa, Islannissa ja Etelä-Koreassa kokeiltiin luonnontieteiden (*science*) osa-alueen testaamista digitaalisesti vuoden 2006 PISA-arvioinnissa (*Programme of International Student Assessment*) (PISA, 2010). Osittain tietokoneperustainen oli myös PISA 2012:n matematiikan ja lukutaitotehtävien testaus (PISA, 2015a), kun taas tieto- ja viestintätekniikan testaaminen oli täysin digitaalinen (PISA, 2015b). Vuoden 2018 PISA-testauksessa valtaosa osallistujamaista teki testin digitaalisesti (PISA, 2018). Vuoden 2019 kansainvälisen matematiikan ja luonnontieteiden hankkeen tiedonkeruu (*Trends in International Mathematics and Science Study)* suunniteltiin alun perin digitaaliseksi (Mullis & Martin, 2017), mutta vain noin puolet osallistujamaista teki digitaalisen testin (Mullis ym., 2020).

Eräissä digitaalisen testauksen pioneerimaissa, kuten Yhdysvalloissa, on käytetty automaattisia pisteitysjärjestelmiä jo kymmenien vuosien ajan kansallisissa testeissä (Mazzeo, 2016). Jo vuonna 2001 näihin liittyviä tutkimustuloksia raportoitiin Euroopan arviointiyhdistyksen (AEA-E) kokouksessa. Toisena hyvänä esimerkkinä Alankomaat on perustanut (täysin) uuden digitaalisen arviointijärjestelmän (van Boxel, 2016). On helppo nähdä, että edellä kuvatuilla kansainvälisillä vaikuttajilla on ollut ja tulee olemaan suuri vaikutus niihin tapoihin, joilla tulevaisuudessa tietoa kerätään kansallisella tasolla. Näin myös Suomessa ja muissa Pohjoismaissa ja Baltiassa. Laajemmasta skandinaavisesta ja baltialaisesta näkökulmasta näyttää siltä, että Ruotsi saattaa olla jonkin verran edellä muita maita tietokonepohjaisen testauksen ja siihen liittyvän kehittämisen osalta. Tähän viittaa Uumajan yliopiston soveltavan kasvatustieteen laitoksen pitkä koulutusmittaushistoria; tutkijat ovat julkaisseet mittausta ja arviointia koskevia raportteja ja artikkeleita vuodesta 1994 alkaen.[15] Viroon perustettiin digitaalinen arviointipankki 2010-luvulla (Voronina, 2016)[16] ja Osloon perustetiin uusi ja aktiivinen osaamisen mittauskeskus (CEMO) osana Oslon yliopiston kasvatustieteiden tiedekuntaa vuonna 2020.[17] Skandinavian maat ovat kuitenkin monin tavoin vasta ottamassa käyttöön digitaalista testausta kansallisella tasolla. Esimerkiksi Suomessa ylioppilastutkinto saatiin täysin digitalisoitua vuosien kehittämisen jälkeen vasta vuonna 2019 (YTL, 2022).

Suomessa on melko pitkä historia kansallisesta oppimistulosten systemaattisesta arviointijärjestelmästä vuodesta 1997 lähtien (ks. Jakku-Sihvonen, 1997, OPH, 1998; ks. myös alaviite 15, Koulutuksen tutkimuslaitoksen roolista aiemmissa vaiheissa). Perinteisten paperi-kynä-testien rinnalle kehitettiin jo Opetushallituksessa digitaalista vaihtoehtoa. Kokeilujen alkusysäyksenä oli tieto PISA- ja TIMMS-tiedonkeruiden digitalisoitumisesta: ennen kansainvälisiä digitaalisia mittauksia katsottiin tärkeäksi varmistaa suomalaisten oppilaiden selviytyminen digitaalisessa testausympäristössä. Ensimmäistä kertaa kansallista digitaalista testausta kokeiltiin vuosien kehittelyn jälkeen Karvissa vuonna 2014 yhdeksännen luokan äidinkielen ja kirjallisuuden arvioinnissa sekä suomeksi (Harjunen & Rautopuro, 2015) että ruotsiksi (Silverström & Rautopuro, 2015). Näissä arvioinneissa noin kolmasosa kouluista osallistui digitaaliseen testaukseen. Vuonna 2015 yhdeksännen luokan matematiikan arvioinnissa kaikki otokseen valitut koulut osallistuivat sekä perinteiseen paperi-ja-kynä-testaukseen että digitaaliseen testaukseen (Julin & Rautopuro, 2016) ja oppilaat tekivät puolet tehtävistä paperilla ja puolet tietokoneella. Tällä kerralla Karvi sai opettajilta poikkeuksellisen paljon negatiivista palautetta tästä kouluille ja opettajille raskaaksi osoittautuneesta järjestelystä.

15 Ks. https://www.umu.se/en/research/. Suomessa Jyväskylän yliopiston Koulutuksen tutkimuslaitos, on ollut aktiivinen vuodesta 1968 lähtien (ks. https://ktl.jyu.fi/en/research)—ei kuitenkaan niinkään testiteorian alan julkaisujen osalta.

16 Ks. https://www.riha.ee/Infos%C3%BCsteemid/Vaata/eis

17 Ks. https://www.uv.uio.no/cemo/english/. Yksikkö on ollut *erittäin* aktiivinen testiteoriaan liittyvien kansainvälisten julkaisujen tuottajana. Vuosien 2020–2022 aikana yksikkö on tuottanut yli 100 tieteellistä julkaisua (ks. https://www.uv.uio.no/cemo/english/about/news-events-and-publications/latest-100-publications/index.html).

Ensimmäinen täysin digitaalisesti toteutettu kansallinen oppimistulosarviointi oli Karvin vuonna 2016 järjestämä ammatillisen koulutuksen kestävän kehityksen osaamisen ja asenteiden arviointi (Räkköläinen ym., 2017). Tätä seurasivat perusopetuksen oppimistulosarviointeina vuonna 2018 suomen kielen arviointi ruotsinkielisissä kouluissa ("Finska") 6. luokan lopussa (Åkerlund, Marjanen, & Lepola, 2019), matematiikan ja äidinkielen ja kirjallisuuden lähtötason kartoitus oppilaiden tullessa kouluun 1. luokan alussa (Ukkola & Metsämuuronen, 2019; Metsämuuronen & Ukkola, 2019) sekä englannin kielen arviointi 7. luokan alussa (Härmälä ym., 2019). Näistä ammatillisen koulutuksen hanke (Räkköläinen ym, 2017) tehtiin yhteistyössä Turun yliopiston ViLLE-tiimin kanssa. Perusopetuksen oppimistuloksia varten Karvissa kehitettiin oma oppimistulosten digitaalinen arviointialusta, jonka kehitystyön toteutti NordicEdu Oy. Karvin arviointialustaa käytettiin sittemmin 9. luokan äidinkielen ja kirjallisuuden kahden oppimäärän (suomen kieli ja kirjallisuus sekä svenska och litteratur) arvioinneissa (Hellgren & Marjanen, 2020; Kauppinen & Marjanen, 2020), 3. luokan alussa matematiikan ja äidinkielen ja kirjallisuuden arvioinnissa (Ukkola & Metsämuuronen, 2021, 2023) ja 9. luokan englannin A-oppimäärän arvioinnissa (Härmälä ym., 2022) ja Finskan arvioinnissa (Åkerlund ym., 2022), perusopetukseen valmistavan opetuksen yhteydessä suomen kielen arvioinnissa (Venäläinen ym., 2022) sekä 9. luokan ruotsin A- ja B-oppimäärien arvioinnissa. Alusta kehitettiin erityisesti kielten arvioinnin tarpeisiin. Se pystyi vastaamaan myös nuorimpien oppilaiden matematiikan tarpeisiin, mutta sen ominaisuudet eivät olleet riittäviä oppilaiden monipuoliseen matemaattisten taitojen arviointiin 9. luokalla. Vuodesta 2021 lähtien Karvissa on yhteistyössä Opetushallituksen ja opetus- ja kulttuuriministeriön kanssa kehitetty uutta digitaalista arviointijärjestelmää, jonka 1. versiota on tarkoitus käyttää ensimmäisen kerran vuonna 2024.

Vuoden 2021 matematiikan oppimistulosarviointi toteutettiin Karvin alustan sijaan Turun yliopiston oppimisanalytiikan keskuksen ViLLE-alustalla. Tähän mittaukseen liittyviä erityispiirteitä kuvataan tässä artikkelissa.

3.2 Digitaalisen testaamisen periaatteita

3.2.1 Adaptiiviset, lineaariset ja staattiset testit

Digitaalinen testaus voidaan toteuttaa teknisesti kolmella päätavalla: staattisesti, lineaarisesti ja adaptiivisesti. Kehittynein digitaalisen testauksen muoto on adaptiivinen testaus (*Computer Adaptive Testing*, CAT; ks. mekaniikasta ja historiasta esimerkiksi Linacre, 2000). Adaptiivisessa testauksessa algoritmi valitsee osiopankista yksittäiset tehtävät eli osiot, jotka sopivat testattavan osaamisen tasoon. Tämä edellyttää käytännössä ns. osio-vaste-teorian (*item response theory*, IRT) ja siihen liittyvän mallinnuksen käyttämistä. IRT-mallinnuksen avulla lasketut ja pankitetut vaikeustasoparametrien (ns. *b*-parametrien) arvot ovat toisiinsa linkittyneitä ja vertailukelpoisia.

Adaptiivista testausta voidaan tehdä myös kiinteillä osiosarjoilla tai moduuleilla, jotka muodostavat pieniä osatestejä adaptiivisen algoritmin sisällä. Tämän kaltainen testijärjestelmä oli käytössä esimerkiksi ensimmäisissä PIAAC-testeissä. Monissa maissa on digitaalisen testauksen myötä siirrytty ainakin osittaiseen adaptiiviseen testaukseen, sillä adaptiivisuuden on havaittu mahdollistavan tehokkaamman ajankäytön itse testitilanteessa; testiin tarvittava aika saattaa puolittua, mutta tarkkuus säilyy samana (mm. Van der Linden & Glas, 2000; Wainer, 2000; Weiss & Kingsbury, 1984; ks. myös Vie ym. 2017). Vaikka osaamisen testaamiseen on nykyään monia mahdollisuuksia, on adaptiivisen testauksen soveltaminen erityisen haasteellista tilanteessa, jossa on tarkoituksena tuottaa oppilaan osaamisesta diagnostista tietoa opettajalle. Tämä johtuu siitä, että adaptiivinen testaaminen voi toimintalogiikkansa vuoksi antaa vain rajallisesti tietoa osasta testattavien tiedosta ja osaamisesta. Kun oppilaat tekevät toisiinsa nähden paljon erilaisia tehtäviä, opettajien on vaikea tietää, millaisissa tehtävissä oppilas onnistui ja millaisissa ei (ks. kritiikistä mm. Kreiner, 2021). Tästä syystä Pohjoismaista esimerkiksi Tanska on siirtymässä pois adaptiivisesta testauksesta perinteisempään, lineaariseen testaukseen (Regeringen, 2021).

Lineaarinen testaus on adaptiivista testausta karsitumpi muoto. Tässä algoritmi valitsee osiopankista kullekin testattavalle eri osiot, mutta ei ota erityisesti huomioon testattavan osaamisen tasoa (vrt. edellä adaptiivinen testaus). Testi rakentuu siten, että algoritmi valitsee testiosioita osiopankista vaikeustasoparametrin mukaisesti vaikeutuvaan järjestykseen, helpoimmista vaikeampiin. Osioita ei välttämättä aina ole tarpeen valita täysin niiden vaikeusjärjestyksen mukaisesti.

Yksinkertaisin digitaalisen testauksen muoto on digitalisoida perinteinen paperi-kynä-testi, jolloin se saatetaan esimerkiksi pdf-muotoon. Tällöin ei tietenkään kaikkia digitaalisen testauksen mahdollisuuksia—kuten tiettyjä osiotyyppejä—voida hyödyntää testauksessa. Tämänkaltainenkin staattinen testi voi toisaalta olla hyvin valmisteltu ja sisältää useita tehtäväsarjoja samaan tapaan kuin (hyvin valmistelluissa) perinteisissä paperi-kynä-testeissä voi olla. Hieman edistyneemmässä staattisessa testauksessa digitaalisen alustan mahdollisuuksia ja erityisiä osiotyyppejä käytetään testausmahdollisuuksien laajentamiseen, vaikka itse testi voi olla identtinen kaikille testien suorittajille. Kaikki Suomessa perusopetuksen piirissä toteutetut oppistulosarvioinnit, niin kansalliset ja kansainväliset (ks. yllä oleva kirjallisuus) kuin myös vuoden 2021 matematiikan mittaus, ovat olleet tämän tyyppisiä staattisia testejä eikä tulevan alustankaan ensimmäinen versio poikkea tästä.

3.2.2 Dynaamiset ja staattiset testit

Dynaamisissa tai interaktiivisissa osaamistesteissä oppijoille annetaan testauksen aikana oppimista edistävää palautetta, jotta he voivat parantaa tulostaan seuraavassa testissä (ks. Feuerstein, Rand, & Hoffman, 1979; Grigorenko & Sternberg, 1998; Metsämuuronen & Mattsson, 2013; Roediger & Karpicke, 2006a; Sternberg & Grigorenko, 2002; ks. keskustelu Haywood, 1992; Hansen, 2011). Jälkimmäisen osalta Haywood (1992) ehdotti yleiseksi termiksi ’interaktiivinen arviointi’, sillä yleisemmin käytössä oleva termi ’dynaaminen arviointi’ saattaa assosioitua liikaa tiettyyn menetelmään, erityisesti Feuersteinin ja kollegoiden (1979) menetelmään. Terminä dynaaminen arviointi—ja näin myös dynaaminen testaus—on kuitenkin yleisemmin käytössä, vaikka muitakin nimiä on ehdotettu (ks. vaihtoehdoista mm. Hansen, 2011).

Dynaamisia testejä käytetään erityisesti, kun halutaan opettaa asiaa testauksen avulla; virheelliset vastaukset korjataan, jolloin testattavan oppimispotentiaalia tai sen perustana olevia valmiuksia voidaan parantaa. Vaikka toistuvalla testaamisella itsellään on toistuvasti osoitettu olevan mieleen palauttamista (eli oppimista) vahvistava vaikutus (ks. aikajärjestyksessä esimerkiksi Tulving, 1967;

Roediger & Karpicke, 2006a, 2006b; Metsämuuronen, 2013; Metsämuuronen & Mattson, 2013; Grevig & Richter, 2018), testin aikana suoriutumisesta palautetta saavilla oppijoilla lopulliset testitulokset ovat parempia kuin ilman palautetta (ks. Butler, Karpicke & Roediger, 2007; 2008; Butler & Roediger, 2008; Metcalfe, Kornell & Finn, 2009; Metsämuuronen & Mattsson, 2013; Vojdanoska, Cranney & Newell, 2009).

Staattisissa testeissä testattaville ei anneta testauksen aikana palautetta suoriutumisesta. Yleensä palaute annetaan testin jälkeen kokonaisuutena esimerkiksi testin summapistemäärän tai siitä muodostetun arvosanan muodossa. Kansallisissa ja kansainvälisissä laajoissa testauksissa staattista testausta käytetään yleisesti, sillä tehtävien oikeat vastaukset on usein tärkeä salata. Näin tehtäviä voidaan myöhemmin käyttää ns. linkkitehtävinä, joiden avulla eri vuosien tehtäväsarjojen vaikeustasot voidaan linkittää ja vertaistaa toisiinsa. Linkkitehtävien vapauttaminen mahdollistaisi sen, että opettajat alkaisivat opettaa kyseisten tehtävien oikeat ratkaisut ja näin seuraavien vuosien oppilaiden tuloksia olisi mahdotonta verrata uskottavasti aiempien vuosien oppilaiden tuloksiin. Tämä perustelu ei kuitenkaan täysin päde Karvin 9. luokan testeihin, jossa oppilas hyvinkin voisi saada omat tulostietonsa, sillä sama oppilas tuskin osallistuu siihen uudestaan.

Vaikka dynaamista testausta usein pidetään vastakkaisena staattiselle testaamiselle, Haywoodin (1997) ajatus on huomion arvoinen: hyvässä tapauksessa dynaaminen testaus täydentää perinteistä standardoitua tai staattista testausta eikä suinkaan korvaa perinteisiä menetelmiä. Dynaamista testausta ja perinteistä staattista testausta voidaan käyttää yhdessä erilaisten tietojen löytämiseen, ellei dynaamista testausta käytetä peräti asian opettamisen muotona, kuten esimerkiksi ViLLE-ympäristön opintopolussa (OAK, 2017).

Karvin vuoden 2021 matematiikan osaamista mittaava testi oli tyypillinen staattinen testi sen molemmissa merkityksissä: Kaikki saman tehtäväsarjan suorittaneet oppijat saivat identtisen testin eikä tehtävistä saanut palautetta testauksen aikana.

3.3 Digitaalisen arvioinnin toteutukseen liittyviä teknisiä kysymyksiä

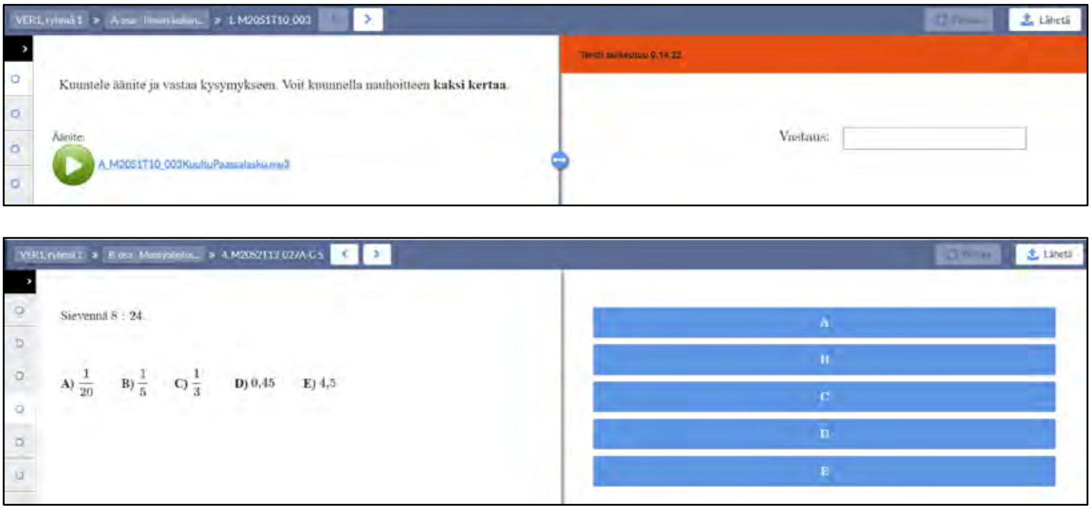
Tehtävien laatimisen prosessia kuvataan toisaalla (Metsämuuronen & Nousiainen, 2023; ks. luku 2). Tässä yhteydessä kuvataan lähemmin mittaukseen liittyviä teknisiä ratkaisuja.

3.3.1 Tehtävien rakentaminen: ideointi, layout ja tekniset ratkaisut

Koska Karvin arviointialusta ei tukenut kaikki matematiikan tehtävätyyppejä ja alusta oli varattu muihin, samaan aikaan tehtäviin mittauksiin, arvioinnin tekniseksi toteuttajaksi valittiin Turun yliopiston ViLLE-tiimi (myöhemmin Turun yliopiston Oppimisanalytiikan tutkimuskeskus, TRILA). Keväällä 2019 TRILA esitteli ViLLE:n ominaisuuksia tehtävälaatijoiden ryhmälle. Mahdollisina uudenlaisina, digitaaliseen ympäristöön soveltuvina ratkaisuina esiteltiin esimerkiksi erilaisia tapoja esittää animaatioita ja interaktiivisia GeoGebra-appletteja tehtävänannoissa.

Myös vastausten antamista GeoGebran avulla esiteltiin. Tämän tarkoituksena oli inspiroida tehtävien laatijat käyttämään digitaalisen alustan tarjoamia mahdollisuuksia kansallisen arvioinnin tehtävissä.

Kun tehtävät oli laadittu, niiden asettelu suunniteltiin keskenään yhtenäiseksi niin, että jokaisessa tehtävässä oli aina vasemmalla puolella ruutua varsinainen tehtävänanto ja oikealla puolella ruutua vastauskentät (ks. Kuvio 11). Tällä pyrittiin saamaan tehtävien eteneminen mahdollisimman intuitiiviseksi ja keskenään yhteneväksi: Tehtävän lukeminen etenee kuten kirjan aukeama ja vastauskentät ovat kaikissa tehtävissä samassa kohdassa ruudun oikealla puolella.



KUVIO 3.11. Tehtävien yleinen asettelu: tehtävänanto vasemmalla, vastaukset annetaan oikealta puolelta

Matemaattisten lausekkeiden syöttäminen tietokoneelle ei ole aivan ongelmatonta. Esimerkiksi murtolukujen, neliöjuurien ja tiettyjen symbolien kuten ° syöttäminen näppäimistöltä ei ole aivan suoraviivaista. Tätä varten digitaalisia tehtäväsarjoja varten kehitettiin ViLLE:n ”matikkapalkki”, jonka avulla oppilaat saivat hiirellä klikkaamalla syötettyä näitä tyypillisestä näppäimistöstä puuttuvia merkintöjä vastauskenttiin (Kuvio 12). Palkki oli jaettu kahteen osaan: operaatioihin ja symboleihin. Operaatiot-osassa oli peruslaskutoimitukset sekä murtolukumerkintä, toiseen korotus, mielivaltaiseen potenssiin korotus, neliöjuuri ja mielivaltainen juuri. Symboli-osassa olivat perusmerkinnät kuten prosentti-, aste- ja kertomerkki.

KUVIO 3.12. Matematiikkaeditori

3.3.2 Vastausten automaattinen pisteitys

Automaattista pisteitystä hankaloittaa se, että vastaus ei välttämättä ole täysin yksiselitteisesti juuri tietynlainen merkkijono. Jos oikea vastaus olisi vaikkapa 50 %, niin oikeaksi vastaukseksi hyväksytään yleensä myös vaikkapa sellainen vastaus, jossa oppilas on kirjoittanut ”50 prosenttia” tai sellainen, jossa luvun ja prosenttisymbolin välillä ei ole välilyöntiä: ”50%”. Oppilas saattaa myös kirjoittaa vastauskenttään jotain muuta, kuten ”Vastaukseni on 50 %” tai ”8 / 4 = 50 %”. Kun tehtävien pisteitys on automaattista, miten määritellään oikean vastauksen tulkinta niin, että algoritmi hyväksyy kaikki oikeaksi kelpaavat vastaukset eikä mitään vääriä vastauksia? Joissain tapauksissa—kuten esimerkiksi kielten opetteluun ja testaamiseen tarkoitetussa *Duolingo*-applikaatiossa (ks. https://www.duolingo.com/)—oikean vastauksen määrittelee miljoonien käyttäjien tuottama data. Tässä arvioinnissa tällaista ei ollut käytössä, ja niinpä määrittely tehtiin manuaalisesti.

Oikean vastauksen arvioimista varten vastauksesta poimittiin kaksi asiaa: Lukuarvo ja mittayksikkö. Lukuarvoksi tulkittiin viimeinen vastauksessa ollut lukuarvo ja yksiköksi kaikki, mitä tämän viimeisen lukuarvon jälkeen oli kirjoitettu. Jokaiseen tehtävään määriteltiin myös kaikki oikeiksi hyväksytyt mittayksiköt suomeksi, ruotsiksi ja englanniksi. Jos oikea yksikkö oli vaikkapa ”cm”, niin yksiköksi hyväksyttiin yksikön ”cm” lisäksi myös ”senttimetriä”, ”senttimetri”, ”senttiä”, ”sentti”, ”centtimetriä”, ”centtimetri”, ”centtiä”, ”centti”, ”centimeters”, ”centimeter”, ”centimetres”, ”centimetre”, ”centimetrar”. Mekanismia, jolla kirjoitusvirheet olisi havaittu, ei toteutettu. Oikeaksi vastaukseksi hyväksyttiin myös sellaiset vastaukset, joista yksikkö puuttui kokonaan, jos yksikkö implikoitiin tehtävänannon kysymyksessä selvästi. Tällainen kysymys olisi voinut olla esimerkiksi muotoa ”Kuinka monta senttimetriä pitkä köysi on?”. Välilyönneistä vastauksissa ei välitetty lainkaan ja joissakin tehtävissä mahdollistettiin myös eri mittayksiköinä vastaaminen määrittelemällä tietyt kirjaimet kertoimiksi, kuten c = 0,01 ja d = 0,1. Tällöin esimerkiksi vastaus ”15cm” tulkittiin vastaukseksi ”15 * 0,01 m” eli ”0,15 m.

Näin ollen esimerkiksi vastaus ”Vastaukseni on 1 cm + 5 cm = 6 senttiä” tulkittiin muotoon ”6 cm” ja tämän jälkeen sitä verrattiin oikeaksi määriteltyyn vastaukseen.

3.3.3 Testin rakenne ja testissä eteneminen

Testi rakennettiin kolmesta tehtäväosuudesta ja kahdesta harjoitteluosuudesta:

- Harjoittelu A-osuuteen
- A-osuus
- Harjoittelu B- ja C-osuuksiin
- B-osuus
- C-osuus

(ks. osuuksien sisältöä Liitteessä raportin lopussa). A-osuuteen valittiin päässälaskutehtäviä, B-osuuden monivalinta- ja yhdistämistehtäviä ja C-osuuteen vaativampia ongelmanratkaisutehtäviä. Näiden lisäksi mittaukseen kuului myös oppilaskysely, mutta sitä ei käsitellä tässä yhteydessä. Jokaisen osan lopussa oli ohjeet seuraavaan osioon siirtymiseksi. Lisäksi C-osuuden lopussa oli ohjeet annettujen vastausten tarkastamiseksi.

Koko testi alkoi A-osuuden harjoitusosuudella, jonka tarkoitus oli antaa oppilaille turvallinen ja stressitön ympäristö erilaisten vastausten syöttämisen harjoitteluun. Siinä harjoiteltiin eri vastaustapoja, joita olivat kuultuun kysymykseen vastaaminen, avoimeen vastauskenttään vastaaminen erikoismerkkien kanssa ja ilman sekä monivalintatehtävään ja yhdistä parit -tyyppiseen tehtävään vastaaminen. A-osuuden harjoittelun lopuksi oppilas sai ohjeet seuraavaan osioon siirtymiseksi.

Varsinaisen A-osuuden kesto oli rajoitettu 15 minuuttiin. A-osuus koostui yksinomaan päässälaskutehtävistä, joissa kysymys esitettiin valmiiksi äänitettynä. Osuudessa ei saanut käyttää laskinta apuna. Äänityksen pystyi kuuntelemaan enintään kaksi kertaa. Tämä sama rajoite oli myös A-osuuden harjoittelussa. Osa sulkeutui 15 minuutin kuluttua sen avaamisesta, eikä oppilas päässyt enää palaamaan päässälaskutehtäviin.

B- ja C-osuuksien harjoitteluosassa oppilas sai harjoitella tehtävätyyppejä, joita ei vielä ollut tehnyt: välivaiheiden merkitseminen varsinaisen vastauksen lisäksi ja vastauksen merkitseminen lukusuoralle *GeoGebra*-appletin avulla. B-osuudessa oli tehtäviä, joihin vastattiin monivalinnoilla ja lyhyillä vastauksilla. C-osuudessa oli monivalintojen ja lyhytvastausten lisäksi myös laajempia tuottamistehtäviä, joiden vastauksiin vaadittiin pelkkien vastausten lisäksi myös perusteluja ja välivaiheita.

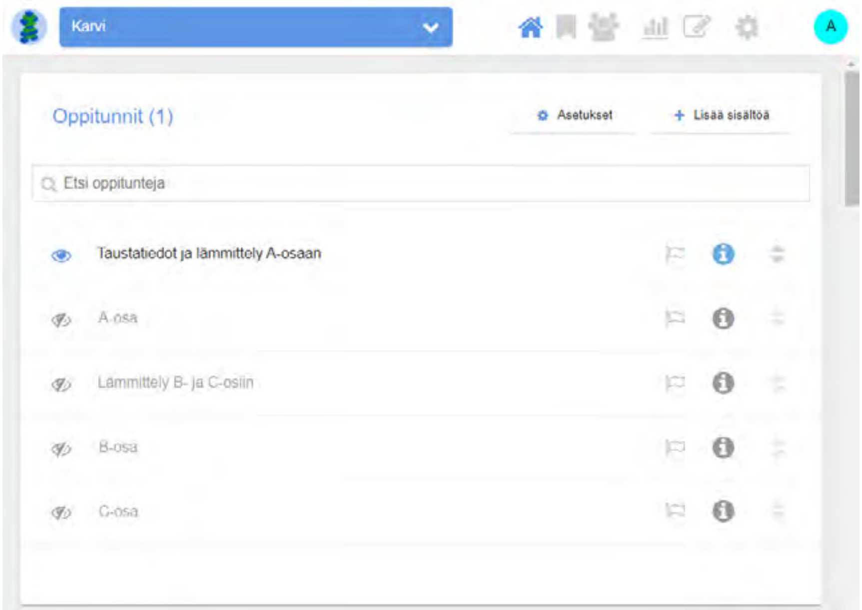
B- ja C-osuuksien sisällä oppilaat saivat tehdä siihen kuuluvat tehtävät haluamassaan järjestyksessä. Klikattuaan osuuden auki oppilaille aukesi lista osuuden tehtävistä (Kuvio 13). Tehtävän nimeä klikkaamalla oppilas pääsi suorittamaan kyseisen tehtävän. Jokaisen tehtävän lopuksi oppilaat pystyivät joko etenemään suoraan seuraavaan tehtävään tai palaamaan takaisin tehtävälistaan. Osuuden viimeisen tehtävän jälkeen oppilaille annettiin ohje tarkastaa vielä, että kaikki tehtävät oli tehty. Omia vastauksia pystyi vielä tarkastelemaan ja muuttamaan niin halutessaan. Oppilaille annettiin myös ohjeet siitä, mitä tehdä sen jälkeen, kun hän oli saanut kyseisen osuuden suoritettua.

KUVIO 3.13. Tehtävälista, josta klikkaamalla oppilas pääsi siirtymään tehtävään

3.3.4 Arviointijärjestelmä opettajan näkökannalta

Opettaja huolehti tehtäväversioiden avaamisesta oppilaille, versioiden sulkemisesta testiajan päätyttyä, tuottamistehtävien pisteittämisestä ohjeiden mukaisesti sekä tulosten lataamisesta järjestelmästä. Karvi ja TRILA loivat kullekin oppilaalle etukäteen tunnukset ViLLE-järjestelmään, ja kouluille lähetettiin oppilaiden henkilökohtaiset kirjautumistunnukset ja salasanat ennen arviointia.

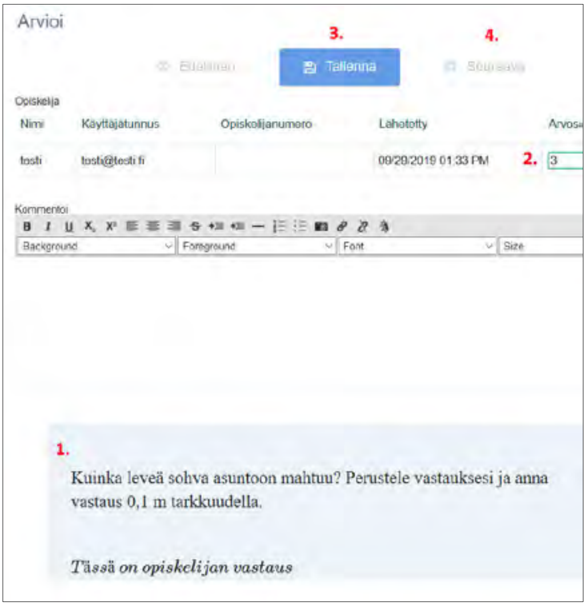
Järjestelmän opettajanäkymä rakennettiin niin, että siellä oli helppokäyttöisyyden vuoksi näkyvissä vain tärkeimmät toiminnot ja elementit, kuten pisteitysnäkymä, ohjesivu sekä tulosten lataussivu (Kuvio 14). Opettajat saivat etukäteen myös kuvalliset ohjeet järjestelmän käytöstä. Etusivulla opettaja pääsi avaamaan tehtäväversion kaikki osat ja samalla sivulla myös sulkemaan ne arvioinnin lopuksi, jotta oppilaat eivät päässeet muuttamaan vastauksia enää arvioinnin jälkeen. Opettaja saattoi seurata oppilaiden etenemistä pisteitysnäkymässä, johon kirjautuivat automaattisesti arvioitavien tehtävien pisteet sitä mukaa, kun oppilas oli vastannut tehtävään. Tämä oli tärkeä työkalu etäopetuksessa olevien oppilaiden seuraamisessa, kun opettaja ei pystynyt fyysisesti tarkastamaan, eteneekö oppilas tehtävissä tai onko hän esimerkiksi jumittunut johonkin tehtävään. Opettaja ei kuitenkaan nähnyt itse arviointitehtäviä tai oppilaiden antamia vastauksia, ainoastaan pistemäärät tehtävittäin.



KUVIO 3.14. Opettajanäkymä

Kukin tehtäväversio sisälsi versiosta riippuen 6–9 tehtävää, jotka opettajien tuli pisteittää annettujen ohjeiden mukaisesti. Muut tehtävät pisteitettiin automaattisesti edellä kuvatulla tavalla. Opettaja latasi tehtävien pisteitysohjeet järjestelmästä pdf-muotoisena tiedostona, minkä jälkeen hän avasi pisteitysnäkymässä jonkin tuottamistehtävään annetun oppilasvastauksen ja pisteitti sen ohjeiden mukaan (ks. Kuvio 15 pisteitysikkunasta).

Pisteitysnäkymä rakennettiin järjestelmään niin, että opettaja saattoi tarkastaa yhden tehtävän kaikki oppilasvastaukset peräkkäin, jolloin saavutettiin ajallista etua pisteityksessä, kun opettajan tarvitsi seurata kerrallaan vain yhteen tehtävään annettuja oppilasvastauksia ja pisteitysohjetta. Tämä sekä nopeuttaa opettajan työtä että saattaa tuottaa myös yhdenmukaisempia pisteitä kuin oppilas kerrallaan pisteittämällä. Saman koulun opettajat pystyivät myös jakamaan pisteitystyön tehtävien perusteella esimerkiksi niin, että yksi opettaja tarkasti esimerkiksi kolme ensimmäistä tehtävää kaikilta oppilailta ja toinen opettaja tarkasti muut tehtävät. Mahdollista oli tarkastaa tehtävät myös oppilaittain esimerkiksi niin, että kukin opettaja tarkasti omien opetusryhmiensä oppilaiden suoritukset. Tämä kuitenkin vaati enemmän mekaanista työtä, kun kaikki koulusta osallistuneet oppilaat näkyivät samalla nimilistalla, eikä opetusryhmätietoa ollut saatavilla, mikä olisi helpottanut oppilaiden järjestämistä opetusryhmän mukaan.



KUVIO 3.15. Opettajanäkymän pisteitysikkuna

Kun kaikki oppilasvastaukset oli pisteitetty, opettaja pystyi lataamaan Excel-tiedostona koulunsa oppilaiden tulokset järjestelmästä ja tarkastelemaan esimerkiksi yksittäisestä tehtävästä saatuja pistemääriä tai vaikka koko päässälaskuosuuden pistemääriä. Lisäksi hän näki, kuinka monta prosenttia maksimipistemäärästä kukin oppilas oli saavuttanut koko testissä ja osatesteissä. Kullekin oppilaalle muodostettiin myös kaksi[18] arvosanaehdotusta arviointisuoriutumisen perusteella, ja opettaja sai ladattua arvosanaehdotukset järjestelmän opettajanäkymästä.

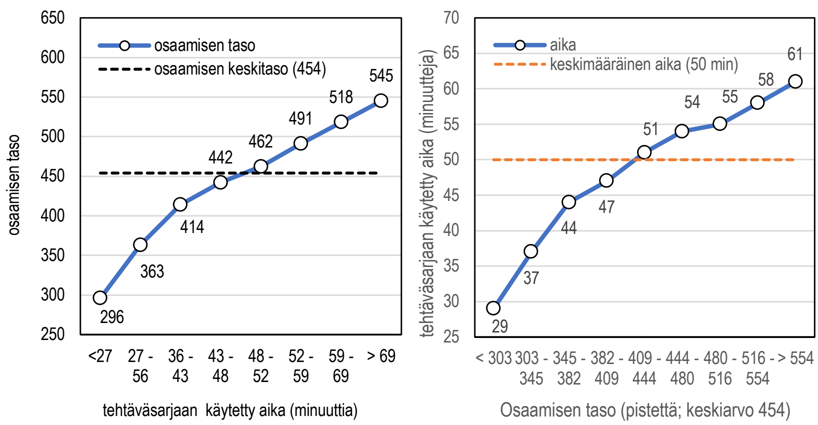
3.4 Testin teknisiin ominaisuuksiin liittyviä huomioita

Testialusta tuki tietokoneita ja tablettilaitteita ja vähimmäisresoluutio leveyssuunnassa oli 1024px. Järjestelmä oli optimoitu toimimaan mittausajanhetkellä käytössä olleilla *Mozilla Firefox*- sekä *Google Chrome*-selaimilla ja myös uusimmat versiot *Safari*- ja *Edge*-selaimista olivat toimivia. Sen sijaan Internet Explorerin käyttöä järjestelmä ei tukenut.

Tehtäväsarjojen tehtäviin käytettyyn aikaan liittyvä tarkastelu rajataan tässä vain siihen tehtäväsarjaan, joka julkaistaan koko raportin liitteessä. Tätä tehtäväsarjaa kutsutaan nimellä VER1. Tässä tarkastelussa käsitellään koko tehtäväsarjaa—myös niitä tehtäviä, joita myöhemmin ei laskettu mukaan kokonaispisteisiin ja niitä, joita ei julkaista. Oppilaathan kuitenkin tekivät kaikki tehtävät varsinaisessa tiedonkeruussa.

18 Raportin liitteessä julkaistavan tehtäväsarjan yhteydessä on esitelty myös arvosanaehdotus, joka perustuu otokseen osallistuneiden oppilaiden (n = 12 000+) viimeisen vahvistetun matematiikan arvosanan jakaumaan. Vastaava, osioiden painottamattomaan summaan perustuva arvosanaehdotus valmisteltiin myös varsinaisessa testauksessa. Lisäksi varsinaisessa testauksessa muodostettiin painotettuun summaan liittyvä arvosanaehdotus.

Yksittäiseen tehtävään käytetty maksimiaika vaihteli välillä 12–149 sekuntia, ja keskiaika oli 53 sekuntia. Keskimäärin tehtäväsarjan suorittamiseen käytettiin aikaa 50 minuuttia. Paremmin suoriutuneet oppilaat käyttivät testiin enemmän aikaa kuin heikommin suoriutuneet, ja enemmän aikaa tehtäviin käyttäneet oppilaat saivat tyypillisesti enemmän pisteitä kuin vähän aikaa käyttäneet. Jos oppilas käytti testiin vähemmän aikaa kuin 27 minuuttia, testitulos jäi keskimäärin tasolle 196 pistettä, ja jos aikaa käytettiin yli 69 minuuttia, osaaminen oli keskimäärin tasolla 545 pistettä kansallisen keskiarvon ollessa 454 pistettä (Kuvio 16). Vastaavasti, jos oppilas sai pistemääräkseen alle 303 pistettä, hän oli käyttänyt testiin keskimäärin 29 minuuttia kun taas oppilas, joka sain testissä yli 554 pistettä, oli käyttänyt aikaa keskimäärin 61 minuuttia.

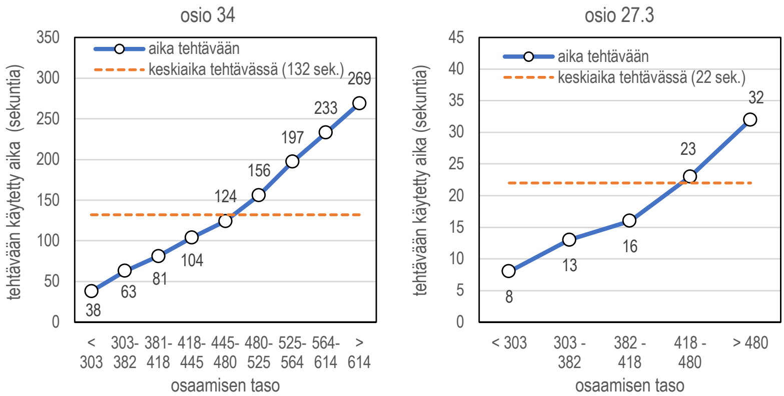


KUVIO 3.16. Ajankäytön yhteys tehtäväsarjassa menestymiseen (VER1)

Kaksi osiota tehtäväsarjassa VER1 osoittautui mielenkiintoiseksi erottelukyvyn ja ennustearvon näkökulmasta, tehtävät 34 ja 27.3 (ks. Liite raportin lopussa). Lineaarisessa regressioanalyysissa näissä kahdessa osiossa vastaamiseen käytetty aika ennusti lähes 40 % (R^2_{adj}= 0,37) koko testin kokonaispistemäärän vaihtelusta. Näistä tehtävä 34 oli vaikeahko, syksyn 2018 lyhyen matematiikan oppimäärän ylioppilaskoetehtävä, jossa tuli laskea levyn pinta-ala annettujen tietojen pohjalta. Periaatteessa tehtävä oli yksinkertainen, joskin hieman työläs suorakulmaisen kolmion ominaisuuksiin perustuva pulma. Parhaiten menestyneet oppilaat (pistemäärä yli 614 pistettä) käyttivät tehtävän ratkaisemiseen keskimäärin 4,5 minuuttia (269 sekuntia), kun keskiosaajat käyttivät tehtävään aikaa 2 minuuttia (124 sekuntia) (Kuvio 17).

Tehtävä 27.3. puolestaan oli yksinkertainen, nopeasti ratkaistavissa ollut monivalintatehtävä, jossa piti tunnistaa ja laskea kuution särmien yhteispituus. Parhaimmin menestyneet (> 480 pistettä)[19] käyttivät tehtävään aikaa keskimäärin puoli minuuttia (32 sekuntia), keskitasoisilta oppilaita aikaa meni 23 sekuntia ja heikoimmat oppilaat (< 303 pistettä) ohittivat tehtävän 8 sekunnin aikana. Näyttää siis siltä, että parhaimmin menestyneet oppilaat varmistelevat suoritustaan jopa aivan helpoissa tehtävissä.

19 Koska tehtävä 27.3 oli helppo, "parhaiden" oppilaiden ei tarvinnut olla kuin keskitasoisia ja tätä parempia. Luokittelussa käytetty DTA ei pysty erottamaan toisistaan keskitasoisia ja tätä paremmin suoriutuneita oppilaita, koska he kaikki käyttivät vain vähän aikaa tehtävän ratkaisemiseen toisin kuin osiossa 34.



KUVIO 3.17. Ajankäyttö kahdessa tehtävässä ja sen yhteys kokonaisosaamiseen (VER1)

Tytöt käyttivät hieman enemmän aikaa testin suorittamiseen (52 minuuttia) kuin pojat (48 minuuttia). Ero on merkitsevä ($p < 0{,}001$), mutta ei merkittävän suuri ($f = 0{,}12$).

3.5 Digitaalisen testauksen haasteita vuoden 2021 mittauksen näkökulmasta

3.5.1 Yleisiä näkökulmia

Kansallisen arvioinnin viitekehyksessä digitaalinen testaus teknisenä suorituksena itsessään ei ole ongelmallista nykyteknologialla; käytettävissä on toimivia ohjelmistoja ja käytänteitä laajojen oppilasryhmien testaamiseen. Neljä suurinta haastetta ovat teknologiaan liittyvät haasteet, tasa-arvohaasteet, muuttuneeseen testikäyttäytymiseen liittyvät haasteet ja etätestaukseen liittyvät uskottavuushaasteet, joita pohditaan tässä yhteydessä. Teknologiaan liittyvät haasteet liittyvät siihen, että täysin yllättäviä, paperi-kynä-testissä esiintymättömiä, uusia haasteita saattaa tulla esiin testitilanteessa. Tasa-arvoon liittyen joudumme kysymään, voivatko/osaavatko tai haluavatko kaikki testattavat näyttää parasta osaamistaan käyttämillämme digitaalisilla testialustoilla. Testikäyttäytymiseen liittyen on mielekästä kysyä, käyttäytyvätkö oppilaat oleellisesti poikkeavalla tavalla digitaalisessa ympäristössä kuin perinteiseen paperi-kynä-testiin osallistuessaan. Uskottavuuteen liittyen joudumme kysymään, voimmeko luottaa testin tuottamaan arvioon osaamisesta ja toisaalta, mitä voimme tehdä tuloksen uskottavuuden lisäämiseksi.

Toisentyyppisiä haasteita, kuten teknologian jatkuvaa kehittymistä ja lisääntyviä kustannuksia, on käsitellyt mm. Brown (2019), testiin sitoutumattomuutta, ahdistuneisuutta ja testeissä huijaamista on tutkinut mm. Wise (2019), tietokoneruudulta lukemisen haasteita on kuvannut mm. Nardi ja Ranieri (2019) ja mahdollisia vertailukelpoisuuden haasteita paperi-kynä-testien ja digitaalisten testin välillä eli ns. moodiefektiä (*Mode effect*) ovat selvittäneet mm. Jerrim ja kollegat (2018). Osaa näistä käsitellään tässä luvussa. Luvussa 3.6 käsitellään tarkemmin pidemmälle ulottuvia megatrendejä ja digitaalisen testauksen tulevaisuutta ja siihen liittyviä haasteita.

3.5.2 Yllättävät teknologiset haasteet

Jacobsen (2020) on koonnut digitaaliseen testaukseen liittyviä teknologisia haasteita. Digitaaliseen testiin osallistuvat tarvitsevat ensinnäkin enemmän teknistä asiantuntemusta kuin paperi-kynä-testiin osallistuvat. Tämä ei Suomessa näytä olevan suuri ongelma—viimeistään etäopiskelu sai aikaan digiloikan niin opettajilla kuin oppilailla (ks. pohdintaa esimerkiksi Goman, ym., 2021; Lavonen & Salmela-Aro, 2022; Metsämuuronen & Seppälä, 2021).

Toiseksi, jos odottamattomia teknisiä ongelmia ilmenee, testin jatkaminen on haasteellista. Testin valvojalla ei välttämättä ole mahdollisuuksia tai osaamista reagoida riittävän nopeasti odottamattomiin haasteisiin kuten esimerkiksi sähkökatkoihin tai tietoverkkoyhteyksien katkeamiseen. Näin kävi vuoden 2021 matematiikan testin osalta seutukunnalla, jossa operaattorin runkoverkko kaatui juuri testauksen aikana ja osalla oppilaista jäi päässälaskuosuus aikarajoituksen vuoksi suorittamatta. Toisin sanoen, jos opettaja oli jo käynnistänyt aikarajatun päässälaskuosuuden, sitä ei voitu lopettaa ja aloittaa alusta. Päässälaskuosuus oli testin osa-alueista ainoa, joka oli aikarajoitettu (ks. edellä testinsarjan teknisiä ratkaisuja).

Kolmanneksi koska testin kulkua ei aina pystytä kontrolloimaan, oppilaat saattavat edetä testissä omaan tahtiinsa ja ehkä jättää osan tehtävistä suorittamatta pelkästään testissä tapahtuneen navigointivirheen vuoksi. Näin voi käydä myös, mikäli testattava ei ymmärrä ohjeita ja ei hahmota toimintalogiikkaa tai rakennetta, ymmärtää jotain väärin, eikä ehkä motivoidu jatkamaan tehtävien tekemistä. Näin voi tietenkin käydä myös paperi-kynä-testissä, mutta digitaalisessa testissä eri testiosuuksissa navigointi ja puuttuvien vastausten tarkistaminen on kankeampaa kuin paperi-kynä-testissä. Tässä mittauksessa kontrollointia ei nähty tarpeelliseksi toisin kuin esimerkiksi aiemmassa 1. luokan tiedonkeruussa (Ukkola & Metsämuuronen, 2019), jossa testissä eteneminen oli ohjattu siten, että seuraavaan tehtävään ei päässyt ennen kuin oli vahvistanut aiemman tehtävän ratkaisun. Kontrolli olisi saattanut tuntua 9.-luokkalaisista nuorista liian rajoittavalta; tehtäväsarjassa ei olisi pystynyt liikkumaan vapaasti omien mieltymisten mukaisesti.

Neljänneksi yleensä digitaalisessa testauksessa ei ole mahdollista hallita käyttäjän työskentelytapaa ja testin tekemisen intensiteettiä. Testin valvojan on haasteellista tietää, suorittavatko oppilaat testaushetkellä ainoastaan testiä vai myös jotain muuta oheistoimintaa kuten television katselua tai mobiililaitteen kanssa puuhailua tai vastaavaa. Tämä linkittyy suoraan myös testin luotettavuuskysymyksiin (ks. luku 3.5.5 ja edellä luku 2.2.4).

3.5.3 Tasa-arvokysymykset

Näyttää siltä, että jotkut oppilaat ja opettajat eivät olleet teknisesti eivätkä henkisesti valmiita COVID-19-pandemian pakottamaan digiloikkaan (Goman ym., 2021). Toisaalta moni oppija, jolla oli korkea motivaatio ja vahva tuki kotoa, sai etua etäopinnoista. Kotien vaikutus on tutkimuksissa korostunut eräänä etäopiskeluun liittyvien ongelmien avainselittäjänä (ks. mm. Blikstad-Balas ym., 2022; Lavonen & Salmela-Aro, 2022; Sjögren ym., 2021). Emme myöskään tiedä, missä määrin huoltajilla digitalisoidun maailman kansalaisten vaaditut taidot ovat kehittyneet vastaamaan koulutuksen ja digitaalisen testauksen tarpeita. Tämä ei näytä olevan ongelma peruskouluikäisillä lapsilla ja nuorilla Suomessa, joilta käytännössä kaikilta löytyy internetyhteyksillä varustettu mobiililaite, ja siihen liittyvät toiminnot ovat heille ainakin päällisin puolin tuttuja. Tietenkin käyttäjäkohtaiset erot voivat olla suuria.

Tasa-arvoon testitilanteissa on kiinnitetty huomiota esimerkiksi alkuperäiskansoihin liittyvän vinouman näkökulmasta (”*cultural bias*”; ks. Litts, Searle, Brayboy, & Kafai, 2020; ks. myös keskustelu alkuperäiskansojen oman kulttuurisen materiaalin tuomisesta osaksi mittaamista ja opetusmateriaalia, Eglash ym., 2020), joka voisi olla relevantti näkökulma myös Suomessa esimerkiksi saamelais-, romani- tai maahanmuuttoyhteisöissä. Esimerkiksi huonosti valmistellut tehtävät tai tehtävän viitekehykset saattavat olla kulttuurisesti vinoutuneita ja vieraita tehtävään vastaajalle. Voidaan kuvitella asia päinvastaisesti: jos esimerkiksi matematiikan tehtävänä käytettäisiin ettoamisen ja pykällyksen yhteydessä havaittujen kermikköjen määrän laskemista paliskuntien keskiarvotietona, joku valtaväestöön kuuluva testattava sanoisi, että testi on vinoutunut suosimaan saamelaisia ja erityisesti poronhoitaja-saamelaisia.

Relevantti näkökulma on myös saavutettavuushaaste (ks. Shaheen, 2021): monet teknologiset ratkaisut eivät sovellu erilaisille erityisryhmille kuten esimerkiksi näkö-, kuulo-, hienomotoriikkarajoitteisille tai hahmotushäiriöisille oppilaille. Näihin voidaan löytää ratkaisuja, mutta kansallisella tasolla digitaalinen testaaminen on niin alkuvaiheissaan, että erityisryhmien erityistarpeet tulevat huomioiduksi vasta ajan myötä, vaikka tahtoa asiaan olisikin, eikä jokaisen oppilaan yksilöllisiä tarpeita pystytä huomioimaan. Vaikka siis esimerkiksi lainsäädäntö ohjaa käytettävyyden ja saavutettavuuden huomioimiseen (https://www.finlex.fi/fi/laki/alkup/2019/20190306), käytännössä ohjeita tullaan antamaan vielä pitkään siitä, kenelle arvioinnissa käytetty testaus soveltuu; esimerkiksi erityiskoulut ja kansainväliset koulut rajattiin matematiikan arvioinnin ulkopuolelle, koska arviointitesteistä ei ole järkevää tehdä hyvin erityisiä versioita kaikille oppilaille soveltuviksi.

Tasa-arvoon liittyen voidaan oikeutetusti kysyä, kuinka voisimme olla varmoja siitä, että digitaalinen testausjärjestelmä ei suosi niitä, joilla on mahdollisuus sijoittaa tietokoneisiin ja ohjelmistoihin ja maksaa lisenssimaksuja näihin asioihin erikoistuneille yrityksille. Tämä ei välttämättä ole ongelma kouluissa, jossa kaikille periaatteessa annetaan samanlaista opetusta. Kysymys saattaa olla aiheellinen erilaisista kotiolosuhteista tulevilla oppijoilla: onko kaikissa perheissä varaa laitteisiin ja ohjelmistoihin, joilla keskeisiä taitoja voidaan harjoitella? Voisiko olla niin, että digitaalinen testaus suosii tahattomasti jotain ryhmää: rikkaampia tai koulutetumpia perheitä, tyttöjä tai poikia, maan joitakin alueita tai jotakin kieliryhmää? Näitä ei juuri ole tutkittu Suomen oloissa—eikä juuri muuallakaan, sillä laajamittaisen digitaalisen testauksen aika on vasta alkamassa. Jos suosimista tapahtuu, kuinka mahdolliset epätasa-arvoon liittyvät seikat tulisi tarkistaa ja kenen toimesta? Mikä on hallituksen, opetus- ja kulttuuriministeriön, koulutuksen järjestäjän, perheen ja yksilön vastuu, kun kaikille oppijoille tulisi tarjota tasavertaiset mahdollisuudet päästä omaan potentiaaliinsa (ks. keskustelu Metsämuuronen & Lehikko, 2022; Metsämuuronen & Seppälä, 2021)?

3.5.4 Testikäyttäytymisen muutokseen liittyvät kysymykset

Keskeinen haaste digitaaliseen testaukseen siirtymisessä on paperi-kynä-testien ja digitaalisten testin tulosten vertailukelpoisuus. Tätä ns. moodiefektiä (*Mode effect*) on tutkittu paljon (ks. mm. Fisbein ym., 2018 TIMSS-testien osalta; IELTS, 2007 kielitestien osalta; Jerrim ym., 2018 PISA-testien osalta). Kun PISA-testauksessa siirryttiin digitaaliseen testaukseen vuonna 2015, huoli kohdistui tulosten vertailukelpoisuuteen. Tämä on oleellinen kysymys, sillä PISA:n kantava ajatus on nimenomaan osaamisen muutos eri mittauskertojen välillä. Kolmen maan (Saksa, Ruotsi ja Irlanti) esitestausaineiston perusteella havaittiin, että huoli oli aiheellinen. Edes korjausten avulla ei saavutettu tulosten täyttä vertailukelpoisuutta (ks. Jerrim ym. 2018). Myös TIMSS-aineiston analyysissa päädyttiin johtopäätökseen, että osioanalyysissa jouduttiin tekemään korjauksia (Fishbein ym., 2018). Toisin sanoen, testin moodin vaihtamisella *on* vaikutusta osaamisen arviointiin.

Välineestä aiheutuva osaamisen muutos

Hyvin varhaisissa *mode effect* -tutkimuksissa havaittiin, että sekä sisällön että välineiden käytön tuttuus, kilpailuhenkisyys ja sukupuoli olivat tekijöitä, jotka erottelivat paperi-kynä-testin ja digitaalisen testin suorituksia (ks. Clariana & Wallace, 2002; ks. kirjallisuutta ja tarkempia tuloksia IELTS, 2007). Näyttää siltä, että ilmiö on kuitenkin muuttunut tietotekniikan kehittymisen ja taitojen lisääntymisen myötä, ja se saattaa vaihdella yhteiskuntakohtaisesti. Esimerkiksi vuoden 2019 TIMSS-aineistossa ei Alankomaissa havaittu eroja eri testityyppien välillä (Hamhuis, Glas, & Meelissen, 2020), vaikka koko aineistossa parametreja jouduttiin korjaamaan, kuten edellä todettiin (ks. Fishbein ym. 2018). Toisessa tutkimuksessa, jossa testattavat saivat itse valita, kumpaa testimuotoa he käyttivät, digitaalisen testimuodon valinneet menestyivät *paremmin* kuin paperi-kynä-testin valinneet (Nardi & Ranieri, 2018); tämän osalta muna-kana-haaste on ilmeinen: valitsivatko paremmat oppilaat mieluummin digitaalisen testin? Suomessa vuoden 2015 matematiikan mittauksessa osaaminen digitaalisessa testissä oli merkittävästi heikompaa kuin paperi-kynä-testissä (Julin & Rautopuro, 2016) ja kansallinen trendi vuoden 2021 mittauksessa jatkoi nimenomaan digitaalisessa testauksessa ilmennyttä laskevaa trendiä (Metsämuuronen & Nousiainen, 2021).

Vuoden 2021 matematiikan mittauksessa havaittiin aivan ilmeinen *mode effect*, jota kuvataan toisaalla tarkemmin (Metsämuuronen & Nousiainen, 2023; ks. luku 2): oppilaiden testikäyttäytyminen poikkesi oleellisesti aiemmista mittauksista. Ensiksi puuttuvia tietoja oli yleisesti ottaen selvästi enemmän kuin aiemmissa paperi-kynä-testeissä. Osittain tämä selittyi teknisillä ongelmilla, kuten mahdollisilla häiriöillä koulun internetverkossa. Erityisesti kouluarvosanaltaan hyvin suoriutuneiden oppilaiden puuttuvien tietojen aiempaa suurempi määrä johti joidenkin oppilaiden poistamiseen aineistosta, koska saatu pistemäärä ei ilmeisestikään kuvannut heidän todellista osaamistaan. Käytännössä, kun aiemmissa paperi-kynä-testeissä arvosanan 10 saaneet oppilaat jättivät vastaamatta 1–2 prosenttiin tehtävistä, vuoden 2021 digitaalisessa testauksessa parhaat oppilaat jättivät vastaamatta keskimäärin 6 prosenttia tehtävistä—osuus oli siis yli kolminkertaistunut. Arvosanan 9 saaneilla oppilailla tyhjiä vastauksia oli keskimäärin lähes kaksinkertainen määrä aiempaan nähden. Yksittäisillä, arvosanan 10 saaneilla oppilailla oli aiemmissa paperi-kynä-testeissä puuttuvia tietoja korkeimmillaan 15–20 % tehtävistä. Vuoden 2021 aineistossa luku oli 45 % eli oleellisesti korkeampi. Pääteltiin, että ainakin osa näistä oli seurausta epäonnistuneista arviointijärjestelyistä. Huomataan myös, että vuoden 2021 tehtäväsarjat olivat kokonaispistemäärän osalta yhtä pitkiä kuin aiempien vuosien tehtäväsarjat.

Puuttuvien tietojen tuoma haaste

Toinen, spesifimpi haaste syntyi siitä, että osa oppilaista *ei halunnut tai osannut vastata perustelutyyppisiin tehtäviin*. Jos tehtävässä piti sekä antaa oikea vastaus (1 piste) että perustella se (1 piste), aiempaa huomattavasti useammalta oppilailta jäi perustelu antamatta. Opettajilta saatujen tietojen mukaan monet oppilaista olisivat kyllä osanneet varmasti antaa perustelun—mihin viittasi myös oikea ratkaisu, mutta syystä tai toisesta eivät olleet kirjanneet perustelujaan. Ilmiö oli hieman tyypillisempää heikoille ja keskitasoisille oppilaille, joskin luvussa 4 (Metsämuuronen & Suomilammi, 2023) huomataan, että myös parhaissa osaamisryhmissä perusteluja jätettiin antamatta selvästi aiempaa useammin. Seikalla oli vaikutusta osioparametrien määräytymiseen—samaan tapaan kuin raportoitiin PISA- ja TIMSS-hankkeissa (ks. Jerrim ym., 2018; Fishbein ym., 2018). Mainitusta syystä kalibrointivaiheessa jouduttiin poistamaan neljä linkkitehtävää (ks. luku 2, Metsämuuronen & Nousiainen, 2023). Yksityiskohtaisemmin linkkitehtävissä ilmenneitä eroja eri vuosien välillä käsittelevät luvussa 4 Metsämuuronen ja Suomilammi (2023).

Puuttuvien tietojen syntymisen mekanismit yleensä jaetaan neljään ryhmään: täysin satunnaisiin puuttuviin tietoihin (*Missing completely at random*, MCAR), satunnaisesti puuttuviin tietoihin (*Missing at random*, MAR), ei-satunnaisiin puuttuviin tietoihin (*Missing not at random*, MNAR) ja rakenteellisesti puuttuviin tietoihin (*Structurally missing*, SR). Kaikki voivat esiintyä aineistoissa myös eri kombinaatioina. Näistä haasteellisia ovat kaksi jälkimmäistä: tietoa ei voi mielekkäästi korvata perinteisillä mekanismeilla, koska oikeaan osuvaa arvoa ei ole helppoa mallintaa. Vuoden 2021 aineistossa haasteellisia nimenomaan nämä mekaanisista syistä syntyneet ei-satunnaiset puuttuvat tiedot. Osa puuttuvista tiedoista, kuten kuunteluosuudessa kokonaan puuttuvat sarjat, mallitettiin verrokkimenetelmää käyttäen. Osalla oppilaista (1 %) oli kouluarvosanaan nähden liian paljon puuttuvaa tietoa, minkä vuoksi heidät poistettiin aineistosta. Näitä käsittelevät tarkemmin luvussa 2 Metsämuuronen ja Nousiainen (2023).

Keskeinen kysymys on, kuinka testijärjestelyt ja itse testit ja testiosiot olisi tulevaisuudessa järkevä tehdä niin, että edellä kuvattujen ”turhien” puuttuvien tietojen määrä pysyisi mahdollisimman pienenä; tätä pohditaan luvussa 3.6. On ilmeistä, että jos testijärjestelmä itsessään ei houkuttele vastaamaan annettuihin tehtäviin, järjestelmää on varmaankin syytä kehittää. Tällä ei kuitenkaan ole välttämättä suurtakaan vaikutusta vastaamatta jättämiseen; varsin karkeillakin testimenettelyillä voidaan saada oppilaat vastaamaan annettuihin tehtäviin, jos he ovat motivoituneita vastaamaan. Toisaalta mikäli ongelmat paikallistuvat matematiikkaeditorin käytössä ilmenneisiin vaikeuksiin, on järkevää varata riittävästi aikaa sen harjoittelemiseen ennen varsinaista testausta.

Alhainen motivaatio low stake -testauksessa

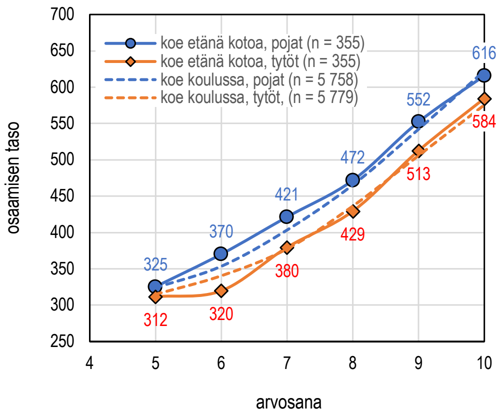
Testijärjestelmään liittyvät seikat voivat olla selvästi toissijaisia testiin vastaamisen suostuttelussa. Kokemukset Karvin oppimistulosarviointien kaltaisista ns. *low-stake* -testauksista, joilla ei ole juuri merkitystä esimerkiksi jatko-opintojen kannalta—toisin kuin esimerkiksi ylioppilaskokeella, joka on tyypillinen ns. *high-stake* -testi, joka vaikuttaa oleellisesti jatkokoulutukseen hakeutumiseen—osoittavat, että *mikäli opettaja ei kykene motivoimaan oppilaitaan osoittamaan parasta osaamista, tehtävien suorittamiseen saatetaan asennoitua välinpitämättömästi*. Tästä syystä opettajia on perinteisesti rohkaistu käyttämään Karvin testiä oman arviointinsa tukena, jolla voi olla päättöarvosanaa nostava vaikutus, mikäli testissä suoriutuminen on aiempaa näyttöä korkeammalla tasolla. Digitaalisiin kansallisiin arviointeihin liittyy tässä mielessä ilmeinen haaste, että yhtäältä oppilaan oikeusturvan vuoksi arvosanan perustelun tulee olla varmennettavissa, mutta toisaalta kansallisten testien tehtävät eivät ole julkisia, jotta näiden vertailukelpoisuus mahdollistuisi, kuten edellä keskusteltiin. Joka tapauksessa jokainen oppilas on nähnyt, millaisista tehtävistä tehtäväsarja on muodostunut. Tällöin asia ei välttämättä muodostu ongelmalliseksi, mikäli arvosanaehdotuksen perustelu on muutoin avointa ja testi on rakennettu opetussuunnitelman perusteiden mukaiseksi—mihin aina kansallisissa mittauksissa pyritään. Asiaan liittyviä haasteita ja ratkaisuja tullaan miettimään Karvissa tulevien arviointien yhteydessä.

3.5.5 Etätestaus ja uskottavuusongelma

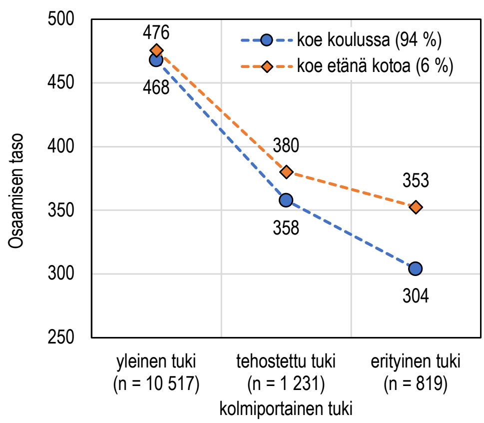
Testitulos ei ole koskaan ”totuus” absoluuttisessa mielessä. Osaamisen mittaamiseen liittyy oppijalähtöisten tekijöiden lisäksi aina useita muita virhelähteitä, joista yksi on ns. pisteittäjävirhe (*rater error*). Digitaalisessa testaamisessa on mahdollista käyttää automaattista pisteitystä osassa tehtäviä, jotka aiemmin on pisteittänyt opettaja, ja tämä vähentää pisteittäjävirheen määrää. Tältä osin digitaaliset testit voivat yhtäältä olla luotettavampia kuin opettajien tai sensoreiden pisteittämät testit.

Toisaalta erityisesti etätestauksessa *testivilpin mahdollisuus lisääntyy*. Erityisesti etätestauksessa on mielekästä kysyä, mistä tiedämme, että oikeat testattavat todella tekivät testin? Mistä tiedämme, että testattava ei lukenut oikeaa vastausta kirjasta tai Internetistä tai että asiantuntevampi oppilas, sisar tai huoltaja ei auttanut testitilanteessa, mikäli yksilösuoritus oli arvioinnin kohteena?

Vuoden 2021 matematiikan aineistossa keskiosaaminen oli 18 pistettä parempaa niillä oppilailla, jotka tekivät testin etänä kotoa käsin (464 pistettä; n = 710) kuin niillä, jotka tekivät testin ainakin jokseenkin kontrolloiduissa olosuhteissa kouluissa (447, n = 11 537). Ero on merkitsevä mutta ei merkittävä. Etätestaus suosi erityisesti poikia (+30 pistettä) kun taas tyttöjen osalta ero ei ollut merkittävä (+6 pistettä). Näyttää siltä, että erityisesti heikommin suoriutuneet pojat (arvosana 6 ja 7) ovat etätestauksessa tuottaneet kontrolloitua testiä hieman paremman tuloksen (Kuvio 18). Heikosti suoriutuneet tytöt (arvosana 6) puolestaan olisivat ehkä hyötyneet kontrolloidusta testijärjestelystä; kotona tehtynä tulos oli selvästi heikompi kuin koulussa tehtynä. Toinen ryhmä, joka erityisesti hyötyi etätestauksesta, olivat erityistä tukea saaneet oppilaat (Kuvio 19): tulos oli merkittävästi korkeampi kotoa käsin etänä tehtynä (+49 pistettä) kuin kontrolloidusti tehtynä. Oppilasmäärät ovat kuitenkin niin pieniä, että näillä ei ole juuri merkitystä kansallisen tuloksen näkökannalta.



KUVIO 3.18. Etänä ja koulussa suoritetussa testissä suoriutuminen tytöillä ja pojilla



KUVIO 3.19. Etänä ja koulussa suoritetussa testissä suoriutuminen kolmiportaisen tuen tasoilla

Yhtäältä Karvin testien *low-stake* luonne ei ole suosinut huijaamista—testivilpistä ei juuri ole hyötyä. Toisaalta haluaisimme oppilaiden tekevän testin vakavasti, jotta saisimme tietää oppilaiden todellisen osaamisen tason. Näyttää siltä, että uskottava etätestaus edellyttää viestintää ja testin kontrollointia (ks. Hannu ym., 2010). Etätestien kontrolloinnin mekanismeja vielä kehitellään—opettajat tarvitsevat näitä joka tapauksessa omien koulukokeidensa kontrolloituun toteuttamiseen, jos ja kun etäkoulu tulee tulevaisuudessa osaksi koulutuksen monimuotoista toteuttamista eli sitä, että oppimisesta tulee ajasta ja paikasta riippumatonta, ”kaikkiallistumista”, kuten esimerkiksi Nyyssölä ja Kumpulainen (2020) tätä kutsuvat (ks. luku 3.6). On ilmeistä, että haluamme saada tämänkaltaisen virhelähteen mahdollisimman pieneksi, joskin on mahdollista, että siitä ei täysin päästä eroon—testivilppiä, huijaamista ja lunttausta on aina ollut eikä siitä ole helppo päästä eroon.

3.6 Näkymiä digitaalisen testaamisen tulevaisuuteen

Digitaalinen testaus on tullut jäädäkseen, ja siksi joitain potentiaalisia tulevaisuuden trendejä voi olla hyvä pohtia tässä yhteydessä kansallisen osaamisen arvioinnin viitekehyksessä. Karvin oppimistulosarviointien käyttöön ollaan kehittämässä uutta arviointijärjestelmää, jonka on tarkoitus valmistua vuoden 2024 mittauksia varten. Uuden alustan kehittämisessä on huomioitu joustavuus, jotta myös tulevaisuuden tarpeet voidaan helposti huomioida myöhemmissä testeissä. On melko helppo ennustaa, että tulevaisuuden digitaaliseen testaukseen vaikuttavat samat megatrendit kuin muuhunkin muutokseen. Tulevaisuutta on tunnetusti hankala ennustaa, mutta sivistyneitä arvauksia voidaan tarjota keskustelun tueksi.

Keskeisten koulutusta koskevien megatrendien kartoittamisessa hyödynnetään ennakointitietoa kolmesta lähteestä. Yhtäältä osana kansallista koulutuksen kehittämistä, kansallinen osaamisen ennakointiverkosto kokoaa säännöllisin väliajoin koulutusalan toimijoita pohtimaan koulutuksen tulevaisuutta (ks. mm. OPH, 2018; Nyyssölä & Kumpulainen, 2020). Osittain foorumin skenaariotyöstä ja osittain muiden toimijatahojen tuottamista ajatuksista voi saada käsityksen siitä, mihin suuntaan koulutuksen sosiaalisen ympäristön arvioidaan muuttuvan. Foorumin näkemyksiä ja ennusteita ovat tiivistäneet Nyyssölä ja Kumpulainen (2020).

Foorumin näkemyksen mukaan tulevaisuuden isoa kuvaa vuonna 2035 määrittävät digitalisaatio, työn murros ja ekologisen ajattelun voimistuminen. Koulutuksen kehitystä yleisellä tasolla leimaa oppimisen ajasta ja paikasta riippumattomuus eli kaikkiallistuminen ja yksilöllisten oppimisympäristöjen ja opetusteknologian laaja hyödyntäminen. Alueelliset erot kasvavat. Vuoteen 2035 mennessä peruskouluverkon ennustetaan väestökehityksen myötä harventuneen osin kustannussyistä ja osin sen vuoksi, että ennaltaehkäisevät ja diagnosoivat järjestelmät sekä digitaalisten opetus- ja oppimisjärjestelmien kehitys ovat tehneet sen mahdolliseksi. Vaikka kouluverkon harveneminen jatkuu yleiselle tasolla tulevaisuudessa, kehityssuuntaa ennustetaan muutettavan eri alueilla poliittisilla päätöksillä, jotta koulutuksen saavutettavuus osana julkisten palveluiden saavutettavuutta turvataan koko maassa. On mahdollista, että oppimisen digitalisoituminen, etäopetusratkaisujen lisääntyminen sekä yhteisopettajuuden lisääntyminen vähentävät opettajien tarvetta. (Nyyssölä & Kumpulainen, 2020.)

Toinen, erityisesti teknologiaan liittyvien megatrendien peruslähde on Heikkilän artikkeli (2021), johon on koottu erityisesti teknologian megatrendejä. Näitä ovat mm. tekoälyn kehittyminen, esineiden internetin ekosysteemi kehittyminen, pilvipalveluiden kehittyminen ja *Big data*n laajamittaisempi hyödyntäminen. Opetuksessa ja koulutuksessa uudenlaista älykästä seurantaa ja ohjausta sekä *big data*n käyttöä voidaan kehittää yhtäältä yksilön reaaliaikainen oppimisen tueksi (esimerkiksi seurataan keskittymistä tai aktivoidaan palkkioilla sopivin väliajoin kuten pelimaailmassa) ja toisaalta pidempiaikaisen seurannan tukena, jota sekä yksilön että populaation tasolta kerätyllä tiedolla voidaan kehittää oppimistehtäviä, suunnitelmia tai oppimisprosesseja. Näitä käsitellään luvussa 3.6.5.

Kolmas, erityisesti pidemmän aikavälin potentiaalisten trendien lähde on tulevaisuustutkimukseen ja strategiseen suunnitteluun erikoistuneen *Futures Platform* -ajatuspajan tuottamat analyysit koulutuksen pitkäaikaisista trendeistä.[20] Seurattavia trendejä ovat mm. metaversumiin (*metaversum*) liittyvät kehityspolut, tekoälyn ja lisätyn todellisuuden (*augmented reality*) tuomat mahdollisuudet koulutuksessa sekä yhdistelevään tai sulautettuun oppimiseen (*blended learning*) liittyvät trendit. Näitä ovat käsitelleet mm. Sandal (2021), Veltheim (2021) ja Ray (2022), ja niitä esitellään tarkemmin luvussa 3.6.7.

Tulevissa luvuissa edellä mainittuja tulevaisuuden vaihtoehtoja ja -kuvia käsitellään APESTE-viitekehyksessä (Metsämuuronen, 2001b; perustuen Meristö, 1993) "asiakkaissa", politiikassa, taloudessa, sosiaalisessa ympäristössä, teknologiassa ja ekologisissa tekijöissä odotettavien muutosten näkökannalta. Kaikkia mahdollisia seikkoja ei tietenkään ole mahdollista käsitellä tässä yhteydessä, mutta joitain kiinnostavia näkökulmia kustakin suunnasta otetaan esille.

3.6.1 Asiakkaista lähtevät muutostarpeet

Digitaalisiin testeihin liittyy kahdenlaisia asiakkaita: instituutioasiakkaita ja henkilöasiakkaita. Kun koulutuksen maailma muuttuu, myös asiakkaat ja heidän ajattelunsa muuttuvat—ja tietenkin myös päinvastoin: kun oppijat ja koulutuksen hallinto muuttuvat, myös maailma muuttuu heidän kanssaan. On helppo ennustaa, että osaamisen arvioiminen ja diagnosointi, johon yleensä liittyy osaltaan testaamista, ei ennusteiden mukaan vähene. Päinvastoin sen merkitys voi jopa kasvaa, kun teknologiset ratkaisut mahdollistavat etäkoulu-tyyppiset ratkaisut.

Instituutioasiakkaiden odotukset

Instituutioasiakkaat odottavat testausjärjestelmiltä käyttövarmuutta, ennakoivuutta, helposti käyttöön otettavuutta ja intuitiivisuutta siinä mielessä, että osioiden ja testien laatiminen ja niiden hallinnointi ei saa olla liian monimutkaista ja epäintuitiivista. Testausjärjestelmissä tulee olla mahdollisuus osioiden pankitukseen (*item banking*) ja osioiden pankista valitsemiseen (*item retríeval*) ja testiversioiden statistiikkaan liittyvien ennakkotietojen kuten osioiden ja koko testin vaikeustasojen, sisältöalueiden, erottelukykyjen ja reliabiliteettien tarkastelemiseen. Järjestelmästä tulee pystyä myös tulostamaan testejä mahdollisia paperi-kynä-testiversioita silmällä pitäen. Järjestelmän tulee pystyä automaattiseen pisteitykseen ja mahdollisesti suoraan muuttamaan pistemäärät arvosanaehdotuksiksi tai diagnostisissa mittauksissa kertomaan, millä osaamisalueilla olisi tarve lisäharjoitteluun. Tämä edellyttää uudenlaisia algoritmeja tai jo olemassa olevien algoritmien liittämistä järjestelmään. Edellä kuvatun kaltaiset ominaisuudet eivät ole varsinaisesti tulevaisuutta, vaan toteutuvat pitkälti jo nykyisissä ja kehitteillä olevissa järjestelmissä.

Suomen oloissa kansallisen tason testauksessa tulevaisuuteen sijoittuvat adaptiivisen testin mahdollisuudet, joita monissa maissa jo käytetäänkin. Tämän on havaittu tehostavan testiin käytettävää aikaa; testiin tarvittava aika saattaa puolittua, mutta tarkkuus säilyy samana kuten

20 Tulevaisuustutkija Tuomo Kuosa *Futures Platform* -ajatuspajasta luki tekstin ulkopuolisena arvioijana ja antoi hyödyllisiä vinkkejä siitä, millaiset tekijät tulevat pidemmällä aikavälillä viitoittamaan koulutuksen tulevaisuutta.

edellä todettiin (ks. mm. Van der Linden & Glas, 2000; Wainer, 2000; Weiss & Kingsbury, 1984; ks. myös Vie ym. 2017). Adaptiivisen testauksen käyttö ei kuitenkaan ole ongelmatonta, kuten Tanskan esimerkki osoittaa (ks. Kreiner, 2021, Regeringen, 2021); erityisesti diagnostisena työkaluna adaptiivinen testaus ei nykymuodossaan ole paras vaihtoehto. Siihen liittyviä algoritmeja olisi ehkä syytä kehittää siihen suuntaan, että osion valinnassa huomioituvat paremmin myös erilaiset tekijät, joiden avulla olisi mahdollista antaa osaamisen tason lisäksi myös laadullista tietoa testattavan osaamisesta.

Tulevaisuudessa myös suuri osa avovastauksista pisteitetään automaattisesti. Tämä on ollut käytössä jo yli 20 vuotta esimerkiksi USA:ssa (ks. kirjallisuutta mm. Dikli, 2006), ja Suomessakin keskustelua on käyty ainakin vuodesta 1997 lähtien, jolloin kansallinen arviointitoiminta alkoi Opetushallituksen erillisyksikkönä ennen Karvin perustamista. Suomen kielen syntaksi on haasteellinen, mikä on hidastanut automaattisen pisteitysjärjestelmän kehittymistä. Paraikaakin on kuitenkin käynnissä hankkeita, jossa suomen kielen syntaksia pyritään ottamaan haltuun tekoälyä varten (ks. eri kieliresursseista esimerkiksi Jauhiainen, Lennes, & Marttila 2019; ks. myös https://www.kielipankki.fi/tyokalut/). Jossain vaiheessa näitä työkaluja otettaneen osaksi kansallista oppimistulosarviointia.

Yksilöasiakkaiden odotukset

Yksilöasiakkaat—kansallisen oppimistulosarvioinnin tapauksessa oppilaat, opiskelijat ja opettajat—tuovat oman haasteensa testijärjestelmien kehittämiseen. Oppilaiden näkökannalta haaste on kehittää testijärjestelmiä suuntaan, joka automaattisesti ohjaisi oppilaita tekemään testit motivoituneesti ja vakavissaan. Mahdollisesti tämä johtaa siihen, että testijärjestelmään lisätään niin videomaisia, pelillisiä ja tarinallisia kuin myös immersiivisiä piirteitä. Nämä tekijät ovat oleellisia silloin, kun itse e-oppimisympäristöä—ei vain testi-ympäristöä—kehitetään pelilliseen suuntaan: oppimisesta tehdään peli, luodaan oppilaiden tarpeisiin mielekästä sisältöä, pyritään saavuttamaan oppijoiden huomio ja pitämään sitä yllä (esimerkiksi rakentamalla opetukseen yllätyksellisyyttä kuten esimerkiksi virtuaalisen laatikon muodossa, joka satunnaisesti valitsee opettajan kysymykseen vastaavan oppijan nimen) sekä kun tietoisesti rakennetaan e-oppimisympäristöjä sellaisiksi, että se johtaa tavoitteelliseen oppimiseen (ks. Miller, 2021).

Liikkuvaan kuvaan liittyvät tekniset haasteet olivat edellisen sukupolven digitaalisessa testauksessa ilmeisiä, koska esimerkiksi videokuvan mahdollistavat väylänopeudet olivat pieniä. Tämä ei ole suuri ongelma enää, vaikka osalla kouluista tietoverkko vielä vuosien 2020 ja 2021 perusopetuksen arvioinneissa oli selvästi vanhentunut (ks. luku 3.6.5. Teknologian muutoksesta nousevista muutostarpeista). Pelillisyyttä ja tarinallisuutta käytettiin jo Karvin aiemmassa arviointialustassa ja tätä edeltäneissä paperi-kynä-tehtäväsarjoissa. Näitä elementtejä voidaan lisätä esimerkiksi siten, että tehtävät liittyvät toisiinsa taustatarinan muodossa (ks. esimerkiksi vuoden 2005 3. luokan mittauksen yhteydessä koko mittauksen taustatarina eläintarhassa olemisesta, Huisman, 2005; Huisman & Silverström, 2005). Samanlaista tarinallisuutta suunniteltiin myös 6. luokan Finskan mittaukseen Karvin arviointialustassa (Åkerlund ym., 2019). Vuodesta 2018 lähtien luokilla 3–9 arviointien pohjana oli kartta, jossa oppilas pääsi ”seikkailemaan” tehtäviin haluamassaan järjestyksessä, ja näin saatiin edes pientä autonomian tuntua. Myös väritykset oli suunniteltu niin, että niistä ilmenivät jo suoritetut tehtävät. Ensimmäisen luokan mittauksessa koulutulokkaat saivat valita pelihahmon, joka ”kertoi tehtävät oppilaille” eli josta painamalla sai ääniohjeistuksen. Pelilliseen oppimisympäristöön voidaan yhtäältä lisätä testaukseen liittyviä piirteitä (ks. kirjallisuutta mm. Kaarakainen, Kivinen, & Hutri, 2015). Toisaalta testien rakentaminen peleiksi ei välttämättä ole reaalinen vaihtoehto laajassa mittakaavassa kehitystyöhön liittyvän kalleuden vuoksi. Pelillisyyttä on kuitenkin tyypillisesti esimerkiksi tehtävissä, joissa kellon käydessä tulee löytää ongelman oikea ratkaisu tietyssä ajassa.

Opetusteknologioiden yhteydessä immersiivisyydellä tarkoitetaan sitä, että teknologia mahdollistaa läsnäolon tunteen virtuaalisessa tilassa (Cummings & Bailenson, 2016). Immersiivisille, upottaville virtuaalisille oppimisympäristöille on tyypillistä stereoskooppiset, 3D- tai 360-näkymät sekä muuta aistinvaraista tietoa ja vuorovaikutteisuutta hyödyntävät elementit (ks. Kuuluvainen ym., 2021). Jos oppilaat ja opiskelijat jo muutenkin ovat tai ovat siirtymässä enenevässä määrin immersiiviseen oppimisympäristöön (ks. Kuuluvainen ym., 2021), on helppo ennustaa, että myös osaamisen testaus tapahtuu tällaisessa järjestelmässä. Vaihtoehtoisesti teknologian kehittyessä immersiiviä elementtejä voidaan lisätä testiosioihin. Jo nykyisissä kielen osaamisen testeissä immersiivisiä elementtejä on sisällytetty esimerkiksi interaktiivisten videoiden muodossa, joilla testataan palvelutilanteessa kielen ymmärtämistä.

Immersiiviset elementit testijärjestelmässä mahdollistavat sen, että siirrytään lähemmäksi ns. primääritestausta, luonnollisissa ympäristöissä tapahtuvaa ongelmanratkaisua sekundääri- tai tertiääritestauksen sijaan. Jälkimmäisissä todellinen reaalimaailman ongelma—kuten esimerkiksi prosenttilaskun osaaminen, kun arvioidaan hinnanalennusten suuruuksia ostotilanteessa—muutetaan hypoteettiseksi, karkeilla kuvilla havainnollistetuksi tai sanallisesti ilmaistuksi tehtäväksi. Immersiivisessä versiossa testattava voidaan viedä virtuaalisesti kauppaan, erilaisia vaihtoehtoisia ostettavia tuotteita voidaan esitellä hintoineen, ja testattava voi valita tuotteet, jotka hän tehtävässä annetulla rahalla voisi hankkia.

Sen lisäksi, että immersiivisyys saattaisi tuoda tehtäviin reaalisen elementin, sillä voi olla motivoiva vaikutus testin suorittamiseen samalla tavalla kuin immersiivisen virtuaalisen oppimisympäristön on havaittu motivoivan oppilaita (ks. esimerkiksi Sattar ym., 2020).

Houkuttelevuuden haaste

Yksi tulevaisuuden haaste on kehittää testijärjestelmiä sellaisiksi, että ne houkuttelevat testattavia tuottamaan tuottamis-, avo- tai perustelutehtäviin vastauksia, kuten edellä keskusteltiin. Kansallisten äidinkielen ja kirjallisuuden arviointien yhteydessä on esimerkiksi huomattu, että pojilla on taipumusta tyytyä niukkoihin vastauksiin, mikä tuottaa alhaisempia pistemääriä testitilanteessa. Tämä piirre havainnollistui erityisesti vuoden 2021 matematiikan mittauksessa: vastaamatta jättäminen muutti koko testin luonnetta monen tehtävän osalta, kun perusteluosaan ei tehtävissä saatu vastauksia. Kyse voi olla osittain motivoitumattomuudesta, osittain osaamattomuudesta ja osittain testilaiskuudesta.

Kansallisen arvioinnin tarkoituksena on selvittää osaamisen tosiasiallinen tila. Mikäli oppilas ei halua näyttää osaamistaan testijärjestelmästä johtuvista syistä, järjestelmiä on syytä kehittää. Tietenkin aina testauksen yhteydessä on syytä miettiä, millä muilla keinoilla oppilaat saadaan suostuteltua näyttämään parasta osaamistaan. Lopullinen syy puuttuviin tai niukkoihin vastauksiin ei läheskään aina ole testijärjestelmässä itsessään; ilman ulkoista motivaatiota, kuten esimerkiksi sillä, että testituloksella voi olla vaikutusta arvosanaan, tekniset ratkaisut voivat jäädä tehottomiksi.

3.6.2 Politiikan muutoksesta nousevat muutostarpeet

Nykyinen aluepolitiikka ja tähän liittyvät megatrendit johtavat siihen, että pienempiä maaseutukouluja lakkautetaan ja opetus tapahtuu yhä suuremmissa yksiköissä alueiden keskuspaikkakunnilla ja kaupungeissa. Yhtäältä kaupungistuminen megatrendinä on johtanut siihen, että kasvukeskuksista etäällä olevien perheiden ja lasten yhtäläiset mahdollisuudet esimerkiksi kielten opiskeluun ovat heikentyneet; monipuolista kielten opetusta ei ole mahdollista järjestää kuin suuremmissa kaupungeissa, joissa on riittävä oppilaspohja opetusryhmien järjestämiseen (ks. Metsämuuronen & Seppälä, 2021). Toisaalta on myös poliittista tahtoa pitää koko Suomi asuttuna sekä edistää koulutuksen digitalisaatiota (ks. esimerkiksi Taivassalo, 2019). Nämä kaikki yhdessä johtanevat siihen, että etäopetukseen liittyvät ratkaisut sekä yleistyvät että muodostavat osan lähitulevaisuuden koulutuspalettia (ks. keskustelu Metsämuuronen & Lehikko, 2022; Metsämuuronen & Seppälä, 2021; Nyyssölä & Kumpulainen, 2020). Kun etäopetukseen liittyvät ratkaisut lisääntyvät, lisääntyy tarve myös etätestaamiseen, ja tämän myötä myös etätestaamiseen liittyvät luotettavuus ja uskottavuuskysymykset on otettava vakavasti, kuten edellä todettiin.

Etätestauksen yhteydessä on mielekästä kysyä, kuinka testattava voidaan identifioida uskottavasti? Tämä ei ole niinkään ongelmallista Karvin arviointikokeiden kaltaisissa *low-stake*-testauksissa, mutta ovat oleellisia testattavan tulevaisuuden kannalta tärkeissä *high-stake*-testauksissa. Kansainvälisissä testausjärjestelmissä on käytössä biometrisia tunnistusmenetelmiä: äänitunnistusta (mm. TOEFL 2022), sormenjälki-, käsiala- ja kasvotunnistusta (mm. IELTS, 2013, 2022) tai verkkokalvotunnistusta (mm. Shdaifat ym. 2020). On mahdollista, että myös ei-biometristen tunnistusmenetelmien käyttö lisääntyy. Tällöin toimiva ratkaisu voi olla esimerkiksi vahva tunnistus pankkitunnuksien tai mobiilivarmentimien avulla ainakin vanhemmilla oppijoilla.

Vaikka testattava saataisiin tunnistettua yksiselitteisesti, mutta testi tehdään etäolosuhteissa, mistä tietäisimme, että testattava ei lukenut oikeaa vastausta kirjasta tai Internetistä tai että asiantuntevampi oppilas, sisar tai huoltaja ei auttanut testitilanteessa? Tämä kysymys saattaa johtaa kahteen erilaiseen ratkaisuun: yhtäältä tiukentuneisiin kontrollimekanismeihin kuten esimerkiksi suljettujen järjestelmien kehittämiseen tai videoseurannan edellyttämiseen testijärjestelmissä tai toisaalta uuden tyyppisten tehtävien rakentamiseen, joissa tietoa *pitääkin* hakea kirjoista tai Internetistä tai tuottaa yhdessä muiden testattavien kanssa tai joissa tehtävät ovat niin soveltavia ja juuri tätä tarkoitusta varten luotuja, ettei niihin ole suoria vastauksia muualla. Tämänkaltaisissa tehtävissä osaaminen ymmärretään hyvin erilaisena kuin perinteisessä osaamistestissä; osaaminen käsitetään tällöin proseduraalisena taitona tai sosiaaliseen vuorovaikutukseen perustuvana osaamisena, eikä yksittäisen tiedon tai opittavan asian arvioiminen ole fokuksessa.

Jälkimmäisessä tapauksessa tulevaisuuden testijärjestelmiin saattaa liittyä algoritmisia ominaisuuksia, joilla arvioidaan esimerkiksi oikeiden hakusanojen määriä tai vuorovaikutuksen määrää ja laatua tiedonhankintaprosessin aikana.

3.6.3 Ekonomisten tekijöiden muutoksesta nousevat muutostarpeet

Testijärjestelmien kehitys on pitkälti riippuvainen rahoituksesta, joka puolestaan määrittyy tuotteen kaupallistamismahdollisuuksien perusteella. Yhtäältä on vaikea ennustaa, ottaako esimerkiksi Sitra tai muut suuret kansalliset tai kansainväliset rahoittajat kuten Maailmanpankki testausjärjestelmien kehittämisen rahoituskohteiksi. COVID-19-pandemian seurauksena syntyneen globaalin koulutuskriisin näkökulmasta (ks. YK, 2020) tämä voi olla mahdollista, ellei jopa todennäköistä kehittyvissä maissa, mutta epätodennäköistä Suomen olosuhteissa. Toisaalta kansallisten ratkaisujen kehittäminen—kuten Ylioppilastutkintojärjestelmän ja Karvin uuden digitaalisen järjestelmän kehittäminen—on ollut riippuvaista budjettirahoituksesta. On kuitenkin epävarmaa, rahoitetaanko myös jatkossa järjestelmien uudistamista ja modernisointia tai edes ajantasaisuutta. Tällöin kehityskustannukset on mahdollisesti kyettävä rahoittamaan kaupallisilla sovelluksilla.

Kaupallisten sovellusten kehittäminen on asiakkaiden tarpeista lähtevää (ks. edellä): ei ole kannattavaa kehittää jotain, mikä ei tuota mahdolliselle asiakkaalle lisäarvoa ja mistä asiakasta ei valmis maksamaan. Toisaalta joskus asiakas on halukas maksamaan siitä, että tietty haluttu ominaisuus toteutuu testijärjestelmässä. Näin esimerkiksi Karvin vuoden 2021 testausta varten ViLLE-järjestelmään kehitettiin uusia ominaisuuksia Karvin toiveet huomioiden. Yleisesti ottaen kuitenkin testijärjestelmiä kehitettäessä on tunnistettava asiakkaan tarpeet.

Opettajien mahdollisesti vähenevä määrä (ks. edellä Nyyssölä & Kumpulainen, 2020) ja jäljelle jääneiden suurentunut työmäärä, opetuksen ja oppimisen paikasta ja ajasta riippumattomuus eli ”kaikkiallistuminen” myös opettajan työssä, saattavat johtaa siihen, että merkityksellisen lisäarvon tuottajaksi nousevat automaattiseen pisteitykseen, opettajan kannalta helppokäyttöisyyteen ja testien pituuteen ja tehokkaaseen ajan käyttöön liittyvät elementit. Tämä voi tarkoittaa adaptiiviseen testaukseen liittyvien elementtien lisäämistä testijärjestelmiin—näiden on todettu vähentävän testin pituuden puoleen ja vähentävän opettajan työmäärää, mikäli automaattisen pisteitykseen liittyvät seikat on huomioitu, kuten edellä todettiin. On myös mahdollista, että diagnostisiin ja oppimista edistäviin elementteihin liittyvät kehittämistarpeet lisääntyvät; näillä elementeillä saattaa olla kansallisesti merkittävää koulutuksen kehittämiseen vaikuttavaa lisäarvoa. Näitä pohditaan tarkemmin luvussa 3.6.5.

3.6.4 Sosiaalisen ympäristön muutoksesta nousevat muutostarpeet

Yksilöllistyneet oppimisympäristöt, oppimisen kaikkialla läsnä olemisen ja saavutettavuuden vaade sekä teknologian laaja hyödyntäminen vienevät testaamista siihen suuntaan, että paikkasidonnaisten testausmenetelmien osuus tulee vähenemään. Tämänkaltaisia sovelluksia tietenkin on jo nyt käytössä, mutta on mahdollista, että tulevaisuuden testijärjestelmiin alun perinkin on järkevää suunnitella mobiilitestaamisen mahdollistavia elementtejä. Mobiilitestit ovat kuitenkin varsin rajallisia pelkästään jo ruututilan vuoksi ja saattavat edellyttää hyvin toisenlaista testaamisajattelua kuin laajemmalle ruudulle tarkoitettavat testit, joihin voidaan vaivatta lisätä kuvallisia ja tekstillisiä elementtejä testattavan vastausnäkymän siitä kärsimättä. Mobiilitestauksessa haasteeksi saattavat tulla tuottamistyyppiset tehtävät ja jotka ovat monipolvisempia tai enemmän taustatietoa sisältäviä.

Tulevaisuuden diagnosoivat järjestelmät ja digitaalisten opetus- ja oppimisjärjestelmien kehitys vienevät kehitystä siihen suuntaan (tai ovat seurausta siitä), että testijärjestelmiä kehitetään diagnosoivaan suuntaan. Tällöin jo osioiden valmistelemisen yhteydessä voi olla järkevää pankittaa myös vääriin vaihtoehtoihin liittyvää metatietoa siitä, minkälaisia virheitä testattava teki, kun valitsi tietyn vastausvaihtoehdon esimerkiksi monivalintatehtävässä. Vaihtoehtoisesti algoritmeja kehitetään niin, että summapisteitä laskettaessa automaattisesti annetaan ohjeita siihen, millä alueilla olisi järkevää nähdä vaivaa, jotta testattavan osaaminen kehittyy tavoitteiden suuntaisesti. Tämänkaltaisia kaupallisia android-pohjaisia järjestelmiä on jo lukuisia kuten esimerkiksi Duolingo kielten testaamiseen (https://www.duolingo.com/) ja tehtäväsarjat ylioppilastutkintoon valmistautuville kokelaille (https://www.sanomapro.fi/sarjat/yo-kertaus/). Nämä ovat kuitenkin vielä harvinaisia kansallisen tason toimijoiden keskuudessa.

3.6.5 Teknologian muutoksesta nousevat muutostarpeet

Teknologian muutoksessa on jo pitkään ollut kolme perustrendiä: kaiken pienentyminen eli miniatyrisoituminen, teknologian sulautuminen kaikkeen ja tekoälyn liittyminen yhä voimakkaammin teknisiin ratkaisuihin (ks. mm. Metsämuuronen, 2001a; Metsämuuronen & Lehikko, 2022; Sitra 2020). Miniatyrisoituminen on näistä ilmeinen eikä sitä ole tarpeen käsitellä enempää. Heikkilä (2021) on koonnut kirjallisuudesta keskeisiä teknologian megatrendejä, ja näillä voi olla merkitystä myös testijärjestelmien kehittymisen kannalta. Kun digitaalisen testauksen tulevaisuuden näkymiä peilataan edellä kuvattuihin muutostrendeihin, on helppo nähdä tiettyjä muutossuuntia, joita jo edellä on tuotu esiin; näillä on taipumusta vaikuttaa samaan suuntaan.

Ensiksi Heikkilän (2021) mukaan tekoälyn (*Artificial Intelligence*, AI) kehittyminen nousee useissa raporteissa keskeisimmäksi megatrendiksi; AI on jo nyt käytössä ohjelmistokehityksessä ja sulautetuissa järjestelmissä. On helppo ennustaa, että testijärjestelmiin tullaan tulevaisuudessa kytkemään tekoälyä joko kehittyneiden algoritmien muodossa tai siten, että seuraavan sukupolven ratkaisut ovat aina edellistäkin kehittyneempiä. Tällä alueella kehitys on nopeaa. Tätä kirjoitettaessa (helmikuu 2023) vilkasta keskustelua on käyty esimerkiksi ChatGPT-sovelluksesta, joka kykenee tuottamaan älykkäitä tai älykkään oloisia vastauksia jopa vaikeisiin eettisiin ongelmiin. Tammikuussa 2023 uutisoitiin, että Jyväskylän yliopiston kauppakorkeakoulu loi ensimmäiset toimintaohjeet tekoälysovellusten ja kielimallien käytölle opintotehtävissä (https://www.jyu.fi/fi/ajankohtaista/arkisto/2023/01/jyvaskylan-yliopiston-kauppakorkeakoulu-sallii-tekoalyn-kayton-opiskelussa). Keskeinen perustelu oli, että ”jos nyt kielletään kielimallin käyttö ChatGPT:n kautta, niin kiellämmekö myös Googlen hakukoneen tai Wordin oikoluvun käytön, kun ne alkavat tuottaa vastauksia ja korjausehdotuksia samalla teknologialla?” Digitaalisen testaamisen näkökannalta älykkäät algoritmit voivat liittyä esimerkiksi yhä laadukkaammin tapahtuvaan avovastausten automaattiseen pisteittämiseen tai entistä älykkäämpien adaptiivisten ratkaisujen kehittämiseen. Laadukkaat oppimisdiagnostiikan työkalut edellyttävät myös älykkäitä ratkaisuja.

Toiseksi esineiden internetin (*Internet of Things*, IoT) ekosysteemi on kehittymässä 5G-verkon myötä. IoT mahdollistaa sen, että älykkäät sensorit integroidaan internetiin, jolloin laite voi reagoida muuttuvaan ympäristöön ja olla yhteydessä siihen, ja kaikkeen tähän liittyy kehittynyttä analytiikkaa. Jos IoT saavuttaa testausjärjestelmät tai kehitetään mielekkäitä ja riittävän edullisia teknisiä sovelluksia, on mahdollista, että testijärjestelmän osat reagoivat mielekkäästi eri komponenttien ja ohjelmistojen välillä, kokoavat informaatiota ja tuottavat diagnostista testattavista tietoa yli erilaisten testausjärjestelmien—opettajien valmistelemista koulukokeista, opettajajärjestöjen testeistä sekä kansallisista ja kansainvälisistä arviointitesteistä—oppijan hyödyksi. Älykkäissä järjestelmissä saattaisi olla mahdollista yhdistää kansalliseen matematiikan testitulokseen tietoa esimerkiksi matematiikkavaikeuksia koskevasta kansallisesta vertailuaineistosta (ks. FUNA, 2019; Räsänen ym., 2021). Osa tähän liittyvistä ratkaisuista saattaa olla pikemmin raja-pintahaasteita kuin varsinaisia esineiden internetiin liittyvä haasteita. Tämän kaltaiseen testaamiseen on kuitenkin pitkä matka kuljettavana; monia teknisiä ja eettisiä, mm. tietoturvaan ja kerätyn datan hyödyntämiseen tai yhdistelyyn liittyviä ongelmia on voitettavana.

Kolmanneksi pilvipalvelut mahdollistavat työn ja palveluiden yhdistymisen asiakkaaseen ajasta ja paikasta riippumatta. Tämän suhteen harppaus tapahtui COVID-19-pandemian seurauksena, mutta siitä ei tulla luopumaan, sillä rinnakkaisina trendeinä ovat Heikkilän (2021) mukaan ”*stay at home*” ja ”*work at home*”. Testaamisen yhteydessä on ilmeistä, että pilvipalveluja tullaan tarvitsemaan enenevässä määrin, jos oppimisesta ja testaamisesta todella tulee paikasta riippumatonta.

Neljänneksi ns. *Big data*n hyödyntäminen lisääntyy. Heikkilän (2021) mukaan viimeisen parin vuoden aikana on ehkä syntynyt 90 % kaikesta ihmiskunnan luomasta digitaalisesta datasta. ”Datan määrä kasvaa räjähdysmäisesti, koska ihmiset viettävät enemmän vapaa-aikaa ja tekevät töitä verkossa. Sosiaalinen media, pelaaminen, suoratoistopalvelut, videot ja kuvat, kaikki kerryttävät dataa ja sen määrä jatkaa kasvamistaan. Datan sanotaankin olevan uusi öljy ja ne yritykset, joilla on tuorein ja laajin data ja parhaat analysointityökalut, ovat vahvimpia yrityksiä.” (Heikkilä, 2021, s. 1.)

*Big data*n hyödyntämisessä on kaksi osittain vastakkaista näkökulmaa. Yhtäältä asiaan liittyy useita ratkaisemattomia eettisiä haasteita ja riskejä: haluammeko, että meistä on olemassa katkeamaton sanojen, tekojen ja suoritusten historia, jonka joku voi varastaa tai sabotoida toisenlaiseksi tulevaisuutemme kustannuksella?[21] Tämänkaltaisia haasteita ovat käsitelleet yleisellä tasolla mm. Kitchin (2013) sekä Price ja Cohen (2019). Toisaalta on haasteellista käyttää *Big data*n järkevästi. Jokaisesta oppilaasta koostuu jo nyt paljon tietoa koulunkäynnin aikana eri vuosilta. Kansallisella tasolla tämä oppilaan oppimisen polun seuraamisessa syntyneen tiedon hyötykäyttö on puuttunut kokonaan tai se on hyvin alkutekijöissään koulun ja koulutuksen järjestäjän tasolla pedagogisissa ratkaisuissa. Järkevissä tulevaisuuden digitaalisissa testausjärjestelmissä, joiden halutaan tuovan lisäarvoa oppilaille, opettajille ja koulutuksen järjestäjille, oppilasta koskevaa oppimisdiagnostista tietoa on järkevää koota mielekkäästi. Tässä eräänä edellä kävijänä Suomessa on ollut Turun yliopiston Oppimisanalytiikan keskuksen ViLLE-järjestelmä, joka kokoaa oppilaista kehitystietoa ja tarjoaa tätä opettajan ja oppilaan käyttöön noin kolmasosassa Suomen kouluja. Vastaavanlaisia digitaalisen ja perinteisen opetuksen rajapinnassa tapahtuvan sopeuttavan opetuksen ja oppimisen

21 Tästä huomiosta erityinen kiitos tutkija Anu Lehikolle Lapin yliopistosta, joka on erityisesti perehtynyt virtuaaliseen oppimiseen, immersiiviseen teknologiaan ja uusien teknologioiden tuomiin haasteisiin. Lehikko luki tekstin ulkopuolisena arvioijana ja antoi useita hyödyllisiä korjausehdotuksia ja tarjosi tarkentavaa lähteistöä.

(*blended learning*) ja siihen linkittyvän oppimisdiagnostiikan alustojen ekosysteemi on kuitenkin vasta kehittymässä. Tähän tarkoitukseen syntyneitä kansainvälisiä ohjelmistoja ja alustoja (*Blended learning maganement systems*) on esitellyt mm. Papas (2021; ks. myös esimerkiksi https://qridi.com/fi/edu). Vastaavanlaisia digitaalista, pitkän aikavälin diagnostiikkatietoa olisi mahdollista koota myös eri oppikirjakustantajien ja opettajajärjestöjen digitaalisiin testeihin liittyen.

3.6.6 Ekologisen ajattelun muutoksesta nousevat muutostarpeet

Yleisesti ottaen kestävään kehitykseen liittyvä ajattelu jaetaan neljään alueeseen: ekologiseen, taloudelliseen, sosiaaliseen ja kulttuuriseen kestävään kehitykseen (ks. esimerkiksi. Räkköläinen ym. 2017). Näistä ekologinen kestävä kehitys oli ensimmäinen, joka selvästi kirjattiin YK:n Brundtlandin komission raporttiin (YK, 1987). Jo varhaisessa vaiheessa kestävä kehitys jaettiin kuitenkin kolmeen pääryhmään: ekologiseen, taloudelliseen ja sosiaaliseen (mm. Heikkurinen, 2014; Salonen, 2010; Soini, 2013). Vuodesta 2010 lähtien kulttuurinen kestävä kehitys on ollut yksi neljästä keskeisestä kestävän kehityksen osatekijästä erityisesti YK:n ja UNESCO:n toiminnassa (mm. YK, 2010). Katajamäki (2011) tiivistää kolmen viimeksi mainitun osa-alueen keskeiset piirteet seuraavasti: taloudellisesti kestävä kehitys ottaa luonnon tarpeet huomioon, sosiaalisesti kestävä kehitys takaa ihmisten oikeudenmukaisen kohtelun riippumatta heidän syntyperästään, varallisuudestaan tai asuinpaikastaan, ja kulttuurisesti kestävä kehitys kunnioittaa ihmisten ja eläimien perusoikeuksia.

Kestävään kehitykseen liittyvä ajattelu on muuttunut viimeisten vuosien aikana radikaalisti. Yhtäältä ympäristöhuoli on muuttunut osittain ehkä jopa ympäristöahdistukseksi ilmaston lämpenemisen ja siihen liittyvän keskustelun myötä. Toisaalta olemme entistä tietoisempia esimerkiksi omista etnisistä juuristamme, kulttuurin omimisesta ja kätketyistä epätasa-arvoa ylläpitävistä rakenteista kulttuureissamme.

Mitä tulee kestävän kehityksen ekologiseen ja taloudelliseen ulottuvuuteen, oppilaat ja opettajat ovat tulleet yhä tietoisemmiksi ympäristökatastrofien mahdollisuudesta ja uhasta ja luonnonkulutuksen haitallisista vaikutuksista maailman taloudelliseen tasapainoon. On kuitenkaan vaikea kuvitella, että tällä olisi juuri ratkaisevaa merkitystä digitaalisen testauksen tulevaisuuden suhteen, kun jo on siirrytty paperi-kynä-maailmasta pitkälti digitaaliseen maailmaan. Ehkä päinvastoin: ekologisen ajattelun muutos vie samaan suuntaan kuin muutkin muutosvoimat: digitaalinen testaus on ekologisesti järkevää ja osa sellaisia koulutuksen ratkaisuja, joilla säästetään luontoa, mikäli sähköntuottamiseen rakennetut ratkaisut ovat kestäviä. Tietenkin kaikessa kehittämistyössä ja valinnoissa on järkevää ottaa huomioon taloudellinen ja ekologinen kestävyys esimerkiksi turhien kokousmatkojen välttämisenä ja etätyöskentelyn kehittämisessä ohjelmistoja suunniteltaessa ja toteutettaessa. Tähän voi tulevaisuudessa liittyä myös eettisiä näkökulmia, mikäli koodaustyötä annetaan sellaisiin kehittyviin maihin, joissa ekologiset seikat ja siihen liittyvät periaatteet eivät välttämättä ole samalla ymmärretty tai sovellettu kuin Suomessa. Olisiko kuitenkin parempi suosia koodaamisen tilaamisessa maita, joissa ekologiset seikat otetaan vakavasti, vaikka se olisikin hieman kalliimpaa? Tämänkaltaisista tekijöistä saattaa tulevaisuudessa tulla kilpailutekijöitä, kun testausratkaisuja kaupallistetaan.

Mitä tulee sosiaaliseen ja kulttuuriseen kestävyyteen, tasapuolisuuteen ja oikeudenmukaisuuteen, on tietenkin selvää, että tulevaisuuden testijärjestelmien tulee olla sellaisia, jotka mahdollistavat sen, että testit ovat tasavertaisesti saavutettavia ja samalla tavoin ymmärrettävissä riippumatta testattavan kulttuurisesta taustasta. Tulevaisuudessa on oltava entistäkin tietoisempia sen suhteen, että tehtävien laadinnassa vältetään kulttuurisia stereotypioita niin kielellisesti kuin kulttuurisestikin; on oltava nykyistäkin tarkempana esimerkiksi sen suhteen, millaisia tekstejä lukutaitoa mittaaviin testeihin valitaan. Suomessa sosiaaliseen ja kulttuuriseen kestävyyteen liittyvät seikat voivat koskea sellaisia ryhmiä kuin esimerkiksi saamelais- ja romani- ja maahanmuuttotaustaisia yhteisöjä, kuten edellä todettiin. Saavutettavuus koskee myös erilaisia erityisryhmiä kuten näkö-, kuulo- ja liikuntarajoitteisia testiin osallistujia (ks. keskustelu edellä ja kirjallisuus Eglash ym., 2020; Litts ym., 2020; Shaheen, 2021).

3.6.7 Pidemmän aikavälin signaaleja ja niiden vaikutuksia digitaaliseen testaamiseen

Futures Platform -ajatuspalan perustaja, tulevaisuustutkija Tuomo Kuosa antoi tätä artikkelia varten vinkkejä sellaisista megatrendeistä, jotka saattavat viitoittaa pidemmällä aikavälillä koulutuksen tulevaisuutta. Näitä tarkastellaan tässä luvussa lyhyesti digitaalisen testaamisen näkökannalta. Seurattavia trendejä ovat mm. metaversumiin (*metaverse*) liittyvät kehityspolut (mm. Sandal, 2021; Bryant, 2021), tekoälyn ja lisätyn todellisuuden (*augmented reality*) tuomat mahdollisuudet koulutuksessa (mm. Veltheim, 2021; Holstein, 2017) sekä yhdistelevään tai sulautettuun oppimiseen (*blended learning*) liittyvät trendit COVID-19-pandemian jälkeisessä maailmassa (mm. Govidrarajan & Srivastava, 2020; Ray, 2022).

Metaversumi tulevaisuuden trendinä

Metaversumiin liittyvät kehityspolut linkittyvät internetin seuraavaan kehitysvaiheeseen, ”Web 3.0” (Kerner & Gillis, 2022; Sanders, 2021). Kun ”Web 1.0” oli ensisijaisesti staattisten verkkosivustojen kokoelma informaation jakelua varten ja ”Web 2.0” toi mukaan interaktiivisia elementtejä, käyttäjien tuottamaa materiaalia ja osallistuvuuden elementin, ”Web 3.0” tuo mukaan myös tekoälystä hyötyvää semanttista sisältöä, spatiaalisia elementtejä, virtuaalisten maailmojen kehittämisen ja immersiivisiä eli uppouttavia kokemuksia, jotka sekoittuvat fyysisiin realiteetteihin (Kerner & Gillis, 2022). Nykyisistä teknologioista esimerkiksi Teams- ja Zoom-kokoukset ovat osa tätä metaversumikehitystä. Näistä on jo olemassa virtuaali-realistisia sovelluksia, esimerkiksi Facebookin *Horizon Workrooms*, jossa ihmiset kokoontuvat virtuaalisiin kokouksiin kuin he osallistuisivat fyysisesti tilanteeseen. Vaikka termi ”Web 3.0” nykyisessä merkityksessään kehitettiin jo vuonna 2014 (ks. Kerner & Gillis, 2022), suurelle yleisölle metaversumin käsite lienee tutuin Facebookin omistaja Zuckerbergin ennustuksesta, että Facebookista on tulossa ”metaversumiyhtiö” viidessä vuodessa (tai tätä kirjoitettaessa neljässä; ks. Bryant, 2021); tähän viittaa myös yhtiön vuonna 2021 lanseeraama uusi nimi ”Meta”. Onkin todennäköistä, että suuret sosiaalisen median jättiläiset tulevat dominoimaan myös ”Web 3”:n toimintaa. Mutta on odotettavaa, että sen sisään tulee rakentumaan laaja ekosysteemi myös pienemmille toimijoille ja sisällöntuottajille (Sanders, 2021; ks. myös Bryant, 2021).

Digitaalisen testaamisen näkökannalta on mahdollista, että jos koulutuksessa kyetään mielekkäästi hyödyntämään syntyviä uusia ”maailmoja”, metaversumeita, tavoitteellisen opetuksen tukena, lienee järkevää ajatella, että näihin kehitetään elementtejä, joilla voidaan testata, onko tämän tavoitteen osaaminen saavutettu. Jos nykyisiä Teams tai Zoom-oppimistilanteita pidetään osana metaversumien hyödyntämistä, digitaalista testaamista jo tehdäänkin osana monimuotoista opettamis-, oppimis- ja testausprosessia. Osana metaversumien ajatusta ehkä juuri onkin päästä pakenemaan realismia, jolloin kouluopetus tai tavoitteelliset suoritukset eivät välttämättä istu metaversumien kaikkiin käyttötarkoituksiin. Myös tämän tyyppisten ei-tavoitteellisten metaversumien kehitystä on järkevä seurata, sillä rakennetut metaversumiratkaisut saattavat hyvinkin toimia eri oppiaineiden sisällön oppimisen kontekstina, mikäli näissä tulee eteen esimerkiksi sosiaalisiin suhteisiin, (rahan) vaihdantaan tai kommunikointiin liittyviä tilanteita, jotka tulee mielekkäästi ratkaista, ja joissa syntyneitä ratkaisuja voisi mielekkäästi kvantifioida oppilaan arvioinnin tueksi.

Tekoäly ja lisätty todellisuus tulevaisuuden trendeinä

Toinen tulevaisuuden seurattava trendi liittyy opettajuuden muutokseen tekoälyn ja lisätyn todellisuuden kehittymisen myötä. Tähän liittyvää keskustelua on koonnut esimerkiksi Veltheim (2021). Tulevaisuustutkijat—kuten esimerkiksi Marianna Mäkiteeri Futures Platformista ja Thomas Frey DaVinci -insituutista Veltheimin (2021) siteeraamina—ennustavat, että tulevaisuudessa tullaan näkemään entistä yksilöllistetympiä koulutusratkaisuja (”*hyper-individualized education*”, kuten Fray asian ilmaisee). Tämä saattaa johtaa siihen, että tulevaisuudessa tekoäly, joka tuntee oppijan paremmin kuin kukaan, tietää, mitä osaamista puuttuu, ja kykenee parhaiten ohjaamaan oppijan tarvittavan tiedon äärelle. Veltheim (2021) esittää, että pidemmällä tulevaisuudessa tekoälystä tulee opettaja, ja ihmisen rooli on olla mentori ja valmentaja. Tästä näkökulmasta katsoen perinteinen testaaminen, yhden oikean ratkaisun ja asioiden muistamisen ajan on sanottu tulleen tiensä päähän (näin mm. Krishnan, 2020 ja Mäkiteeri Veltheimin, 2021 siteeraamana).

Äärimmäisen yksilöllistettyihin opetusratkaisuihin on kuitenkin pitkä matka, jos se koskaan osoittautuukaan järkeväksi ajatukseksi etenkään oppivelvollisuusikäisillä oppijoilla, kun otetaan huomioon, että kaikilla kansalaisilla ei välttämättä ole (eikä ole järkevääkään olla) käytössä kaiken tietävää tekoälyä, joka auttaisi kansalaista esimerkiksi matematiikan osalta arkisessa taloudenpidossa, hankintojen budjetoinnissa tai pankkineuvotteluissa. On uskottavaa, että oppivelvollisuusikäisten oppijoiden on vielä vähintään kymmenien kohorttien päähän opittava samoja asioita, muistettava asioita ulkoa ja osattava käyttää tätä tietoa arkipäivän tilanteissa—ja myös osallistuttava tämä opittavan tiedon testaamiseen. Huomattakoon, että Khrishnan (2020), joka muutoin kritisoi nykyistä koulusysteemiä aikansa eläneeksi, nostaa suomalaisen koulujärjestelmän esimerkiksi hyvästä systeemistä.

Digitaalisen testauksen tulevaisuuden näkökannalta mahdollinen tekoälyyn perustuva ”hyperyksilöllinen koulutus”, jos koskaan toteutuu, ei sinällään tuo ylimääräistä haastetta, mikäli tulevaisuudessakin yhteisesti sovitut tavoitteet kirjataan kaikille oppijoille lähtökohtaisesti samoiksi. Tällöinkin jonkun on kyettävä päättelemään jollain järkevällä tavalla objektiivisesti se, hallitseeko oppija tavoitteeksi asetetun asian vai ei. Osaamisen arviointi jossain muodossa, esimerkiksi testaamalla, on tähän hyvä vaihtoehto. Toki myös tekoälyyn ja algoritmeihin voidaan kirjoittaa osaamisen kehittymisen tarkkailuun liittyviä elementtejä. Näitä joudutaan kuitenkin varmaan odottamaan vieläkin pidempään kuin itse tekoäly-opettajan kehittymistä.

Opettajan työn näkökannalta tekoälyä konkreettisempi ja todennäköisemmin nopeammin käytännön sovelluksia tuottava työväline saattaa olla lisättyyn todellisuuteen (*augmented reality*) liittyvät opetusratkaisut. Hiljattain raportoitiin (Molina ym., 2019; ks myös Johns Hopkins, 2021), että ensimmäinen lisätyn todellisuuden avulla tehty aivoleikkaus suoritettiin onnistuneesti Johns Hopkins sairaalassa Yhdysvalloissa. Tämänkaltaisilla uutisilla on taipumusta olla tiettyä teknologista edistysaskelta vahvistava vaikutus; ”Jos aivospesialistitkin voivat käyttää lisättyä todellisuutta leikkauksissa, miksen minä opettajana voisi sitä käyttää opetuksen tukena”, saattaisi ajattelu mennä lähitulevaisuudessa. Lisättyä todellisuutta on jo ollut useita vuosia mahdollisuus käyttää opiskelussa esimerkiksi historiallisissa kohteissa ja museoissa, mutta varsinaiset opettajille tarkoitetut sovellukset ovat olleet pienimuotoisia (ks. Holstein, 2017). Veltheimin (2021) mukaan Thomas Frey ennustaa, että seuraavan sukupolven käyttöliittymät sisältävät kehittyneempää teknologiaa kuten älylaseja tai henkilökohtaisia projektoreita, joiden avulla tekoäly voi näyttää hakutuloksia oppijalle suoraan (Veltheim, 2021). Vastaavasti lisätyn todellisuuden opettajasovelluksissa opettaja voi esimerkiksi nähdä omilla laseillaan oppijan virtuaalisia viittauksia tai avun pyyntöjä, käyttäytymistä tehtävien aikana tai osaamisen kehittymistä (ks. esimerkiksi Holstein, 2017).

Digitaalisen testauksen näkökannalta lisätyn todellisuuden tuomat mahdollisuudet voivat liittyä esimerkiksi opeteltavaan aineistoon liittyvien oppimistehtävien ja tähän linkittyvän oppimisanalytiikan hyödyntämiseen simultaanisesti. Käytännössä teknologiaa voidaan kehittää suuntaan, jossa oppijan käytyä läpi digitaalista materiaalia, hän vastaa asiaan liittyviin (testi-)kysymyksiin ja algoritmi arvioi ratkaisun oikeellisuuden ja monivalintatehtävissä väärään vaihtoehtoon johtaneen virheellisen ajattelun. Tämä tieto syötetään takaisin opettajalle, joka saa esimerkiksi virtuaalisesti viestin siitä, kenellä oppilaista on minkäkinlaista ongelmaa esimerkiksi matematiikan tehtäviä ratkaistaessa. Tiedon perusteella opettaja voi täsmä-auttaa oppijaa—ellei tekoäly tee sitä välittömästi. Tämän kaltaisten sovellusten toteutuminen edellyttää kuitenkin esimerkiksi uudenlaista rajapintateknologiaa, ehkä esineiden internetin kehittymistä, pitkälle kehitettyjä oppimisanalytiikan ratkaisuja ja jatkuvaa liikennettä pilvipalveluiden avulla keskustietokoneelle ja takaisin. Tämänkaltainen diagnostiikka olisi mahdollista jo tämän päivän teknologialla, mutta käytännön sovellukset edellyttävät suurta muutosta opetusteknologian käytössä ja hyväksymisessä opetuksen tueksi. On vaikea kuvitella, että tämän suuntaisia ratkaisuja olisi aivan lähivuosina kuitenkaan yleisessä käytössä.

Sulautuva ja yhdistelevä oppiminen ja opettaminen tulevaisuuden trendeinä

Lähempien vuosien opetusratkaisuja ovat erilaiset sulautuvaan tai yhdistelevään oppimiseen (*blended learning*) liittyvät ratkaisut. Sulautuva oppiminen viittaa muodollisen oppimisen ja opettamisen tapaan, jossa oppijat oppivat sekä perinteisen henkilökohtaisen opetuksen että verkkomedian kautta. Termi on ollut tunnettu jo pitkään etäopetuksen yhteydessä (ks. esimerkiksi Osguthorpe & Grahan, 2003), mutta erityisesti COVID-19-pandemia nosti termin uudelleen keskusteluun (ks. esimerkiksi Govindarajan & Srivastava, 2020; Ray, 2022). Ray on koonnut joukon tutkimuksia ja julkaisuja, joissa esitellään sulautuvan opetuksen ja oppimisen uusia muotoja ja niitä vahvistavia prosesseja. Govindarajan ja Srivastava puolestaan pohtivat asiaa korkeakouluopetuksen näkökannalta ja esittävät oleellisen kysymyksen: mitä erityistä etua oppijat saavat siitä, että he asuisivat 4 vuotta korkeakoulun asuntolassa. Vastaavan voisi kysyä Suomen olosuhteissa toisin: mitä erityistä hyötyä oppijat saavat siitä, että he osallistuisivat kaikille oppitunneille tai luennoille fyysisesti—tai edes virtuaalisesti? COVID-19-pandemia opetti, että etätyöskentely ei sovi kaikille, mutta osalle se sopi oikein hyvin (ks. esimerkiksi Goman ym., 2021). Työskentely, jossa sopivasti yhdistellään verkkokursseja ja läsnäoloa tulee todennäköisesti olemaan ainakin osa koulutuksen ratkaisuja tulevaisuudessa.

Digitaalisen testaamisen näkökannalta sulautuva opetus ja oppiminen eivät tuo varsinaisia uusia haasteita, joita ei edellä olisi jo käsitelty. Jos oppiminen on tavoitteellista, siihen on järkevää yhdistää osaamisen kehittymisen tai loppuosaamisen kontrollointimekanismi, joka perinteisesti on osaamisen testaamista jossain muodossa. Tällöin osaamista voidaan testata joko yksittäisten verkkokurssien yhteydessä tai lähiopetuksen aikana.

3.6.8 Yhteenvetoa digitaalisen testauksen tulevaisuuden näkymistä

Vertailuksi sille, millaisia muutoksia digitaalisiin testeihin on odotettavissa vuoteen 2035 mennessä, voidaan muistella, millaisia digitaaliset testit olivat 15 vuotta sitten, jolloin ensimmäisiä laajan mittakaavan digitaalisia testejä toteutettiin PISA-testien yhteydessä (PISA, 2010) ja digitaaliset testit otettiin käyttöön kielten testauksessa (TOEFL, 2007). Yhtäältä monet asiat ovat pysyneet samanlaisina kuin varhaisissa digitaalisissa testeissä. Yhä edelleen käytetään esimerkiksi hyvin samankaltaisia tehtävätyyppejä kuin on käytetty jo 100 vuoden ajan erityisesti monivalinta-, yhdistämis- ja lyhytvastaustehtävissä; näitä tehtäväformaatteja tullaan näkemään tulevaisuuden testeissäkin.

Toisaalta joidenkin seikkojen suhteen vanhemmat digitaaliset testit jo näyttävätkin vanhentuneilta. Ensimmäiset hahmotelmat esimerkiksi silloin edistyksellisistä pelillisistä tehtävistä näyttäytyvät nykykatsannosta käsin karkeina ja pelkistettyinä grafiikaltaan, fonteiltaan ja värimaailmaltaan. Tietenkin oleellista on myös se, että viimeisen viidentoista vuoden aikana väylänopeudet ovat monikertaistuneet, mikä on mahdollistanut hyvinkin vaativat graafiset ratkaisut testijärjestelmien toteutuksessa. Nyt ollaan siirtymässä 5G-maailmaan, mikä mahdollistaa entistä nopeammat yhteydet, ja tällä tulee olemaan tietenkin ilmeinen merkitys esimerkiksi siihen, kuinka vaativaa grafiikkaa tai esimerkiksi millaista videomateriaalia testeissä voidaan käyttää. Tältä osin teknologiset mahdollisuudet ovat muuttunet oleellisesti ja nopeasti viimeisten vuosien aikana, ja odotettavasti kehitys jatkuu samansuuntaisena.

Oleellinen muutos tulevaisuuteen nähden tulee olemaan tekoälyn ja älykkäiden algoritmien tuomien mahdollisuuksien suurempi rooli testauksessa, diagnosoinnissa ja älykkäiden räätälöityjen ratkaisujen tuottamisessa oppimisen edistämisessä—sekä yhtälailla älykkään testivilpin. Näiden kaltaisia mahdollisuuksia oli käytettävissä erittäin rajallisesti vielä 15 vuotta sitten. Vuoteen 2035 mennessä älykkäisiin ratkaisuihin saattaa yhdistyä vähintään kokeiluja älykkäillä immersiivisillä ja interaktiivisilla osio- ja testityypeillä sekä muunlaista oppijaa koskevaa datan keräystä. Tekoäly saattaa muuttaa myös tehtävien laadinnan prosesseja oleellisesti.

Eräs uskottava tulevaisuuskuva on, että digitaalisen testaamisen volyymit kasvavat merkittävästi, kun kouluverkko kapenee, mutta koko Suomi halutaan pitää asuttuna. Tämä lisää niin etäopetuksen kuin etätestaamisen määriä. Tällöin myös etätestaamiseen liittyvien kontrollimekanismien merkitys kasvaa tulevaisuuden testaamisessa; testijärjestelmiin tultaneen lisäämään ratkaisuja, joilla voidaan sekä kontrolloida testattavan autenttisuutta että pienentää testivilpin mahdollisuuksia niin pieneen kuin se on järkevää. Emme voi koskaan päästä täysin eroon testivilpistä—täysin pitävää järjestelmää ei realistisin kustannuksin ole mahdollista rakentaa.

On helppo ennustaa, että tulevaisuuden digitaalinen testaaminen tulee sisältämään aiempaa enemmän diagnostisia työkaluja osaamisen kehittämisen tueksi. Tulevaisuudessa ei enää riitä, että opettajalle ja oppilaalle tuotetaan lopputulos ja pistemäärä (edes osa-alueittain), vaan tietoa on tuotettava myös siitä, millä alueilla testattavan on järkevää lisätä harjoitusta, jotta asetetut tavoitteet saavutetaan. On mahdollista, että vuoteen 2035 mennessä käytössämme on FUNA-testin tapaisia, vakuuttavia diagnostiikkatyökaluja esimerkiksi kielellistä tai matemaattisista oppimisvaikeuksista, joista saatua tietoa saatetaan kyetä yhdistämään testissä saatuun tietoon.

Vuoteen 2035 mennessä testausjärjestelmiin on luultavasti syntynyt oppilaista koottavaan *Big data*an liittyviä sovelluksia opetuksen ja oppimisen kehittämisen tueksi, mikäli asiaan liittyvät eettiset haasteet on saatu ratkaistua. Mikäli tekniset järjestelmät saadaan tukemaan kehitystä, on mahdollista, että kun oppilaista kootaan joka tapauksessa tietoa koko koulu-uran ajalta, näiden kehitystietojen perusteella voidaan rakentaa hyvinkin räätälöityjä opintopolkuja ja pedagogisia ratkaisuja riippuen siitä, minkälaisia haasteita oppilailla on. Tämän tyyppisiä ratkaisuja jo onkin käytössä esimerkiksi Turun yliopiston Oppimisanalytiikan tutkimuskeskuksen ViLLE-järjestelmässä, mutta vuoteen 2035 mennessä kehitystä voidaan odottaa tapahtuvaksi myös esimerkiksi oppikirjakustantajien digitaalisten testien osalta sekä äidinkielen, kielten ja matematiikan opettajien liittojen kansallisissa testeissä (ÄOL, SUKOL ja MAOL), kun ne saadaan digitaaliseen muotoon.

Vaikka tulevaisuuden näkymien suunta lienee melko selvä, muutoksen hitaus on hämmästyttänyt tulevaisuustutkijoita toistuvasti. Ennusteilla on Mannermaan (1991) mukaan taipumus olla pitkällä aikavälillä pessimistisiä ja lyhyellä aikavälillä optimistisia. Asiantuntijoilla on myös taipumus sijoittaa hyödyllinen, järkevä ja taloudellisesti kannattavampi idea läheisempään tulevaisuuteen kuin vähemmän hyödyllinen, vähemmän järkevä ja taloudellisesti kannattamattomampi idea (ks. Metsämuuronen, 2001a, ks. myös Kuusi, 1999). On hyvinkin mahdollista, että vuoden 2035 kansallisissa testeissä yhä edelleen kamppaillaan aivan samanlaisten teknisten ongelmien kanssa kuin nykyisissäkin testeissä.

Toisaalta on mahdollista, että aivan yllättävä teknologinen innovaatio tai vasta laboratoriovaiheessa oleva kehitelmä mullistaa testaamisen täysin. Eräs tällainen voisi olla esimerkiksi infravalokuvaus, jossa ihossa tapahtuvat muutokset voidaan taltioida lämpö- tai kemikaalimuutoksina. Voisiko olla mahdollista, että esimerkiksi oikean vastauksen tai vastausalgoritmin tietäminen ja ajatteleminen pelkästään tehtävän näkemisen perusteella erittäisi oikean kemikaalikomposition ihoon, jonka kuvantamislaite pystyisi rekisteröimään ilman mitään testiäkin? Tai voisivatko Nobel-palkinnon saaneen Moserin tutkimusryhmän tulokset korkeamman ajattelun paikantamisesta aivoissa (Nobel, 2014; Rowlands ym., 2016) johtaa käytännön sovelluksiin, joissa kuvannetaan aivojen tiettyjen osien solutilavuus, joka osoittaisi osaamisen määrän suoraan tai laskennallisesti mallittaen (ks. keskustelu myös esimerkiksi Metsämuuronen & Räsänen, 2018)? Tai kyettäisiinkö aivokuvantamista kuten magnetoenkefalografiaa (MEG) tai *event-related potential* (ERP) -aivovasteita hyödyntämään laajemmin osaamisen ja muistin mittaamisessa (ks. esimerkiksi Näätänen ym., 1997). Yllättäviä ratkaisuja ja ongelmia saattavat tarjota myös esimerkiksi avoimeen internetiin ja ChatGPT-tyyppisiin tekoälysovelluksiin liittyvät sovellukset. Mahdollisuudet ovat monet, ja vasta aika näyttää, löytyykö nykymuotoisten testien korvaajaa—tai halutaanko sellaisia edes löytää.

3.7 Lähteet

Blikstad-Balas, M., Roe, A., Pedersen Dalland, C., & Klette, K. (2022). Homeschooling in Norway during the pandemic. Digital learning with unequal access to qualified help at home and unequal learning opportunities provided by the school. Teoksessa F. M. Reimers (toim.), *Primary and secondary education during Covid-19. Disruptions to educational opportunity during a pandemic* (ss. 177–212). Springer. https://link.springer.com/content/pdf/10.1007%2F978-3-030-81500-4.pdf

Brown, G.T.L. (2019). Technologies and Infrastructure: Costs and Obstacles in Developing Large-Scale Computer-Based Testing. *Education Inquiry, 10*(1), 4–20. https://doi.org/10.1080/20004508.2018.1529528

Bryant, M. (2021). Is Facebook leading us on a journey to the metaverse? *The Guardian*. 26.9.2021. https://www.theguardian.com/technology/2021/sep/26/is-facebook-leading-us-on-a-journey-to-the-metaverse

Butler, A.C., Karpicke, J.D. & Roediger, H.L. 3rd (2007). The effect of type and timing of feedback on learning from multiple-choice tests. *Journal of Experimental Psychology, Applied, 13*(4), 273–281. http://www.ncbi.nlm.nih.gov/pubmed/18194050

Butler, A.C., Karpicke, J.D. & Roediger, H.L. 3rd (2008). Correcting a metacognitive error: feedback increases retention of low-confidence correct responses. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(4), 918–928. https://doi.org/10.1037/0278-7393.34.4.918

Butler, A.C. & Roediger, H.L. 3rd (2008). Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. *Memory & Cognition, 36*, 604–616. http://www.ncbi.nlm.nih.gov/pubmed/18491500

Clariana, R. & Wallace, P. (2002). Paper–based versus computer–based assessment: key factors associated with the test mode effect. *British Journal of Educational Technology, 33*(5), 593–602. https://doi.org/10.1111/1467-8535.00294

Cummings, J. J., & Bailenson, J. N. (2016). How immersive is enough? A meta-analysis of the effect of immersive technology on user presence. *Media Psychology, 19*(2), 272–309. https://doi.org/10.1080/15213269.2015.1015740

Dikli, S. (2006). An Overview of Automated Scoring of Essays. *Journal of Technology, Learning, and Assessment, 5*(1). https://ejournals.bc.edu/index.php/jtla/article/view/1640

Eglash, R., Bennet, A., Babbitt, W., Lachney, M., Reinhardt, M. & Hammond-Sowah, D. (2020). Decolonizing posthumanism: Indigenous material agency in generative STEM. *British Journal of Educational Technology, 51*(4), 1334–1353. https://doi.org/10.1111/bjet.12963

Feurstein, R., Rand, Y., Hoffman, M.B. (1979). *The dynamic assessment of retarded performers: The learning potential assessment device: Theory, instruments, and techniques*. University Park Press.

Fishbein, B., Martin, M.O., Mullis, I.V.S, & Foy, P. (2018). The TIMSS 2019 Item Equivalence Study: Examining Mode Effects for Computer-Based Assessment and Implications for Measuring Trends. *Large-scale Assessments in Education, 6*, Article 11. https://doi.org/10.1186/s40536-018-0064-z

FUNA (2019). *FUNA—Toiminnallisen laskutaidon arviointi*. Oppimisanalytiikan keskus, Turun yliopisto. http://oppimisanalytiikka.fi/funa

Goman, J., Huusko, M., Isoaho, K., Lehikko, A., Metsämuuronen, J., Rumpu N., Seppälä, H., Venäläinen, S., & Åkerlund, C. (2021). *Poikkeuksellisten opetusjärjestelyjen vaikutukset tasa-arvon ja yhden vertaisuuden toteutumiseen eri koulutusaloilla. Osa III: Kansallisen arvioinnin yhteenveto ja suositukset*. Julkaisut 8:2021. Kansallinen koulutuksen arviointikeskus. https://karvi.fi/wp-content/uploads/2021/04/KARVI_0821.pdf

Govindrajan, V. & Srivastava, A. (2020). What the shift to virtual learning could mean for the future of higher Ed. *Harvard Business Review*, March 31, 2020. https://hbr.org/2020/03/what-the-shift-to-virtual-learning-could-mean-for-the-future-of-higher-ed

Greving, S. & Richter, T. (2018). Examining the Testing Effect in University Teaching: Retrievability and Question Format Matter. *Frontiers in Psychology,* 9:2412. https://doi.org/10.3389/fpsyg.2018.02412

Grigorenko E.L. & Sternberg R.J. (1998). Dynamic testing. *Psychological Bulletin, 124*(1), 75–111. https://doi.org/10.1037/0033-2909.124.1.75

Hamhuis, E., Glas, C., & Meelissen, M. (2020). Tablet assessment in primary education: Are there performance differences between TIMSS' paper-and-pencil test and tablet test among Dutch grade-four students? *British Journal of Educational Technology, 51*(6), 2340–2358. https://doi.org/10.1111/bjet.12914

Hannu, J., Saikkonen, T., Häkkinen, J. Karttunen, J., & Moilanen, M. (2010). Enabling remote testing: Embedded test controller and mixed-signal test architecture. *Journal of Electronic Testing, 26*, 641–658. https://doi.org/10.1007/s10836-010-5175-6

Hansen, A. (2011). *What does dynamic testing entail?* Translated from Norwegian for the Daffodil project, 2011. Lifelong Learning Programme. EU. https://static1.squarespace.com/static/5c2c66dd297114a6e7b41163/t/5c4b79fc562fa7de1aa49e18/1548450304544/Article+on+Dynamic+Assessment.pdf

Harjunen, E. & Rautopuro J. (2015). *Kielenkäytön ajattelua ja ajattelun kielentämistä. Äidinkielen ja kirjallisuuden oppimistulokset perusopetuksen päättövaiheessa 2014: keskiössä kielentuntemus ja kirjoittaminen*. Julkaisut 2015:8. Kansallinen koulutuksen arviointikeskus. https://karvi.fi/wp-content/uploads/2015/04/KARVI_08151.pdf

Haywood, H.C. (1992). Interactive assessment: A special issue. *Journal of Special Education, 26*, 233–234. https://doi.org/10.1177/002246699202600301

Haywood, H.C. (1997). Interactive assessment. Teoksessa R. L. Tayler (toim.), *Assessment of individuals with mental retardation*. Singular Publishing Group.

Heikkilä, T. (2021). Vuoden 2021 tärkeimmät megatrendit. *Sijoittaja.fi*. https://www.sijoittaja.fi/258979/vuoden-2021-tarkeimmat-megatrendit/

Heikkurinen, P. (2014). Kestävyyden käsitteen ulottuvuudet. *Tieteessä tapahtuu, 32*(4), 10–16. http://ojs.tsv.fi/index.php/tt/article/view/46149

Hellgren, J. & Marjanen, J. (2020). *Svenska och litteratur i slutet av årkurs 9. Resultat av en utvärdering av lärresultat våren 2019*. Publikationer 18:2020. Nationella Centret för Utbildningsutvärdering. https://karvi.fi/wp-content/uploads/2020/10/KARVI_1820.pdf

Holstein, K. (2017). *The Lumilo Project*. Haettu osoitteesta https://medium.com/lumilo/the-lumilo-project-cda82726f9ec (24.2.2023).

Huisman T (2006). *Luen, kirjoitan ja ratkaisen. Peruskoulun kolmasluokkalaisten oppimistulokset äidinkielessä ja kirjallisuudessa sekä matematiikassa*. Oppimistulosten arviointi 7/2006. Opetushallitus. https://karvi.fi/app/uploads/2014/09/OPH_0906.pdf

Huisman T & Silverström C (2006). *Läsa, skriva, räkna. Utvärdering av inlärningsresultat i modersmål och litteratur samt matematik i årskurs 3*. Utvärdering av inlärningsresultat 8/2006. Utbildningsstyrelsen. https://karvi.fi/app/uploads/2014/10/OPH_1006.pdf

Härmälä, M., Huhtanen, M., Puukko, M. & Marjanen, J. (2019). *A-englannin oppimistulokset 7. vuosiluokan alussa 2018*. Julkaisut 13:2019. Kansallinen koulutuksen arviointikeskus. https://karvi.fi/app/uploads/2019/05/KARVI_1319.pdf

Härmälä, M. & Marjanen, J. (2022). *Englantia koronapandemian aikaan. A-englannin oppimistulokset 9. vuosiluokan lopussa 2021.* Julkaisut 22:2022. Kansallinen koulutuksen arviointikeskus. https://karvi.fi/wp-content/uploads/2022/09/KARVI_2222.pdf

IELTS (2007). Weir, C., O'Sullivan, B., Yan, J. & Bax, S. (2007). Does the computer make a difference? The reaction of candidates to a computer-based versus a traditional hand-written form of the IELTS Writing component: effects and impact. *IELTS Research Reports, Volume 7*. British Council and IELTS Australia Pvt Limited. https://www.ielts.org/-/media/research-reports/ielts_rr_volume07_report6.ashx

IELTS (2013). *Biometric systems further tighten IELTS test security*. https://www.ielts.org/news/2013/biometric-systems-further-tighten-ielts-test-security

IELTS (2022). *IELTS Identity Verification*. https://www.idp.com/hongkong/ielts-hk/ielts-registration/identity-verification/?lang=en

Jacobsen, J (2020). *How to conduct remote user tests successfully.* TestingTime. https://www.testingtime.com/en/blog/how-to-conduct-remote-user-tests-successfully/

Jakku-Sihvonen, R. 1997. Johdanto. Teoksessa R. Jakku-Sihvonen (toim.), *Onnistuuko oppiminen—oppimistuloksien ja opetuksen laadun arviointiperusteita peruskoulussa ja lukiossa*. Arviointi 3/1997. Opetushallitus.

Jauhiainen, T., Lennes, M., & Marttila, T. (toim.) (2019), *Suomenkielisen tekoälyn kehittämisohjelma—esiselvitys*. Valtion kehitysyhtiö. https://vake.fi/wp-content/uploads/Vake-suomenkielisen-teko%C3%A4lyn-kehitt%C3%A4minen-esiselvitys-2019.pdf

Jerrim, J., Micklewright, J., Heine, J.-H., Salzer, C., & McKeown, C. (2018). PISA 2015: how big is the 'mode effect' and what has been done about it? *Oxford Review of Education, 44*(4), 476–493. 10.1080/03054985.2018.1430025

Johns Hopkins (2021). Johns Hopkins performs its first augmented reality surgeries in patients. *NeuroLogic*, Winter 2021. https://www.hopkinsmedicine.org/news/articles/johns-hopkins-performs-its-first-augmented-reality-surgeries-in-patients

Julin, S. & Rautopuro, J (2016). *Läksyt tekijäänsä neuvovat. Perusopetuksen matematiikan oppimistulosten arviointi 9. vuosiluokalla 2015*. Julkaisut 20:2016. Kansallinen koulutuksen arviointikeskus. https://karvi.fi/app/uploads/2016/04/KARVI_2016.pdf

Kaarakainen, M.-T., Kivinen, O., & Hutri, H. (2015). Pelit ja pelaaminen sosiaalisena oppimisympäristönä. Teoksessa R. Koskimaa, J. Suominen, F. Mäyrä, J. T. Harviainen, U. Friman, & J. Arjoranta (toim.), *Vuosikirja2015*. Suomen Pelitutkimuksen Seura. https://www.pelitutkimus.fi/vuosikirja2015/artikkeli-pelit-ja-pelaaminen-sosiaalisena-oppimisymparistona

Katajamäki, H. 2011. Maailmasta on kysymys. Sanomalehtiyliopisto kevät 2011. Vaasan yliopisto. Julkaisu n.o 34. *Kestävä kehitys*. http://www.uva.fi/materiaali/pdf/isbn_978-952-476-400-1.pdf

Kauppinen, M. & Marjanen, J. (2020). *Millaista on yhdeksäsluokkalaisten kielellinen osaaminen?—Suomen kielen ja kirjallisuuden oppimistulokset perusopetuksen päättövaiheessa 2019.* Julkaisuja 13:2020. Kansallinen koulutuksen arviointikeskus. https://karvi.fi/wp-content/uploads/2020/08/KARVI_1320.pdf

Kerner, S. M., & Gillis, A. S. (2022). *Web 3.0 (Web3)*. Definition. https://www.techtarget.com/whatis/definition/Web-30

Kitchin, R. (2013). Big data and human geography: Opportunities, challenges and risks. *Dialogues in Human Geography, 3*(3), 262–267. https://doi.org/10.1177/2043820613513388

Kreiner, S. (2021). Haastattelu 6.9.2021. *Folkeskolen*. Riise, A.B., "Svend Kreiner: Derfor var de adaptive test ikke så god en ide, som vi troede". https://www.folkeskolen.dk/evaluering-ledelse-matematik/svend-kreiner-derfor-var-de-adaptive-test-ikke-sa-god-en-ide-som-vi-troede/1361560

Krishnan, K. (2020). *Our education system is losing relevance. Here's how to unleash its potential.* World Economic Forum. Apr. 13, 2020. https://www.weforum.org/agenda/2020/04/our-education-system-is-losing-relevance-heres-how-to-update-it/?utm_source=sfmc&utm_medium=email&utm_campaign=2716680_Agenda_weekly-17April2020&utm_term=&emailType=Newsletter

Kuuluvainen, V., Virtanen, I., Rikkonen, L., & Isotalus, P. (2021). Testing an Immersive Virtual Environment for Decreasing Intergroup Anxiety among University Students: An Interpersonal Perspective. *International Journal of Emerging Technologies in Learning (iJET)*, *16*(16), 202–220. https://doi.org/10.3991/ijet.v16i16.19673

Kuusi, O. (1999). Teknologien kehityksen heikot signaalit. *FUTURA* 2/1999, 8–15.

Lavonen, J. & Salmera-Aro, K. (2022). Experiences of moving quickly to distance teaching and learning at all levels of education in Finland. Teoksessa F. M. Reimers (toim.), *Primary and secondary education during Covid-19. Disruptions to educational opportunity during a pandemic* (ss. 105–124). Springer. https://link.springer.com/content/pdf/10.1007%2F978-3-030-81500-4.pdf

Linacre, J. M. (2000). Computer-adaptive testing: A methodology whose time has come. Teoksessa S. Chae, U. Kang, E. Jeon, & J. M. Linacre, *Development of Computerized Middle School Achievement Tests*. MESA Research Memorandum. Komesa Press. https://www.rasch.org/memo69.pdf

Litts, B.K., Searle, K.A., Brayboy, B.M.J. & Kafai, Y.B. (2020). Computing for all?: Examining critical biases in computational tools for learning. *British Journal of Educational Technology, 52*(2), 842–857. https://doi.org/10.1111/bjet.13059

Mannermaa, M. (1991). *Evolutionaarinen tulevaisuustutkimus. Tulevaisuustutkimuksen paradigmojen ja niiden metodologisten ominaisuuksien tarkastelua*. Tulevaisuuden tutkimuksen seura, Acta FUTURA Fennica n:o 2. VAPK-kustannus, Helsinki.

Mazzeo, J. (2016). Large-scale computer-based tests in the United States: Expectations, experiences, successes and some challenges for the future. Presentation on conference *Future Tests and Test Environments*, June 16, 2016, Umeå, Sweden.

Meristö, T. (1993). Skenaariotyöskentely strategisessa johtamisessa. Teoksessa M. Vapaavuori (toim.), *Miten tutkimme tulevaisuutta?* (ss. 215–221) Tulevaisuuden tutkimuksen seura. Acta *FUTURA* Fennica n:o 5. Tulevaisuustutkimuksen keskus.

Metcalfe J., Kornell, N. & Finn, B. (2009). Delayed versus immediate feedback in children's and adult's vocabulary learning. *Memory & Cognition*, *37*, 1077–1087. http://www.ncbi.nlm.nih.gov/pubmed/19933453

Metsämuuronen, J. (2001a). *Sosiaali- ja terveysalan tulevaisuutta etsimässä*. International Methelp Oy.

Metsämuuronen, J. (2001b). Uuden vuosituhannen haasteet sosiaali- ja terveysalalla. Teoksessa J. Metsämuuronen, *Sosiaali- ja terveysalan tulevaisuutta etsimässä* (ss. 162–175). International Methelp Oy.

Metsämuuronen, J. (2009). *Tutkimuksen tekemisen perusteet ihmistieteissä*. Tutkijalaitos. International Methelp Oy.

Metsämuuronen, J. (2013). Effect of repeated testing to the development of Biblical Hebrew language proficiency. *Journal of Educational and Psychological Development,* 3(1), 10–24. http://dx.doi.org/10.5539/jedp.v3n1p10

Metsämuuronen, J. & Lehikko, A. (2022). Challenges and possibilities of educational equity and equality in the post-COVID-19 realm in the Nordic countries. *Scandinavian Journal of Educational Research*. https://doi.org/10.1080/00313831.2022.2115549

Metsämuuronen, J. & Mattsson, M. (2013). Effect of repeated testing to the development of vocabulary, nominal structures, and verbal morphology. *Journal of Educational and Psychological Development,* 3(2), 89–101. http://dx.doi.org/10.5539/jedp.v3n2p89

Metsämuuronen, J. & Nousiainen, S. (2023).Yleiset menetelmäratkaisut matematiikan oppimistulosten arvioinnissa vuonna 2021. Teoksessa J. Metsämuuronen & S. Nousiainen (toim.), *Matematiikkaa COVID-19-pandemian varjossa II. Menetelmälliset ratkaisut matematiikan 9. luokan arvioinnissa keväällä 2021* (ss. 21–82). Julkaisut 5:2023. Kansallinen koulutuksen arviointikeskus.

Metsämuuronen, J. & Räsänen, P. (2018). Cognitive–linguistic and constructivist mnemonic triggers in teaching based on Jerome Bruner's thinking. *Frontiers in Psycholology*, 9:2543. https://doi.org/10.3389/fpsyg.2018.02543

Metsämuuronen, J. & Seppälä, H. (2021). *COVID-19-pandemia, osaamisvaje ja osaamisen eriytyminen*. Policy Brief 1:2021. Kansallinen koulutuksen arviointikeskus. https://karvi.fi/wp-content/uploads/2021/12/KARVI_Policy_brief_0121.pdf

Metsämuuronen, J. & Suomilammi, M. (2023). Matematiikan osaamisen eriytyminen ja osaamisen heikentymistä selittäviä tekijöitä. Teoksessa J. Metsämuuronen & S. Nousiainen (toim.), *Matematiikkaa COVID-19-pandemian varjossa II. Menetelmälliset ratkaisut matematiikan 9. luokan arvioinnissa keväällä 2021* (ss. 127–172). Julkaisut 5:2023. Kansallinen koulutuksen arviointikeskus.

Metsämuuronen, J. & Ukkola, A. (2019). *Alkumittauksen menetelmällisiä ratkaisuja*. Julkaisut 18:2019. Kansallinen koulutuksen arviointikeskus. https://karvi.fi/wp-content/uploads/2019/08/KARVI_1819.pdf

Miller, P. (2021). Online Learning Trends to Support Students and Teachers in 2022. *EdTech Digest*. Dec 01, 2021. https://www.edtechdigest.com/2021/12/01/online-learning-trends-to-support-students-and-teachers-in-2022/

Molina, C. A., Theodore, N., Ahmed, A. K., Westbroek, E. M., Mirovsky, Y., Harel, R., Orru', E., Khan, M., Witham, T., & Sciubba, D. M. (2019). Augmented reality–assisted pedicle screw insertion: a cadaveric proof-of-concept study. *Journal of Neurosurgery: Spine SPI, 31*(1), 139–146. https://doi.org/10.3171/2018.12.SPINE181142

Mullis, I. V. S., & Martin, M. O. (toim.). (2017). *TIMSS 2019 Assessment Frameworks*. Boston College, TIMSS & PIRLS International Study Center. http://timssandpirls.bc.edu/timss2019/frameworks/

Mullis, I. V. S., Martin, M. O., Foy, P., Kelly, D. L., & Fishbein, B. (2020). *TIMSS 2019 International Results in Mathematics and Science*. Boston College, TIMSS & PIRLS International Study Center. https://timssandpirls.bc.edu/timss2019/international-results/

Nardi, A. & Ranieri, M. (2019). Comparing Paper-Based and Electronic Multiple-Choice Examinations with Personal Devices: Impact on Students' Performance, Self-Efficacy and Satisfaction. *British Journal of Educational Technology*, 50(3), 1495–1506. https://doi.org/10.1111/bjet.12644

Nobel (2014). *Press release. 06.10.2014.* https://www.nobelprize.org/nobel_prizes/medicine/laureates/2014/press.html

Nyyssölä, K. & Kumpulainen, T. (2020). *Perusopetuksen ja kouluverkon tulevaisuusnäkymiä.* Raportit ja selvitykset 2020:25. Opetushallitus. https://www.oph.fi/sites/default/files/documents/Perusopetuksen_ja_kouluverkon_tulevaisuudennakymia.pdf

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M. Huotilainen, M., Iivonen, A., Vaivio. M., Alku, P., Ilmoniemi, R.J., Luuk, A., Allik, J., Sinkkonen, J., & Alho, K. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature, 385*, 432–434. https://doi.org/10.1038/385432a0

OAK (2017). *Ohje opettajille. ViLLE Team.* Oppisanalytiikan keskus. Turun yliopisto. https://ville.cs.utu.fi/opintopolku/villeohje-opintopolku-%20opettajille.pdf

OECD (2013). *Technical Report of the Survey of Adult Skills (PIAAC)*. 2. laitos. OECD. http://www.oecd.org/skills/piaac/_Technical%20Report_17OCT13.pdf

OECD (2016). *Technical Report of the Survey of Adult Skills (PIAAC)*. 2. laitos. OECD. http://www.oecd.org/skills/piaac/PIAAC_Technical_Report_2nd_Edition_Full_Report.pdf

OPH (1998). *Koulutuksellisuuden tuloksellisuuden arviointimalli*. Arviointi 7/98. Opetushallitus.

OPH (2018). *Työllisyyden ja osaamisen muutoksia. Osaamisen ennakointifoorumin skenaariotyön tuloksia*. https://www.oph.fi/sites/default/files/documents/tyollisyyden-ja-osaamisen-muutoksia-oef-vaihe-iii-er3-koulutus-kulttuuri-ja-viestinta_1.pdf

Osguthorpe, R. T., & Graham, C. R. (2003). Blended learning environments: Definitions and directions. *Quarterly Review of Distance Education, 4*(3), 227. https://www.learntechlib.org/p/97576/

Papas, C. (2021). *The 10 Best Blended Learning LMS Solutions (2022)*. https://elearningindustry.com/best-blended-learning-lms-solutions

PISA (2010). *PISA Computer-Based Assessment of Student Skills in Science. OECD*. Publication: 24/8/2010. http://www.oecd.org/edu/school/programmeforinternationalstudentassessmentpisa/pisacomputer-basedassessmentofstudentskillsinscience.htm

PISA (2015a). Main results from the PISA 2012 Computer-based Assessments. Teoksessa OECD, *Students, Computers and Learning. Making the Connection* (ss. 81–103). http://dx.doi.org/10.1787/9789264239555-6-en

PISA (2015b). *Students, Computers and Learning. Making the Connection*. OECD. http://dx.doi.org/10.1787/9789264239555-en

PISA (2018). Annex A5. How comparable are the PISA 2018 computer- and paper-based tests? Teoksessa *PISA 2018 Results (Volume I): What Students Know and Can Do*. OECD. https://www.oecd-ilibrary.org/sites/8f293551-en/index.html?itemId=/content/component/8f293551-en

Price, W. N. II, & Cohen, I. G. (2019). Privacy in the age of medical big data. *Nature Medicine, 25*(1), 37–43. http://dx.doi.org/10.1038/s41591-018-0272-7

Regeringen (2021). *Bred aftale om fremtidigt evaluerings- og bedømmelsessystem*. Børne- og Undervisningsministeriet, 29.10.2021. https://www.regeringen.dk/nyheder/2021/bred-aftale-om-fremtidigt-evaluerings-og-bedoemmelsessystem

Roediger, H.L. 3rd & Karpicke, J.D. (2006a). The power of testing memory. Basic research and implications for educational practice. *Perspectives on Psychological Science*. *1*(3), 181–210. http://dx.doi.org/10.1111/j.1745-6916.2006.00012.x

Roediger, H. L., & Karpicke, J. D. (2006b). Test-enhanced learning: taking memory tests improves long-term retention. *Psychological Science*, *17*(3), 249–255. http://dx.doi.org/10.1111/j.1467-9280.2006.01693.x

Rowland, D. C., Roudi, Y., Moser, M. B., & Moser, E. I. (2016). Ten years of grid cells. *Annual Review of Neuroscience, 8*(39), 19–40. https://doi.org/10.1146/annurev-neuro-070815-013824

Räkköläinen, M, Metsämuuronen J., Holopainen J., & Hievanen R. (2017). *Kestävän kehityksen osaaminen, opetus ja koulutuksen järjestäjän toiminta ammatillisissa perustutkinnoissa*. Julkaisut 12:2017. Kansallinen koulutuksen arviointikeskus. https://karvi.fi/wp-content/uploads/2017/05/KARVI_1217.pdf

Räsänen, P., Aunio, P., Laine, A., Hakkarainen, A., Väisänen, E., Finell, J., Rajala, T., Laakso, M.-J., & Korhonen, J. (2021). Effects of gender on basic numerical and arithmetic skills: Pilot data from 3rd to 9th grade for a large-scale online dyscalculia screener. *Frontiers in Education*, 6:683672. https://doi.org/10.3389/feduc.2021.683672

Salonen, A. O. (2010). *Kestävä kehitys globaalin ajan hyvinvointiyhteiskunnan haasteena*. Tutkimuksia 318. Helsingin yliopisto, Käyttäytymistieteellinen tiedekunta. https://helda.helsinki.fi/bitstream/handle/10138/20067/kestavak.pdf

Sandal, G. (2021). The metaverse is coming: How should organisations prepare? Nov. 30. *Futures Platform*. https://www.futuresplatform.com/blog/metaverse-how-should-organisations-prepare

Sattar, M., Palaniappan, S., Lokman, A., Shah, N., Khalid, U. & Hasan, R. (2020). Motivating medical students using virtual reality based education. *International Journal of Emerging Technologies In Learning (IJET), 15*(02), 160–174. http://dx.doi.org/10.3991/ijet.v15i02.11394

Shaheen, N.L. (2021). Accessibility4Equity: Cripping technology-mediated compulsory education through sociotechnical praxis. *British Journal of Educational Technology, 53*(1), 77–92. https://doi.org/10.1111/bjet.13153

Shdaifat, A. M., Obeidallah, R. A., Ghazal, G., Abu Sarhan, A., & Abu Spetan, N. R. (2020). A proposed iris recognition model for authentication in mobile exams. *International Journal of Emerging Technologies in Learning (iJET)*, *15*(12), 205–216. https://doi.org/10.3991/ijet.v15i12.13741

Silverström, C. & Rautopuro J. (2015). *Språk och skrivande i årskurs 9. En utvärdering av lärresultat i modersmål och litteratur våren 2014*. Publikationer 18:2015. Nationella Centret för Utbildningsutvärdering. https://karvi.fi/wp-content/uploads/2015/09/KARVI_1815.pdf

Sitra (2020). *Megatrendit 2020*. Sitra. https://www.sitra.fi/aiheet/megatrendit/

Sjögren, A., Engdahlc, M., Hall, C., Holmlund, H., Lundin, M., Mühlrad, H., & Öckert, B. (toim.) (2021). *Swedish children and youth during the COVID-19 pandemic*. Working Papers 2021:3. The Institute for Evaluation of Labour Market and Education Policy. https://www.ifau.se/globalassets/pdf/se/2021/wp-2021-03-swedish-children-and-youth-during-the-covid-19-pandemic.pdf

Soini, K. (2013). Kestävä kehitys ja kulttuuri. Teoksessa M. Laine (toim.), *Kestävä kasvatus—Kulttuuria etsimässä* (ss. 12–25). Suomen kulttuuriperintökasvatuksen seuran julkaisuja 6.

Sternberg, R. J. & Grigorenko, E. L. (2002). *Dynamic testing*. Cambridge University Press.

Taivassalo, M. (2019). *Uudistuvat oppimisympäristöt ja digitaaliset ratkaisut oppimisen tukena—esimerkkejä erilaisista oppimisympäristöistä ja -ratkaisuista*. Zoomin alueellinen koulutus Turku 10.9.2019. Opetushallitus. https://www.oph.fi/sites/default/files/documents/uudistuvat_oppimisymparistot_minna_taivassalo.pdf

TOEFL (2007). *Test and Score Data Summary for TOEFL®. Computer-Based and Paper-Based Tests*. Educational Testing Service. https://www.ets.org/Media/Research/pdf/TOEFL-SUM-0506-CBT.pdf

TOEFL (2022). *TOEFL® Test Security: Detection*. https://www.ets.org/toefl/score-users/about/security/detection/

Tulving, E. (1967). The effects of presentation and recall of material in free-recall learning. *Journal of Verbal Learning and Verbal Behavior, 6*, 175–184. http://dx.doi.org/10.1016/S0022-5371(67)80092-6

Ukkola, A. & Metsämuuronen, J. (2019). *Alkumittaus—Matematiikan ja äidinkielen ja kirjallisuuden osaaminen ensimmäisen luokan alussa*. Julkaisut 17:2019. Kansallinen koulutuksen arviointikeskus. https://karvi.fi/wp-content/uploads/2019/07/KARVI_1719.pdf

Ukkola, A. & Metsämuuronen, J. 2021. *Matematiikan ja äidinkielen ja kirjallisuuden osaaminen kolmannen luokan alussa*. Julkaisut 20:2021. Kansallinen koulutuksen arviointikeskus. https://karvi.fi/wp-content/uploads/2021/08/KARVI_2021.pd

Ukkola, A., & Metsämuuronen J. (2023). *Matematiikan ja äidinkielen taidot alkuopetuksen aikana. Perusopetuksen oppimistulosten pitkittäisarviointi 2018–2020.* Julkaisut 1:2023. Kansallinen koulutuksen arviointikeskus. https://karvi.fi/wp-content/uploads/2023/01/KARVI_0123.pdf

van Boxel, M. (2016). *Implementing large-scale computer-based high-stakes testing in the Netherlands*. Presentation on conference Future Tests and Test Environments, June 16, 2016, Umeå, Sweden.

Van der Linden, W. J., & Glas, C.A.W. (toim.). (2000). *Computerized adaptive testing: Theory and practice*. Kluwer.

Veltheim, H. (2021). Future of education after COVID-19: AI becomes the teacher while humans mentor and coach. Jun 1. *Futures Platform*. https://www.futuresplatform.com/blog/future-education-after-covid-19-ai-becomes-teacher-while-humans-mentor-and-coach

Venäläinen, S., Laimi, T., Seppälä, S., Vuojus, T., Viitala, M., Ahlholm, M., Latomaa, S., Mård-Miettinen, K., Nirkkonen, M., Huhtanen, M., & Metsämuuronen, J. (2022). *Kielellisiä taitoja ja koulunkäyntivalmiuksia—Valmistavan opetuksen ja oman äidinkielen opetuksen tila ja vaikuttavuus -arviointi.* Julkaisut 19:2022. Kansallinen koulutuksen arviointikeskus. https://karvi.fi/wp-content/uploads/2022/09/KARVI_1922.pdf

Vie, J.-J., Popineau, F., Bruillard, É., & Bourda. Y. (2017) A review of recent advances in adaptive assessment. Teoksessa Learning Analytics: Fundaments, Applications, and Trends, 94, Springer International Publishing, (ss.113–142). *Studies in Systems, Decision and Control 978-3-319-52976-9. hal-01488284*. https://hal.archives-ouvertes.fr/hal-01488284/document

Vojdanoska M., Cranney, J. & Newell, B.R. (2009). The testing effect: The role of feedback and collaboration in a tertiary classroom setting. *Applied Cognitive Psychology*, *24*(8), 1183–1195. http://dx.doi.org/10.1002/acp.1630

Wainer, H. (toim.). (2000). *Computerized adaptive testing: A Primer*. 2. laitos. Lawrence Erlbaum Associates

Weiss, D.J., & Kingsbury, G.G. (1984). Application of computerized adaptive testing to educational problems. *Journal of Educational Measurement, 21*(4), 361–375. https://doi.org/10.1111/j.1745-984.1984.tb01040.x

Wise S.L. (2019). Controlling construct-irrelevant factors through computer-based testing: disengagement, anxiety, & cheating. *Education Inquiry, 10*(1), 21–33. https://doi.org/10.1080/20004508.2018.1490127

YK (1987). *Report of the World Commission on Environment and Development, General Assembly Resolution 42/187*, 11 December 1987. Yhdistyneet Kansakunnat. http://www.un-documents.net/a42-427.htm

YK (2010). *Culture and development. Assembly Resolution 65/166, 20 December 2010.* Yhdistyneet Kansakunnat. http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CLT/pdf/8_UNGA_Resolution_A_RES_65_166_EN.pdf

YK (2020). *Policy Brief. Education during COVID-19 and beyond.* August 2020. Yhdistyneet Kansakunnat. https://www.un.org/development/desa/dspd/wp-content/uploads/sites/22/2020/08/sg_policy_brief_covid-19_and_education_august_2020.pdf

YTL (2022). *Digitaalisten kokeiden kuvaukset*. Ylioppilastutkintolautakunta. https://www.ylioppilastutkinto.fi/ylioppilastutkinto/digitaalinen-ylioppilastutkinto/digitaalisten-kokeiden-kuvaukset

Åkerlund, C., Marjanen, J., & Lepola L. (2019). *Lärresultat i finska i åk 6. Resultat av en utvärdering våren 2018*. Publikationer 6:2019. Nationella Centret för Utbildningsutvärdering. https://karvi.fi/wp-content/uploads/2019/04/KARVI_0619.pdf

Åkerlund, C., Marjanen, J. & Peltola, E. (2022). *A-finska i åk 9. Utvärdering av lärresultat våren 2021*. Publikationer 4:2022. Nationella Centret för Utbildningsutvärdering. https://karvi.fi/wp-content/uploads/2022/03/NCU_0422.pdf

4

Matematiikan osaamisen eriytyminen ja osaamisen heikentymistä selittäviä tekijöitä

Jari Metsämuuronen, Karvi
Matti Suomilammi, Karvi

4

- 9. luokan aineisto poikkesi aiemmista selvästi siinä, että se muodostui yhden jakauman sijaan kolmesta jakaumasta: heikosti suoriutuneista, keskitasoisesti suoriutuneista ja erittäin hyvin suoriutuneista oppilaista.
- Nämä kolme populaatiota mallinnettiin ja etsittiin tekijöitä, jotka selittävät kuulumista heikosti menestyvään populaatioon näiden oppilaiden tunnistamiseksi ja auttamiseksi.
- Osaamisen kansallinen heikkeneminen viimeisten vuosien aikana näyttää syntyvän kahdella perusmekanismilla. Yhtäältä sekä heikoimpia että parhaimpia arvosanoja saaneiden oppilaiden näytetty osaaminen on laskenut tehtävissä, joissa digitaalisella välineellä on vaikutusta oikean vastauksen antamiseen. Toisaalta näyttää siltä, että heikoimpia arvosanoja saaneiden oppilaiden tarkan lukemisen taito ja viitseliäisyys ovat heikentynyt vuosien varrella.

4.1 Johdattelua teemaan

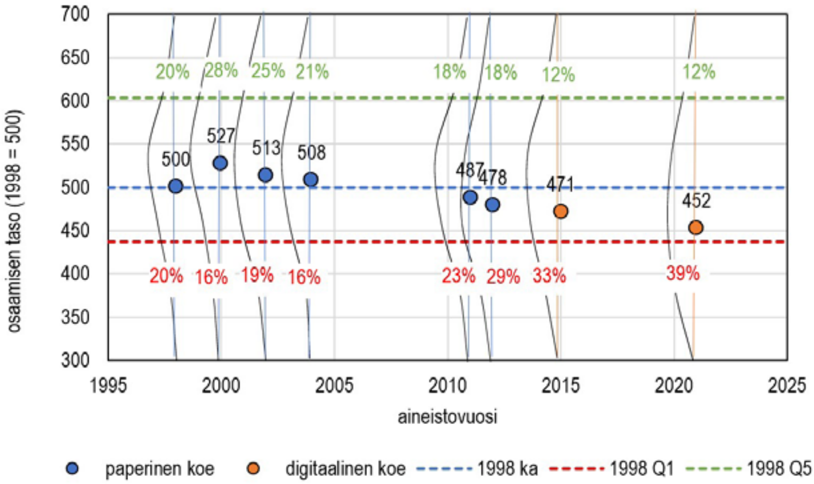
Sekä kansainvälisissä että kansallisissa oppimistulosarvioinneissa on havaittu, että matemaattisen osaamisen taso on ollut laskujohteinen koko 2000-luvun alkuvuosikymmenien ajan (ks. esimerkiksi OECD, 2019; Metsämuuronen & Nousiainen, 2021). Kansallisissa oppimistulosaineistoissa matematiikan keskiosaamisen havaittu lasku sijoittuu samaan taitekohtaan (vuosi 2000) kuin kansainvälisessä *Harmonized learning outcomes* -aikasarjassa (Altinok, Diebolt, & Demeulemeester, 2014; Altinok, Angrist, & Patrinos, 2018; ks. Kalenius, 2023). Tosin osaamisen lasku on Suomessa saattanut alkaa jo aiemminkin; armeija-aineistossa aritmetiikan testitulosten perusteella näyttää siltä, että osaamisen huippu saavutettiin jo 1990-luvun alkupuolella (VATT, 2018; Kalenius, 2023).

Kansallista osaamisen laskua osana kansainvälistä ilmiötä ovat pohtineet aiemmin mm. Hautamäki ja kollegat (2013) sekä hiljattain Vainikainen ja Hautamäki (2022). Keskustelussa eräänä juonteena esiin on tullut ns. käänteinen Flynnin ilmiö (*anti-Flynn effect;* mm. Flynn & Shayer, 2018). James R. Flynnin mukaan nimetty Flynnin ilmiö (*Flynn effect*) tarkoittaa sitä, että älykkyystesteillä mitattu "älykkyys" tai kognitiiviset taidot näyttävät kasvavan sukupolvien myötä pitkällä aikavälillä (mm. Flynn, 1984, 1987, 2009a, 2009b). Havaittu mitatun osaamisen lasku ja älykkyystesteissä heikentyneet tulokset monissa maissa ovat saaneet tutkijat eri maissa pohtimaan, voisiko kyseessä olla käänteinen Flynnin ilmiö: kognitiivinen kapasiteetti olisi alkanut laskea. Toisaalta esimerkiksi Shayer, Ginsburg ja Coe (2007, 2009) esittivät jo 15 vuotta sitten Englannissa osaamisen muutoksen syyksi liiallisen oppilaslähtöisten menetelmien korostumisen, joka suosi matalampaa ajattelun prosessointia. Hollannissa Woodley ja Meisenberg (2013) näkivät heikentyneiden testitulosten taustalla biologisten tekijöiden ohella mm. madaltuneet koulutukselliset tavoitteet ("*worsening educational standards*"). Vainikainen ja Hautamäki (2022) arvelevat oppimaan oppimisen aineistojen perusteella Suomessa osaamisen laskussa kyseessä olevan kognitiivisten taitojen rapautumisen sijaan motivaation puutteesta ja laiskistumisesta testeissä, joilla ei ole suurta merkitystä oppilaan tulevaisuuden kannalta (ns. *low-stake* testit).

Kun matematiikan osaamisen kansallinen taso laskee, se voi teknisesti tapahtua ainakin neljällä tavalla: (1) kaikkien osaaminen laskee tasaisesti, (2) parhaiden oppijoiden osaamisen taso laskee, mutta heikompien ja keskitasoisten osaaminen pysyy samalla tasolla kuin aiemmin, (3) heikompien ja keskitasoisten osaajien osaamisen taso laskee, mutta parhaat osaajat pysyvät aiemmalla tasolla, tai niin, että (4) heikkojen osaamisen taso laskee, mutta keskitasoiset ja parhaat osaajat pysyvät yhtä hyvinä kuin aiemmin. Kansallisten 9. luokan matematiikan aineistojen perusteella kyseessä näyttää olevan useamman vaihtoehdon hybridi. Asiaa tarkastellaan neljästä näkökulmasta.

4.1.1 Osaamisen taso heikkenee kaikissa taitotasoryhmissä

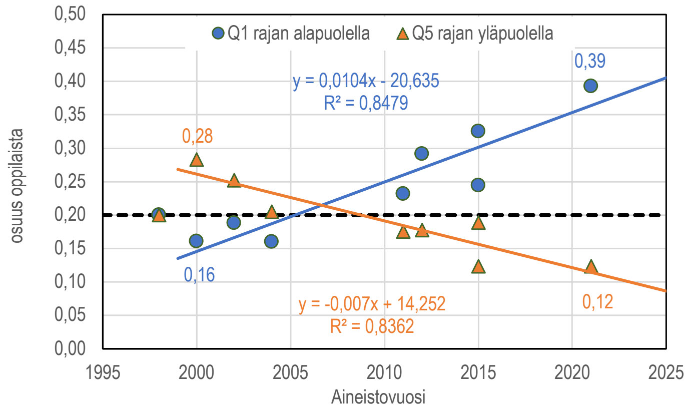
Kansallisten aineistojen perusteella näyttää ensiksikin siltä, että kaikkien oppilaiden osaamisen taso on ollut laskujohteinen vuodesta 2002 lähtien niin, että jos vuoden 1998 jakauman alinta viidennestä eli alakvintiilin osoittamaa osaamisen tasoa (Q1 = 437 pistettä) pidetään standardina, vuoden 2004 mittauksen jälkeen tämän rajan alle jääneiden oppilaiden osuus on lisääntynyt vuosi vuodelta niin, että vuoden 2021 aineistossa *kahteen alimpaan* viidennekseen kuuluvat oppilaat ovat samalla tasolla kuin vuoden 1998 aineistossa alin viidennes; heikkojen oppilaiden määrä on siis *kaksinkertaistunut* (Kuvio 20). Vastaavasti toisessa ääripäässä vuonna 1998 ylimpään viidennekseen eli yläkvintiiliin kuuluvien oppilaiden osaamisen tason (Q5 = 603 pistettä) saavutti vuonna 2021 enää 12 prosenttia oppilaista. Kaikissa osaamisryhmissä osaamisen taso on siis laskenut vuoden 2000 jälkeen oleellisesti.



KUVIO 4.20. Matematiikan 9. luokan aineistojen jakaumat sekä heikoimmin ja parhaimmin suoriutuneiden oppilaiden osuudet

4.1.2 Heikommin suoriutuvien oppilaiden osaamisen taso heikkenee enemmän kuin parhaimmin suoriutuneiden oppilaiden

Toinen näkökulma osaamisen muutoksen tulee siitä, että heikoimmin suoriutuneiden oppilaiden osuus on lisääntynyt hieman enemmän kuin parhaimmin suoriutuneiden oppilaiden osuus on vähentynyt. Jos standardina käytetään osaamiseltaan vuoden 1998 alimpaan viidennekseen kuuluvia oppilaita (437 pistettä tai vähemmän), tämän tasoisten oppilaiden osuus oli vuonna 2000 16 % ja vuonna 2021 39 % eli määrä on lisääntynyt 23 prosenttiyksikköä. Vastaavana aikana osaamiseltaan ylimpään viidennekseen sijoittuvien oppilaiden osuus on laskenut 16 prosenttiyksiköllä eli hieman vähemmän (Kuvio 21). Ero ääriryhmien muutoksissa ei ole suuri, mutta se voi osaltaan selittää kuitenkin sitä, miksi vuoden 2021 matematiikan osaamisen jakauman saattoi tulkita olevan kolmihuippuinen (Metsämuuronen & Nousiainen, 2021): heikoimmin suoriutuneiden oppilaiden joukko on eriytynyt aiempaa selkeämmin keskiryhmästä ja parhaimmin suoriutuneiden oppilaiden joukosta (ks. myöhemmin Kuvio 23).



KUVIO 4.21. Ääriosaamisen eriytyminen matematiikan aineistoissa

4.1.3 Kaikkein heikoimmin suoriutuvien oppilaiden osuus ei ole noussut merkittävästi

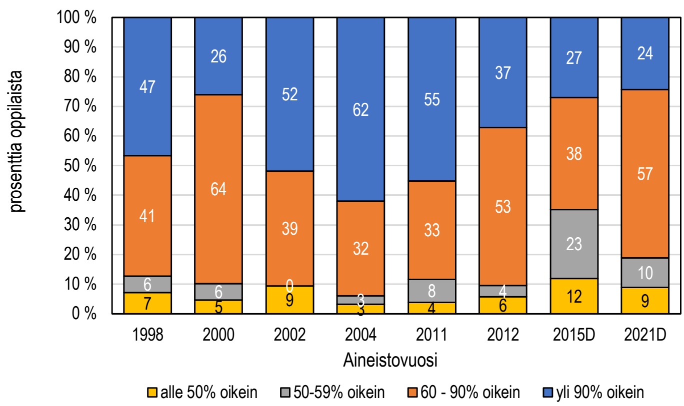
Kolmas näkökulma osaamisen muutokseen tulee siitä, että kansallisten aineistojen perusteella näyttää siltä, että taidoiltaan kaikkein heikoimpien oppilaiden osaamisen taso ei ole juuri muuttunut vuosien varrella, mutta tätä *hieman parempien* eli matemaattisilta taidoiltaan hieman keskitasoa heikommin suoriutuvien oppilaiden osaaminen näyttää laskeneen. Kaikkein heikoimpien oppilaiden osaamista voisi osittain selittää ns. lattiaefekti: heikoimmat eivät pääse enää alemmas. Tämä ei kuitenkaan selitä ilmiötä, sillä hyvin harva oppilas ei vuosien varrella olisi saanut yhtäkään tehtävää oikein ratkaistuksi, eikä asian suhteen ole juuri tapahtunut muutosta vuosien varrella. Tämä johtuu siitä, että monivalintatehtäviä on tehtäväsarjoissa niin paljon, että on vaikea saada nollaa pistettä, jos ylipäätään vastaa tehtäviin.

Ilmiön selvittämistä ja ymmärtämistä varten kaikkiin 9. luokan aineistoihin laskettiin ns. ”helppojen” osioiden summa siten, että kaikkina mittausvuosina laskettiin yhteen ne tehtävät, joissa oikean ratkaisun antoi paperi-kynä-testeissä yli 80 % oppilaista. *Item response theory* (IRT) -mallinnuksen näkökannalta näicden osioiden vaikeustasoparametri sai arvokseen $B = -1{,}000$ tai vähemmän. Raportin liitteenä julkaistavassa tehtäväsarjassa tämän tasoisia tehtäviä oli 15.[22] Tehtävät sisältävät mm. yksinkertaisia matemaattisten termien muistamistehtäviä (tehtävät 8.1–8.5, 11.1–11.2, 12.1–12.3), päässälaskuna suoritettavan kahden desimaalin vähennyslaskun, jossa on kymmenylitys (tehtävä 3.), päättelytehtävän, jossa on laskettava kolmella jaollisten lukujen todennäköisyys (tehtävä 14.), sekä perinteisiä yksinkertaisia sievennys- (tehtävä 9.), prosentti- (tehtävät 10. ja 19.) ja funktiotehtäviä (tehtävä 16.2) sekä polynomin ratkaisutehtäviä (tehtävä 20.). Näistä muistamista edellyttävät tehtävät ovat sellaisia, jotka ovat ratkaistavissa enintään kuudennen luokan oppimäärällä. Vaikka tehtävät ovat vaikeustasoltaan helppoja, edellyttävät ne kuitenkin jonkin verran työtä.

22 Nämä olivat julkaistavan tehtäväsarjan tehtävät 3, 8.1, 8.3, 8.4, 8.5, 9, 10, 11.1, 11.2, 12.1, 12.3, 14, 16.2, 19, ja 20. Raportin liitteessä nämä tehtävät on merkitty *-symbolilla

Systemaattinen muutos heikkojen oppilaiden osuuksissa saattaisi näkyä siten, että erittäin pieniä pistemääriä saaneiden määrä nousisi systemaattisesti. Tätä indikoimaan laskettiin niiden oppilaiden osuudet, joiden tulos jäi alle 50 % maksimipistemäärästä, joka on joissain *standard setting* -menettelyissä perinteisesti ajateltu ns. ”*weak pass*” rajaksi (ks. keskustelu esimerkiksi Metsämuuronen, 2013; Van der Schoot, 2009). Tämä 50 %:n raja tarkoittaa, että oppilas olisi läpäissyt ”helpon” testin pienimmällä mahdollisella pistemäärällä ja tätä pienempi ratkaistujen tehtävien määrä ei vielä riitä vakuuttamaan, että taso olisi saavutettu. Tälle indikaattorille laskettiin kolme rinnakkaista indikaattoria: niiden osuus, jotka saivat kaikki tai lähes kaikki tehtävät oikein (yli 90 % oikein), niiden osuus, jotka juuri pääsivät alarajan yli (50–60 % oikein) ja näiden väliin jääneiden oppilaiden osuudet (60–90 % oikein).

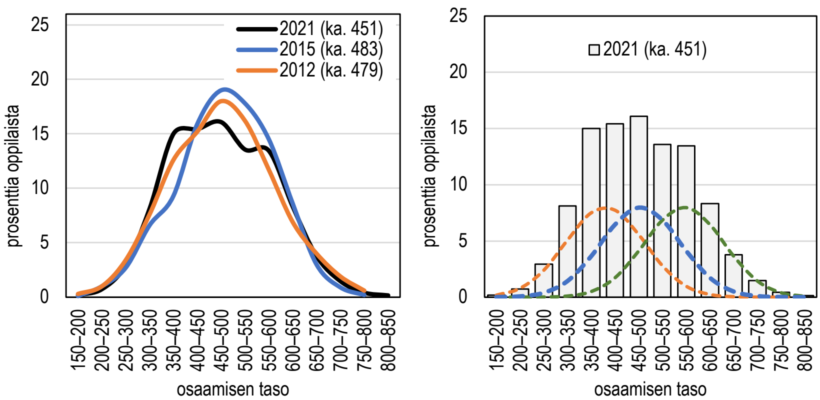
Aineiston perusteella näyttää siltä, että erittäin heikkojen oppilaiden osuus on pysynyt alle 12 prosentin tasolla kaikissa mittauksissa (Kuvio 22). Vuonna 2021 näiden oppilaiden osuus oli 9 %, sama prosenttiosuus oli myös vuoden 2002 aineistossa ja lähes sama myös vuoden 1998 aineistossa (7 %). Se, että aivan heikkojen oppilaiden osuus ei ole kasvanut lienee hyvä uutinen, ellei ajattele, että edelleenkin 9. luokalle asti selviytyy varsin alkeellisilla matemaattisilla taidoilla. Huomiota herättää myös se, että enintään 50 % oikein ratkaisseiden määrässä ei havaita systemaattista muutosta, mutta niiden osuus, jotka saivat tässä helpossa testissä täysiä tai lähes täysiä pisteitä (yli 90 % oikein), vähenee systemaattisesti vuodesta 2004 (62 %) vuoteen 2021 (24 %). Näyttää siltä, että keskitasoiset tai jopa tätä paremmat oppilaat eivät siis joko aidosti osaa tai jostain syystä näytä osaamistaan samalla intensiteetillä kuin aiempina vuosina.



KUVIO 4.22. Kaikkein helpoimpien tehtävien muodostaman ”testin” ratkaisuprosenttijakaumat eri vuosina

4.1.4 Osaamisen ääripäät ovat eriytyneet aiempaa selvemmin toisistaan

Neljäs näkökulma matematiikan osaamisen muutokseen on se, että osaamisen jakauma on muuttunut viimeisimpään mittaukseen tultaessa selvästi (Metsämuuronen & Nousiainen, 2021)—tätä ilmiötä tarkastellaan tarkemmin tässä artikkelissa. Näyttää siltä, että aiemmissa arvioinneissa niiden havaintojen määrä, jotka rikkoivat normaalijakauman, oli selvästi nykyistä vähäisempi. Vuoden 2021 arvioinnissa niiden määrä oli jo sillä tasolla, että aiempi yksihuippuinen normaalijakauma oli muuttunut selvästi kolmihuippuiseksi. Nämä kolme (osittain päällekkäistä) jakaumaa edustavat seuraavia osaamisryhmiä: heikot osaajat, keskitasoiset osaajat ja hyvät osaajat (Kuvio 23). Metsämuurosen (2019) koulutuksellisen epätasa-arvon indikaattoreiden typologiassa tätä kutsutaan jakaumaepätasa-arvoksi (*Population disparity*). Viime vuosien aikana heikoimpien ja parhaiden oppilaiden väliset erot ovat tulleet oleellisesti selkeämmiksi.



KUVIO 4.23. Osaamisen kokonaisjakauma vuonna 2021 ja vuosina 2012 ja 2015 (Metsämuuronen & Nousiainen, 2021)

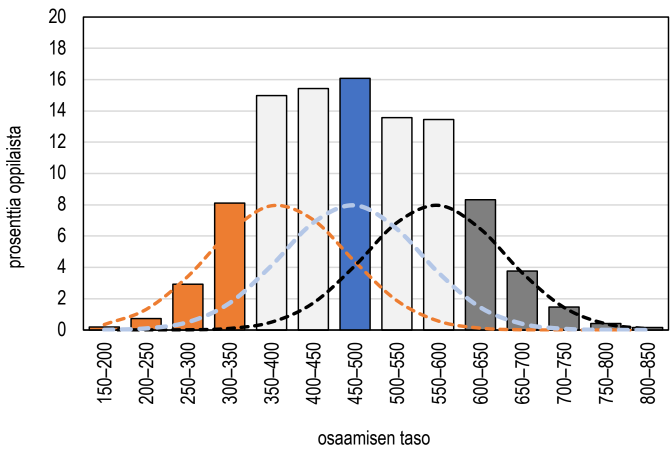
Epävarmaa on, mikä on ilmiön syy, ja nähtäväksi jää, palautuuko jakauma takaisin normaalimmaksi tulevina vuosina. Yhtäältä kuitenkin näyttää siltä, että COVID-19-pandemian ja esimerkiksi etäkoulukäytänteiden ja kotoa saatavan tuen erojen myötä erot oppilasryhmien välillä ovat ainakin tilapäisesti kasvaneet. Toisaalta osaamisen yleinen lasku herättää huolen heikommin suoriutuvien oppilaiden tilanteesta ja heidän mahdollisesta pysyvämmästä suoritustason heikkoudesta. Tähän olisi kyettävä puuttumaan mahdollisimman varhaisessa vaiheessa ja oikeisiin oppilaisiin liittyvillä toimilla. Siksi on mielekästä pyrkiä mallintamaan nämä eri populaatiot, jotta tunnistetaan ne oppilaat, jotka potentiaalisesti kuuluvat heikoimmin suoriutuvaan joukkoon ja joille siten voitaisiin kohdentaa tarpeellista tukea.

4.2 Kolmen populaation mallinnus

Eri osaamisryhmien mallinnus tapahtuu kolmessa vaiheessa. Ensimmäisessä vaiheessa valitaan analyysiin mukaan vain kaikkein selkeimmin eri osaamisryhmiä edustavat oppilaat, joiden avulla toisessa vaiheessa etsitään ääriryhmiä kuvaavat keskeiset erottelevat tekijät. Kolmannessa vaiheessa löydettyjen erottelevien muuttujien avulla mallinnetaan kaikkien oppilaiden ennustettu populaatio.

4.2.1 Vaihe 1: Mahdollisimman puhtaiden populaatioiden valitseminen

Oletetaan, että kokonaispopulaatio muodostuu kolmesta toisiinsa limittyvästä populaatiosta kuvion 24 mukaisesti. Koska populaatiot todennäköisesti ovat selvästi päällekkäin, etsitään ensimmäisessä vaiheessa sellaisia oppilaita, jotka edustavat mahdollisimman puhtaasti omaa populaatiotaan ja joiden välistä eroa voi selittää mahdollisimman tarkasti. Kokonaisjakaumasta valittiin korkeintaan 350 pistettä saaneet oppilaat edustamaan alinta populaatiota (populaatio 1), 450–500 pistettä saaneet edustamaan keskipopulaatiota (populaatio 2) ja yli 600 pistettä saaneet edustamaan ylintä populaatiota (populaatio 3). Näin saatuihin alkuvaiheen populaatioihin kuului $n = 5317$ oppilasta ($n_1 = 1488$, $n_2 = 2017$, $n_3 = 1812$).



KUVIO 4.24. Osaamisen kokonaisjakauma vuonna 2021 ja kolme populaatiota

4.2.2 Vaihe 2: Yksittäisten selittävien muuttujien löytäminen

Toisessa vaiheessa käytettiin päätöksentekopuu-analyysia (DTA) selittämään eroja näiden alustavien ryhmien välillä. Tehtävänä oli löytää keskeisesti erottelevat muuttujat ja niiden jakokohdat, joiden avulla voidaan mallintaa myös analyysin ulkopuolelle jääneet havainnot sopiviin ryhmiin. DTA:n vahvuus on, että se pystyy ryhmittelemään muuttujien arvoja parhaan selitysvoiman saavuttamiseksi. Tämän perusteella useimmista muuttujista muodostettiin myöhempää käyttöä varten 0/1-tyyppisiä dummy-muuttujia, joiden avulla voidaan tarkemmin kuvata, kuinka ryhmät poikkeavat toisistaan.

Taulukkoon 16 on koottu joitain muuttujia, jotka erottelevat ryhmiä toisistaan (ks. tarkemmin tämän osuuden Liite 1). Puuttuvien tietojen vuoksi kaikissa muuttujissa on hieman eri otoskoko; taulukon luvut on laskettu havaittujen määrien perusteella. Kaksi muuttujista, tuen taso ja matematiikan arvosana on käsitelty yksityiskohtaisemmin, jotta taulukon merkinnät tulisivat selkeämmäksi. Tehostettua tai erityistä tukea saaneet oppilaat sijoittuivat testitulokseltaan todennäköisesti heikoimmin menestyneeseen ryhmään; 75 % tehostettua tukea saavista oppilaista ja 84 % erityistä tukea saaneista oppilaista sijoittui alapopulaatioon. Vastaavasti yleistä tukea saaneista noin 40 % sijoittui keski- ja yläpopulaatioon. Matematiikan päättöarvosana selittää myös hyvin sijoittumista eri populaatioihin: matalimpia arvosanoja (4–6) saaneista oppilaista 83 % kuului alapopulaatioon, arvosanan 7 tai 8 saaneista 68 % kuului keskipopulaatioon ja arvosanan 9 tai 10 saaneista 73 % kuului yläpopulaatioon.

Profiloinnissa on hyvä huomioida kaksi seikkaa. Ensiksi, vaikka profiloinnissa pyrittiin mahdollisimman selkeisiin ja toisistaan eroaviin ryhmäjakoihin, ilmenee niissä silti päällekkäisyyksiä. Toiseksi, kun muuttujia tarkastellaan toisinpäin, päädytään toisenlaisiin päätelmiin: kaikkiaan 1441 selkeästi alapopulaatioon luokittuvasta oppilaasta, joilta kolmiportaisen tuen tieto oli saatavilla, 786 eli 55 % oli yleisen tuen piirissä, toisin sanoen, heillä ei ollut tehostetun tai erityisen tuen päätöstä. Vastaavasti keskipopulaatioon kuuluvista 93 % ja alustavasti yläpopulaatioon kuuluvista oppilaista 98 % oli yleisen tuen piirissä. Niinpä taulukko 16 tässä muodossa tarjoaa mielenkiintoisempaa tietoa selittää heikoimpaan osaamisryhmään sijoittumista. Jos nimittäin *havaitaan*, että oppilaalla oli erittäin matala arvosana ja hän sai joko tehostettua tai erityistä tukea, voidaan *ennustaa*, että hän todennäköisesti kuului alimpaan populaatioon.

Yksittäisten muuttujien löytymisen jälkeen kaikki merkitseviksi selittäjiksi havaitut muuttujat—ja joitain lisämuuttujia kuten sukupuoli—mallinnettiin yhtä aikaa DTA:lla. Tässä vaiheessa huomattiin, että vaikka esimerkiksi sukupuolen merkitys erillisenä tekijänä ei ole läheskään yhtä selitysvoimainen kuin joidenkin muiden muuttujien, se erottelee joissain pienemmissä ryhmissä oppilaita toisistaan selvästi. Esimerkiksi niiden oppilaiden ryhmässä, jotka saivat arvosanakseen 8 ja jotka opiskelivat muussa kuin joustavan perusopetuksen ryhmässä tai erityisluokalla, poikia sijoittui tyttöihin nähden yli kolminkertainen määrä ylimpään populaatioon. Tämä selittyi sillä, että pojille tyypillisempää oli pyrkiä lukioon ja siellä pitkän matematiikan opintoihin. Siksi muuttuja otettiin mukaan seuraavassa vaiheessa muodostettuihin regressiomalleihin.

TAULUKKO 4.16. Muuttujia, jotka alustavasti selittävät sijoittumista ei populaatioihin

		alapopulaatio	keskipopulaatio	yläpopulaatio	n
	n (korkeimmillaan)	**1753**	**2004**	**1802**	**5559**
1	3-portaisen tuen taso	tehostettu tai erityinen tuki (75−84%)	yleinen tuki (42%)	yleinen tuki (40%)	
	1.1 yleinen tuki	18,1	41,9	40	4351
	1.2 tehostettu tuki	75,2	19,9	4,8	557
	1.3 erityinen tuki	84,3	11,8	3,9	280
2	Matematiikan arvosana	4,5,6 (83 %)	7,8 (68%)	9, 10 (73%)	
	2.1 arvosana 4–6	82,9	26,0	1,0	1235
	2.2 arvosana 7–8	16,4	68,2	26,1	2130
	2.3 arvosana 9–10	0,8	5,8	72,9	1895
3	S2-status	S2-status (59 %)	ei S2-statusta (39 %)	ei S2-statusta (36 %)	
4	Minkä ikäisenä aloitti koulunkäynnin Suomessa	>7 (61 %)	5−7 (39 %)	5−7 (35%)	
5	Erikois- tai erityisluokka	joustava perusopetus, pienryhmä, erityisluokka (87 %)	ei erikoisluokalla tai liikunta, ilmaisu-, viestintä- tai taideluokka (39−44 %)	stem-, matematiikka-, science-, teknologia-, musiikki-, kieliluokat (53−71 %)	
6	Mitä peruskoulun jälkeen	ammatillinen koulutus, töihin, lisävalmiuksia antavat opinnot (55−70 %)	lukio, lyhyt matematiikka (65 %)	lukio, pitkä matematiikka (69 %)	
7	SES[1] äidin ja isän korkein koulutus yhdessä	molemmilla pk, amm. tai muu tai ei tiedä (37−38 %)	muu kuin ”molemmilla yliopisto” (43−44 %)	molemmilla yliopisto, tai toisella yliopisto ja toisella lukio tai amk (58 %)	
8	SES Käytössäni on… SUM (max 5)	0−1 (58 %)	2−5 (39 %)	4−5 (38 %)	
9	SES Kotonani on… SUM (max 5)	0−1 (54 %)	2−3 (42%)	4−5 (50 %)	
10	Miten viihdyt koulussa	erittäin huonosti (53 %)	melko hyvin (41 %)	erittäin hyvin (46 %)	
11	Poissaolojen määrä viimeisen vuoden aikana	>20 pv (47 %)	6−20 pv. (42–44 %)	0−5 pv. (42 %)	
12	Positiivinen tunnetila matematiikkaa kohtaan	ei koskaan tai hyvin harvoin (42−67 %)	joskus tai harvoin (46−49 %)	usein tai lähes aina (60−72 %)	
13	Negatiivinen tunnetila matematiikkaa kohtaan	lähes aina (55 %)	Joskus tai usein (46−49 %)	ei lainkaan tai harvoin (44−56 %)	
14	Tukiopetuksen määrä luokkien 7, 8, ja 9 aikana	>9 kertaa (72 %)	1−8 kertaa (47 %)	ei lainkaan (53 %)	
15	Puuttuvien vastausten osuus (%)	yli 17,7 prosenttia (50 %)	yli 9,7 prosenttia (43 %)	alle 9,7 prosenttia (45−55 %)	

1) SES = sosioekonominen status

4.2.3 Vaihe 3: Kokonaismallin rakentaminen

Joillain muuttujilla on yksittäisinä muuttujina selkeää tilastollisesti merkitsevää erottelevaa vaikutusta sijoittumisessa kolmeen populaatioon (Taulukko 17). Näillä on kuitenkin paljon yhteistäkin selitysosuutta. Kolmannessa vaiheessa rakennettiin erilaisilla muuttujatyypeillä (ns. dummy-muuttujilla ja alkuperäisillä muuttujilla) ja muuttujien määriä vaihtelemalla erilaisia regressiomalleja, joilla pyrittiin vähentämään selittävien muuttujien määrää ja löytämään ne muuttujat, joilla mallissa on itsenäistä selitysvoimaa. Näistä lopulliseksi malliksi valittu esitellään tässä.

Regressiomallinnus tehtiin kahdessa vaiheessa. Ensin mukaan otettiin aiemmassa vaiheessa kaikki DTA:n perusteella valitut muuttujat ja niille ehdotetut ryhmittelyt ja näihin liittyvät dummy-muuttujat (ks. Liite 1) riippumatta siitä, kuinka paljon niissä oli puuttuvaa tietoa. Oppilailta saatu aineisto oli tältä osin puutteellisempi kuin Koski-rekisteristä saatu tieto. Kaikkiaan 30 muuttujasta 17:llä oli itsenäistä vaikutusta (Taulukko 17). Valtaosalle (81 %) oppilaista saatiin malli kaikilla mukaan otetuilla muuttujilla ja niille, joille mallia ei syntynyt ensimmäisessä vaiheessa, se rakennettiin toisessa vaiheessa niillä muuttujilla, joista kaikki tiedot olivat saatavissa. Molemmat mallit tehtiin lineaarisena *Stepwise*-regressiona, jossa malliin jää jäljelle vain tilastollisesti merkitsevät selittäjät, joilla on itsenäistä selitysvoimaa.

Ensimmäisen vaiheen regressiomalliin jääneet 17 muuttujaa selittävät jakoa kolmeen populaatioon melko hyvin. Ainakin 68 prosenttia ilmiöstä tulee selitettyä.[23] Niille 19 % oppilaista, joille laaja malli ei tuottanut ennustetta johtuen puuttuvista tiedoista, se laskettiin toisessa vaiheessa matematiikan arvosanan, sukupuolen, kolmiportaisen tuen tason, testin puuttuvien tietojen määrän, toisen asteen opintojen suunnitelman ja pienryhmään kuulumisen perusteella. Saatu jatkuva-asteikkoinen ennustemuuttuja muutettiin kolmiportaiseksi normaaleja pyöristyssääntöjä noudattaen.

Mallinnuksen tuloksena syntyy kolme populaatiota kuten kuvissa 23 ja 24 alustavasti hahmoteltiin, mutta ne eivät olekaan samansuuruisia, kuten aluksi ehkä ajateltiin. Näyttää siltä, ja on uskottavaa, että kansallinen aineisto muodostuu ensisijaisesti keskipopulaatiosta (52 % oppilaista), joka muodostaa hieman positiiviseen suuntaan leventyneen normaalijakauman (Kuvio 25). Tästä perusjakaumasta on erillään kaksi pienempää populaatiota, keskitasoa heikompien oppilaiden alapopulaatio (23 % oppilaista) ja keskitasoa parempien oppilaiden yläpopulaatio (25 % oppilaista).

Koska tarkoituksena oli pyrkiä erottelemaan heikoin populaatio keskipopulaatiosta, saatua mallia paranneltiin korjaamalla jakaumia niin, että keskipopulaatiosta olisi hieman kapeampi. Käytännössä kaikki arvosanan 10 saaneet oppilaat sijoitettiin yläpopulaatioon ja arvosanan 5 oppilaat alapopulaatioon riippumatta alkuperäisen mallin tuottamasta sijoituksesta ja alimmissa osaamisryhmissä myös arvosanan 6 saaneet oppilaat sijoitettiin alapopulaatioon. Osaamisluokkien rajoilla heikkoja pistemääriä saaneet sijoitettiin yhtä luokkaa alemmas ja parhaita pistemääriä saaneet yhtä luokkaa ylemmäs. Näin siis osa keskipopulaation oppilaista siirrettiin osaamisen perusteella ala-

23 Metsämuuronen (2022) perustelee, miksi perinteinen selitysaste (R^2) on aina aliarvio, mikäli yksikin malliin mukaan tulevista muuttujista on mitattu eri asteikolla kuin muut muuttujat. Tässä tilanteessa mukana on sekä binäärejä että jatkuvia muuttujia, joten on odotettavaa, että selitysaste on ainakin jossain määrin liian matala todelliseen selitysasteeseen nähden. Siksi käytetään sanamuotoa ”ainakin”.

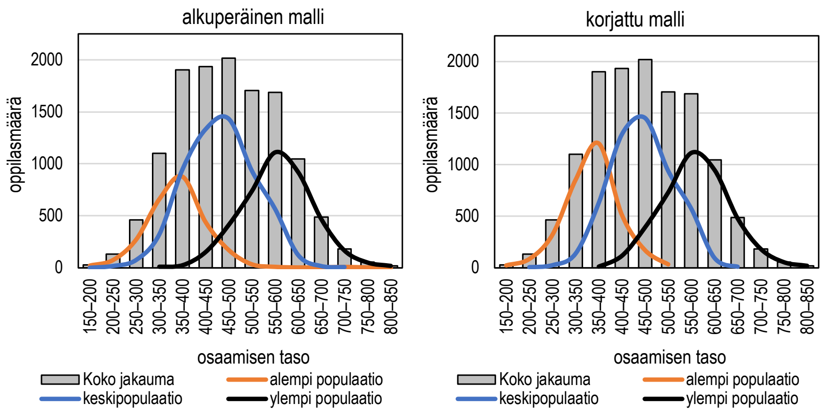
ja yläpopulaatioon. Kuvio 25 havainnollistaa mallien eroja. Tästä eteenpäin käsitellään korjattua mallia ja sen mukaisia oppilasryhmiä. Tässä mallissa keskipopulaatioon kuuluu 46 % oppilaista, alapopulaatioon 21 % ja yläpopulaatioon 33 % oppilaista.

TAULUKKO 4.17. Ensimmäisen vaiheen regressiomalli (81 % oppilaista)

Muuttujat[1]	Standardoimattomat kertoimet		Tilastollinen päättely	
	B	Keskivirhe	t	Merkitsevyys
Vakio	1,961	0,046		
Matematiikan arvosana 9–10 (Dummy)	0,603	0,021	29,034	<,001
Matematiikan arvosana 4–6 (Dummy)	-0,392	0,022	-18,095	<,001
Tehostettu tai erityinen tuki (Dummy)	-0,218	0,023	-9,511	<,001
Peruskoulun jälkeen Lukio, pitkä matematiikka (Dummy)	0,222	0,023	9,729	<,001
Tukiopetuksen määrä luokkien 7, 8 ja 9 aikana (1 = en kertaakaan, 2 = 1–3 kertaa, 3 = 4–8 kertaa, 4 = 9 kertaa tai enemmän)	-0,074	0,009	-8,446	<,001
SES Kotonani on… (mm. taide-esineitä, musiikki-instrumentteja) 0–1 viidestä (Dummy)	-0,167	0,021	-7,930	<,001
Puuttuvien tietojen osuus testissä (%)	-0,007	0,001	-8,141	<,001
S2-status (1 = kyllä, 0 = ei ole)	-0,231	0,028	-8,165	<,001
Positiivinen tunnetila (1 = ei lainkaan, 2 = harvoin, 3 = joskus, 4 = usein, 5 = lähes aina)	0,047	0,009	5,291	<,001
Erikoisluokka Jopo, pienryhmä, erityisluokka (Dummy)	-0,236	0,039	-6,086	<,001
SES Äidin korkein koulutus Yliopisto (Dummy)	0,075	0,018	4,223	<,001
Negatiivinen tunnetila (1 = ei lainkaan, 2 = harvoin, 3 = joskus, 4 = usein, 5 = lähes aina)	-0,034	0,007	-4,555	<,001
SES Kotonani on… (mm. taide-esineitä, musiikki-instrumentteja) 2–3 viidestä (Dummy)	-0,044	0,017	-2,609	0,009
Peruskoulun jälkeen Lukio, lyhyt matematiikka (Dummy)	0,063	0,022	2,923	0,003
Erikoisluokka matematiikka, STEM, yrittäjyys (Dummy)	0,126	0,046	2,750	0,006
SES Äidin korkein koulutus Peruskoulu, Ammatillinen, En tiedä (Dummy)			2,473	0,013
Erikoisluokka Musiikki, Kielet (Dummy)	0,043	0,017	2,202	0,028

1) Selitettävänä Populaatiot123 (vain äärihavainnot); muuttujat järjestetty t-testisuureen itseisarvon mukaiseen järjestykseen

R	R^2	Adj. R^2	Estimaatin keskivirhe
0,826	0,683	0,681	0,428



KUVIO 4.25. Osaamisen kokonaisjakauma vuonna 2021 ja kolme populaatiota

Muistetaan, että valtaosa heikosti suoriutuvista oppilaista ei saa tehostettua tai erityistä tukea. Koulutuksen kehittämisen näkökannalta merkityksellistä voi olla tunnistaa ne oppilaat, jotka kuuluvat alapopulaatioon, jotta heille voidaan tarjota riittävästi tukea. Seuraavissa luvuissa tutkitaan yhtäältä sitä, kuinka alapopulaatioon kuuluvia oppilaita voidaan tunnistaa (luku 4.3) ja toisaalta sitä, millaisissa matematiikan oppimiseen ja osaamiseen liittyvissä sisällöllisissä ja teknisissä asioissa heikko matematiikan taito näkyy verrattuna keskipopulaatioon, millaisia muutoksia asiassa on tapahtunut vuosien aikana ja miten näitä oppilaita voisi ehkä tukea (luku 4.4). Jälkimmäinen tarkastelu on alustavaa ja perustuu ensisijaisesti julkaistavan version avulla tehtävään osaamisen kuvaamiseen.

4.3 Alimpaan populaatioon sijoittuvien oppilaiden tunnistaminen

Alapopulaatioon sijoittuvien oppilaiden ennustamiseen käytetään ensisijaisesti logistista regressioanalyysia (LRA; 1 = alapopulaatio, 0 = muut populaatiot) ja tätä tukemaan DTA-mallitusta. Näistä LRA perustuu lineaariseen mallitukseen. Siinä kaikkia muuttujia käsitellään samanarvoisina ja muuttujilla on mallissa merkitystä vain yhdessä muiden malliin tulleiden muuttujien kanssa. DTA puolestaan perustuu epälineaariseen mallitukseen ja muuttujia käsitellään hierarkkisesti. DTA:n avulla voidaan profiloida oppilaita pienempiin ryhmiin, joista joissain todennäköisyys kuulua alapopulaatioon voi olla oleellisesti korkeampi kuin jossain toisessa ryhmässä. Tämä havainnollistuu tuonnempana.

LRA löytää 30 muuttujan joukosta 13 muuttujaa, joilla aineistossa pystytään profiloimaan alapopulaatioon kuuluvat oppilaat 95 prosentin varmuudella (Taulukko 18). Näistä vaikuttavin ennustetekijä on matematiikan matala kouluarvosana. Heikko arvosana (5 tai 6) lisää valtavasti (4000-kertaisesti) tilastollista ”riskiä” kuulua heikoimpaan osaamisryhmään verrattuna siihen, että oppilas olisi saanut 7 tai sitä korkeamman arvosanan. Suuri riski on tietenkin ilmeistä ja seurausta siitä, että muuttuja on mukana perusmallissa, jossa oppilaat luokiteltiin eri populaatioihin ja siitä, että matalia arvosanoja käytettiin korjaamaan populaatioita eriyttävämpään suuntaan. Siksi ehkä mielenkiintoisempia ovat muut ennustetekijät. Näitä ovat kuuluminen joustavaan perusopetukseen, pienryhmään tai erityisluokalle ja S2-status, joissa molemmissa on noin 7-kertainen riski sekä tehostetun tai erityisen tuen päätös (6-kertainen riski). Myös sosioekonomiseen taustaan linkittyvä kodin vähäinen kulttuuriresurssi (4-kertainen riski) sekä tukiopetuksen suurempi määrä ennustavat matalampaan populaatioon kuulumista selvästi.

TAULUKKO 4.18. Alimpaan populaatioon sijoittuvien oppilaiden ennustaminen

Muuttujat[1,2]	B	S.E.	Wald	df	Sig.	Exp(B)	1/exp
Vakio	-6,682	0,477	195,964	1	<,001	0,001	
Matematiikan arvosana 4–6 (Dummy)	8,296	0,39	453,445	1	<,001	4006,602	
Matematiikan arvosana 7 tai 8 (Dummy)	2,796	0,357	61,351	1	<,001	16,374	
Luokka: Jopo, pienryhmä, erityisluokka (Dummy)	1,988	0,197	101,764	1	<,001	7,303	
S2-status (1 = kyllä, 0 = ei ole)	1,971	0,165	142,931	1	<,001	7,178	
Kolmiportaisen tuen taso tehostettu tai erityinen tuki (Dummy)	1,85	0,118	245,052	1	<,001	6,358	
SES Kotonani on… (mm. taide-esineitä, musiikki-instrumentteja) 0–1 viidestä (Dummy)	1,345	0,135	99,608	1	<,001	3,837	
Tukiopetuksen määrä (1 = en kertaakaan, 2 = 1–3 kertaa, 3 = 4–8 kertaa, 4 = 9 kertaa tai enemmän)	0,613	0,05	151,654	1	<,001	1,845	
sukupuoli 1 = poika, 2 = tyttö	0,485	0,107	20,393	1	<,001	1,624	
SES Kotonani on… (mm. taide-esineitä, musiikki-instrumentteja) 2–3 viidestä (Dummy)	0,344	0,127	7,401	1	0,007	1,411	
Negatiivinen tunnetila (1 = ei lainkaan, 2 = harvoin, 3 = joskus, 4 = usein, 5 = lähes aina)	0,137	0,051	7,293	1	0,007	1,146	
Positiivinen tunnetila (1 = ei lainkaan, 2 = harvoin, 3 = joskus, 4 = usein, 5 = lähes aina)	-0,371	0,059	39,657	1	<,001	0,690	1,449
Peruskoulun jälkeen Lukio, lyhyt matematiikka (Dummy)	-0,679	0,118	33,311	1	<,001	0,507	1,972
Peruskoulun jälkeen Lukio, pitkä matematiikka (Dummy)	-1,843	0,198	86,772	1	<,001	0,158	6,329

1) Logistinen regressioanalyysi, Conditional forward -ratkaisu; selitettävänä sijoittuminen alapopulaatioon (kun vaihtoehtona on sijoittua kahteen muuhun populaatioon); muuttujat järjestetty ”riski”-tekijän mukaan (Exp(B))

Cox & Snell R Square	Nagelkerke R Square	oikein ennustettu
0,567	0,831	94,5 %

Jos oppilas aikoo kaikista heikkoa osaamista indikoivista tekijöistä huolimatta mennä lukioon, tällä on alapopulaatioon kuulumista ehkäisevä vaikutus: jos tavoitteena on lukio ja pitkän matematiikan opinnot, oppilaalla on muista tekijöistä huolimatta 6-kertainen ”riski” päätyä alapopulaatiota korkeampaa populaatioon (Taulukko 3). Lukiossa lyhyen matematiikan opintoihin suuntautuvilla oppilailla ”riski” päätyä ylempiin populaatioihin on kaksinkertainen. Tätä edesauttaa myös matematiikkaa kohtaan tunnettu positiivinen tunnetila. Kiinnostava yksityiskohta on, että ennustemuuttujien yhteydessä tytöillä on poikia hieman suurempi riski sijoittua alapopulaatioon

(1,6-kertainen riski). DTA-mallituksessa arvosanoiltaan keskitasoisilla pojilla on tietyissä erityisryhmissä[24] tyttöjä enemmän intressiä mennä lukioon, mikä nostaa positiivista ”riskiä” sijoittua keskipopulaatioon.

Vaikka yläpopulaation ennustetekijät eivät ehkä ole niin kiinnostavia kuin alapopulaation, voidaan todeta, että ylimpään populaatioon kuulumista ennustaa selkeimmin arvosanojen 9 tai 10 saaminen sekä jatkokoulutusvalintana lukio ja siellä pitkän matematiikan opinnot (15-kertainen ”riski”).

4.4 Alimman ja keskipopulaation rajapintatarkasteluja

Tarkastellaan lähemmin alimman ja keskipopulaation rajapintaa: mikä erottaa toisistaan alimpaan ja keskipopulaatioon kuuluvat oppilaat? Asiaa tarkastellaan kolmesta näkökulmasta: kuvaamalla muuttujia, jotka erottelevat nämä ryhmät toisistaan (luku 4.4.1) ja tarkastelemalla millainen on tyypillinen osaaminen näissä ryhmissä (luku 4.4.2). Jälkimmäistä seikkaa tarkennetaan tutkimalla, kuinka osaaminen on muuttunut vuosien aikana ja kuinka se näkyy linkkitehtävissä (luku 4.4.3). Luvuissa 4.4.1 ja 4.4.2 tarkastelu rajataan vain heikoimpiin ja keskitasoisiin oppilaisiin.

4.4.1 Mitkä tekijät erottavat toisistaan heikot ja keskitasoiset oppilaat

Ala- ja keskipopulaatioon luokittuvien oppilaiden keskeinen erottelija aineistossa on matematiikan arvosana: lähes kaikki arvosanan 4–6 saaneet kuuluivat alapopulaatioon eikä sinne kuulunut yhtään arvosanan 9 tai 10 saanutta oppilasta (Taulukko 19). Sen sijaan korkeita arvosanoja saaneita oppilaita sijoittui keskipopulaatioon. Kuten edellä kolmen populaation tapauksessa, lukioon tähtääminen erottelee selvästi oppilaat ala- ja keskipopulaatioiden välillä; jos oppilas tähtää lukion pitkän matematiikan opintoihin, hänellä on 6-kertainen ”riski” sijoittua keskipopulaatioon ja jos suunnitelmissa on lukion lyhyen matematiikan oppimäärä, ”riski” on 2-kertainen. Alapopulaation suuntaan ennustavat joustavaan perusopetukseen ryhmään, pienryhmä tai erityisluokalla oleminen (8-kertainen riski kuulua alapopulaatioon), S2-status (8-kertainen riski), tehostetun tai erityisen tuen saanti (6-kertainen ”riski”), matala kodin kulttuuriresurssien määrä (3-kertainen riski) ja suurempi tukiopetuksen määrä yleisen tuen piirissä.

4.4.2 Miten heikkojen ja keskitasoisten oppilaiden osaaminen poikkeaa toisistaan

Heikoimpien ja keskitasoisten oppilaiden rajapintaa tarkastellaan seuraavassa kahdesta näkökulmasta. Asiaa tarkastellaan luvussa 4.1.3 käsitellyn ”helpon testin” avulla: millaista osaamista on oppilailla, jotka eivät saa oikein edes puolia kaikkein helpoimmista tehtävistä. Toiseksi tarkastellaan asiaa ala- ja keskipopulaation erojen avulla: millaiset tehtävät erottelevat parhaiten oppilaita näihin kahteen ryhmään.

24 Tällainen erityisryhmä on esimerkiksi ne oppilaat, jotka saivat arvosanakseen 8 ja jotka opiskelivat muussa kuin joustavan perusopetuksen ryhmässä tai erityisluokalla. Tässä ryhmässä poikia sijoittui tyttöihin nähden yli kolminkertainen määrä ylimpään populaatioon. Tämä selittyi sillä, että pojille tyypillisempää oli pyrkiä lukioon ja siellä pitkän matematiikan opintoihin.

TAULUKKO 4.19. Alimpaan ja keskipopulaatioon sijoittuvien oppilaiden ennustaminen

Muuttujat[1]	B	S.E.	Wald	df	Sig.	Exp(B)	1/ Exp(B)
Vakio	-3,656	0,311	138,399	1			
Matematiikan arvosana 4−6 (Dummy)	5,548	0,174	1022,174	1	<0,001	256,774	
Luokka: Jopo, pienryhmä, erityisluokka (Dummy)	2,072	0,2	106,999	1	<0,001	7,939	
S2-status (1 = kyllä, 0 = ei ole)	2,043	0,169	145,404	1	<0,001	7,713	
Tehostettu tai erityinen tuki (Dummy)	1,753	0,121	208,305	1	<0,001	5,77	
SES[2] Kotonani on… (mm. taide-esineitä, musiikki-instrumentteja, kirjoja) 0−1 viidestä (Dummy)	1,034	0,118	77,354	1	<0,001	2,813	
Tukiopetuksen määrä (1 = en kertaakaan, 2 = 1–3 kertaa, 3 = 4–8 kertaa, 4 = 9 kertaa tai enemmän)	0,629	0,051	152,178	1	<0,001	1,876	
sukupuoli 1 = poika, 2 = tyttö	0,509	0,11	21,52	1	<0,001	1,663	
Negatiivinen tunnetila (1 = ei lainkaan, 2 = harvoin, 3 = joskus, 4 = usein, 5 = lähes aina)	0,146	0,052	7,94	1	0,005	1,157	
SES Kotonani on… (mm. taide-esineitä, musiikki-instrumentteja) 4−5 viidestä (Dummy)	-0,355	0,129	7,513	1	0,006	0,701	1,427
Positiivinen tunnetila (1 = ei lainkaan, 2 = harvoin, 3 = joskus, 4 = usein, 5 = lähes aina)	-0,372	0,061	37,649	1	<0,001	0,690	1,449
Peruskoulun jälkeen Lukio, lyhyt matematiikka (Dummy)	-0,677	0,12	31,85	1	<0,001	0,508	1,969
Peruskoulun jälkeen Lukio, pitkä matematiikka (Dummy)	-1,724	0,201	73,585	1	<0,001	0,178	5,618
Matematiikan arvosana 9−10 (Dummy)	-4,501	1,035	18,924	1	<0,001	0,011	90,909

1) Logistinen regressioanalyysi, Conditional forward -ratkaisu; selitettävänä sijoittuminen alapopulaatioon (kun vaihtoehtona on sijoittua keskipopulaatioon); muuttujat järjestetty ”riski”-tekijän mukaan (Exp(B))
2) SES = sosioekonominen status

Cox & Snell R Square	Nagelkerke R Square	oikein ennustettu
0,586	0,797	92,1 %

Heikoimpien oppilaiden osaamisen ominaispiirteitä

Jotta voisi arvioida millainen on kaikkein heikoimmin suoriutuneiden oppilaiden eli ryhmän ”alle 50 % oikein” matemaattinen taso luvussa 4.1.3 kuvatussa helpossa tehtäväsarjassa, tarkastellaan näitä 9. luokan oppilaille periaatteessa helppoja tehtäviä Metsämuurosen (2018) esittelemän matematiikan yleisen viitekehyksen (*common framework in reference of mathematics*, CFM) näkökannalta. Tätä viitekehystä käytettiin myös koulun alkuvaiheen osaamisen tason syventävässä tarkastelussa (Metsämuuronen & Ukkola, 2022).

CFM seuraa nimitysten osalta pitkälti kielten arvioinnissa käytettävän CEF-luokittelun (*Common European Framework of reference for languages* (https://www.coe.int/en/web/common-european-framework-reference-languages) sanamuotoja. ”Helppoon” testiin valitut tehtävät näyttävät heijastavan tasoille A2.1 (Kehittyvä perustaso) ja A2.2 (Toiminnallinen perustaso) määriteltyjä osaamisen sisältöjä (Taulukko 20; ks. tarkemmin Liite 2 tämän osuuden lopussa).

TAULUKKO 4.20. Tiivistetyt tasokuvaukset matematiikan yleisessä viitekehyksessä tasoille A1.3, A2.1 ja A2.2 (Metsämuuronen, 2018)

Taso	Tiivistetty kuvaus osaamisen tasosta "Tällä tasolla oppilas…"
A1.3 Toiminnallinen alkeistaso	• osaa käyttää sujuvasti luonnollisia numeroita alueella $-\infty$ – $+\infty$ • tuntee ja ymmärtää desimaalijärjestelmän paikkajärjestelmänä ja osaa sitä käyttää • ymmärtää yhteen- vähennys-, kerto ja jakolaskun ja osaa käyttää niitä arkielämän tilanteissa • ymmärtää rationaalilukujen käsitteen
A2.1 Kehittyvä perustaso	• käyttää suhdetta, prosentin laskemista ja muita laskentamenettelyjä ratkaistakseen ongelmia arkielämässä vastaan tulevissa tilanteissa • pystyy muodostamaan yksinkertaisen yhtälön ja ratkaisemaan sen joko algebrallisesti tai päättelemällä ratkaistakseen ongelmia arkielämässä vastaan tulevissa tilanteissa • osaa laskea piirin, alueen ja tilavuuden • ymmärtää todennäköisyyden ja satunnaisuuden merkityksen arkielämän tilanteissa • tietää, kuinka koordinaattipisteet määrätään koordinaatistoon
A2.2 Toiminnallinen perustaso	• hallitsee potenssilaskun perusteet ja pystyy kytkemään sen kertolaskuun • hallitsee neliön (toisen potenssin) ja sen yhteyden käytännön tilanteisiin • löytää samanlaiset, yhteneväiset (kongruentit) ja symmetriset muodot ja pystyy näiden avulla tutkimaan kahden kulman ominaisuuksia yksinkertaisessa tilanteessa • lukee erilaisia taulukoita ja diagrammeja ja pystyy määrittämään frekvenssit, keskiarvot, mediaanin ja moodin annetusta aineistosta • tietää, kuinka etsiä lineaarisen funktion nollakohta

CFM:ssa tasoa A2.1 alemmalla tasolla (A1.3, Toiminnallinen alkeistaso) ei vielä käsitteiden osalta hallita esimerkiksi luonnollisten lukujen potensseja, luvun hajottamista tekijöihin tai funktioon liittyviä käsitteitä (ks. tämän osuuden Liite 2). Matemaattisten operaatioiden osalta tällä tasolla ei pystytä käyttämään suhteellisia osuuksia, prosenttilaskuja tai yksinkertaisiakaan yhtälöitä, kolmikulmaiseen kolmioon liittyviä laskuja ja niiden sovelluksia, eikä todennäköisyyteen liittyviä laskuja arkielämän tilanteissa eteen tulevien ongelmien ratkaisemisessa. Matemaattisen ajattelun osa-alueella tällä tasolla ei esimerkiksi kyetä arvioimaan väitteen totuudenmukaisuutta ennakkotietojen perusteella tai muuttamaan arkielämän ongelmaa matemaattiseen muotoon eikä arvioimaan tuloksen mielekkyyttä. Vuoden 2021 aineistossa 9 % oppilaista ei välttämättä osaa siis yksikertaisiakaan prosenttilaskuja tai kykene arvioimaan saamiensa tulosten järkevyyttä.

Tasolla A1.3 oppilaat eivät välttämättä hallitse käsitteitä, joita tarvitaan vaativammissa laskuissa. Tämä rajoittaa esimerkiksi prosenttilaskujen osaamista, sillä tehtävän ratkaisussa oppilas ei pääse eteenpäin, mikäli joku keskeinen termi on jäänyt jo itse tehtävänannossa ymmärtämättä. Toisaalta oppilas ei välttämättä pysty tuottamistehtävissä antamaan riittävän hyvin perusteltuja vastauksia, mikäli hän ei hallitse matematiikan erityissanastoa. Tarkasteltaessa linkkitehtäviä eri vuosilta huomataan esimerkiksi funktiolaskujen alueella muutosta, joka liittyy tehtävien termistöön. Heikko oppilas, joka pystyy ratkaisemaan kuvaajan avulla funktiotehtävän ja valitsemaan vaihtoehdoista oikean vastauksen, ei useinkaan pysty vastaavassa tehtävässä tuottamaan vastausta, koska termistönhallinta on puutteellista. Jos oppilaan matemaattiset perustaidot ovat jääneet heikoiksi, myös lukion matematiikka saattaa tuntua liian hankalalta.

Huomattakoon, että tasoa A2.2 korkeammalla tasolla CFM:ssä siirrytään polynomifunktioiden, trigonometristen funktioiden, vektorilaskennan, analyyttiseen geometrian ja differentiaalilaskennan maailmaan ja tätä vaativampaan ainekseen, joita tyypillisesti Suomessa opetetaan lukiossa eikä näillä tasoille sijoittuvia tehtäviä ollut vuoden 2021 tehtäväsarjoissa. Lienee selvää, että tasoille A1.3 ja A2.1 jääville oppilaille ei käytännössä ole mahdollista selviytyä lukion pitkän matematiikan opinnoista ilman mahdollisia siltakursseja tai kertausta.

Millaisissa tehtävissä paremmin suoriutuvat oppilaat todennäköisesti onnistuvat?

Eräs näkökulma profiloida heikkoja oppilaita suhteessa keskipopulaatioon on katsoa, millaisissa tehtävissä eri populaatioihin kuuluvat oppilaat todennäköisemmin onnistuvat ja epäonnistuvat. Taulukkoon 21a on koottu julkaistavan version tehtävistä 13 selkeimmin populaatioita erottelevaa tehtävää ja taulukkoon 21b on koottu tehtävien ominaispiirteitä. Taulukkoihin on valittu sellaiset osiot, jotka selvimmin lisäävät riskiä sijoittua heikoimpaan osaamisryhmään.

Huomataan, että useimmat tehtävät eivät juuri kykene ennustamaan alapopulaatioon kuulumista, mutta kylläkin kuulumista keskipopulaatioon, joskin korkeimmillaankin ”riskit” ovat kaksinkertaisia. Tästä poikkeuksen tekevät erittäin vaikeat tehtävät 30 ja 34, joissa useimmiten vain parhaat oppilaat löysivät oikean ratkaisun. Näissä tehtävissä keskipopulaatioon kuuluvat oppilaat saivat hieman todennäköisemmin oikean vastauksen, joskin tässäkin populaatiossa vain 1 % onnistui ratkaisemaan tehtävät. Huomattakoon, että ylimmän populaation oppilaistakin vain 12 ja 18 prosenttia sai tehtävät oikein.

TAULUKKO 4.21a. Alimpaan ja keskipopulaatioon sijoittuvien oppilaiden ennustaminen tehtävissä onnistumisen perusteella

Osiot[1]	Julkaistu versio	Ratkaisu-osuus	B	S.E.	Wald	df	Sig.	Exp(B)	1/ Exp(B)
M20S2T13_027A	9	0,826	-0,738	0,141	27,328	1	<,001	0,478	2,09
M00S3T14_172B	32	0,400	-0,680	0,092	54,833	1	<,001	0,507	1,97
M20S3T14_048B	20	0,800	-0,613	0,117	27,323	1	<,001	0,542	1,85
M15S4T15_174	7	0,654	-0,558	0,107	27,262	1	<,001	0,573	1,75
M12S2T11_170A	18	0,668	-0,512	0,107	23,039	1	<,001	0,599	1,67
M20S2T13_027A	10	0,916	-0,508	0,173	8,589	1	0,003	0,602	1,66
M20S3T14_046A	27.1	0,511	-0,503	0,120	17,556	1	<,001	0,605	1,65
M20S1T10_003	1	0,552	-0,499	0,112	19,798	1	<,001	0,607	1,65
M15S5T16_179	15	0,639	-0,472	0,108	19,272	1	<,001	0,623	1,61
M20S5T17_084	29	0,360	-0,451	0,155	8,462	1	0,004	0,637	1,57
M20S2T10_028ABCD	4	0,310	-0,421	0,125	11,256	1	<,001	0,657	1,52
…	…	…	…	…	…	…	…	…	…
MYs17S3T14_196C	34	0,036	1,500	0,291	26,541	1	<,001	4,480	
M20S2T13_103	30	0,045	1,623	0,516	9,881	1	0,002	5,068	

1) Logistinen regressioanalyysi, Conditional forward -ratkaisu; selitettävänä sijoittuminen alapopulaatioon (kun vaihtoehtona on sijoittua keskipopulaatioon); muuttujat järjestetty ”riski”-tekijän mukaan (Exp(B))

	Cox & Snell R Square	Nagelkerke R Square	oikein ennustettu
	0,424	0,575	83,0 %

TAULUKKO 4.21b. Eriytymistä kuvaavien tehtävien piirteitä

Osiot	Julkaistu sarja	yksinkertainen kuvaus	Ratkasu-prosentti alin.	Ratkasu-prosentti keski
M20S2T13_027A	9[1]	yksinkertainen sievennystehtävä	0,68	0,92
M00S3T14_172B	32	yksinkertaisen yhtälön ratkaisutehtävä	0,08	0,37
M20S3T14_048B	20[1]	polynomin ratkaisu sijoittamalla	0,51	0,86
M15S4T15_174	7	yksinkertainen funktion arvon lasku annetulla arvolla	0,35	0,66
M12S2T11_170A	18	yksinkertainen laskujärjestystehtävä	0,43	0,69
M20S2T13_027A	10[1]	ei triviaali prosenttilasku, kahdesta pääteltävä kolmas	0,77	0,96
M20S3T14_046A	27.1	käsiteen tahko muistaminen, yksinkertainen lasku	0,21	0,44
M20S1T10_003	1	päässälasku; pinta-alan perusteella laskettava sivun pituus	0,24	0,55
M15S5T16_179	15	trigonometriaa, kulmien laskenta, yksinkertainen	0,31	0,63
M20S5T17_084	29	suorakulmaisen kolmion kateetin lasku	0,08	0,28
M20S2T10_028ABCD	4	päässälasku, helpon lausekkeen yksinkertaisin muoto	0,18	0,30
…	..	…	..	…
MYs17S3T14_196C	34	yhtälöparin ratkaisutehtävä	0,01	0,01
M20S2T13_103	30	kuvion lukeminen, vaativahko prosenttilasku	0,00	0,01

1) tehtävät olivat mukana ns. ”helpossa” tehtäväsarjassa (ks. luku 4.1.3)

Kun kytketään osioiden sisältö aiemmin esillä olleeseen CFM-viitekehykseen (ks. Taulukko 20; ks. myös tämän osuuden Liite 2), huomataan että laskut ovat tyypillisiä tasoille A2.1 ja A2.2 sijoittuvia tehtäviä ja sielläkin osittain helpoimmasta päästä. Suurimmat erot ratkaisuosuuksissa ovat tehtävissä 20 (polynomin ratkaisutehtävä; 35 prosenttiyksikön ero populaatioiden keskiratkaisuprosenteissa), 15 (kulmien laskentatehtävä; 32 prosenttiyksikköä), 1 (päässälaskuna suoritettava pinta-alan lasku, 31 prosenttiyksikköä), 7 (yksikertaisen funktion ratkaisu sijoittamalla; 31 prosenttiyksikköä) ja 32 (yksinkertaisen yhtälön ratkaisu; 29 prosenttiyksikköä).

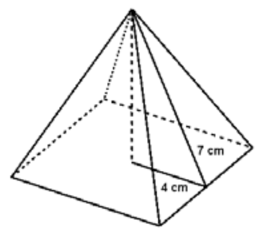
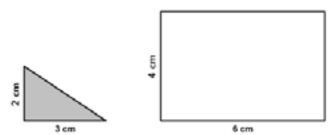
Erottelevimmista tehtävistä 9, 10 ja 20 olivat mukana myös luvussa 4.1 kuvatussa ”helpossa” tehtäväsarjassa. Näissäkin tehtävissä alapopulaatioon sijoittuvista oppilaista yli 50 % löysi oikean ratkaisun. Tehtävät 7, 15 ja 18 ovat linkkitehtäviä, joita käsitellään tarkemmin seuraavassa luvussa.

4.4.3 Osaamisen muutoksen trendejä tehtävätasolla

Aiemmasta tiedetään, että matematiikan osaamisen keskimääräinen taso on ollut laskujohteinen ainakin vuodesta 2000 lähtien. Tätä asiaa tarkastellaan seuraavassa julkaistavien linkkitehtävien avulla. Näiden lisäksi analysoitiin muitakin linkkitehtäviä eri vuosilta. Julkaistavat linkkitehtävät on koottu taulukkoon 22.

TAULUKKO 4.22. Julkaistavat linkkitehtävät ja virheellisten ratkaisujen mahdolliset syyt

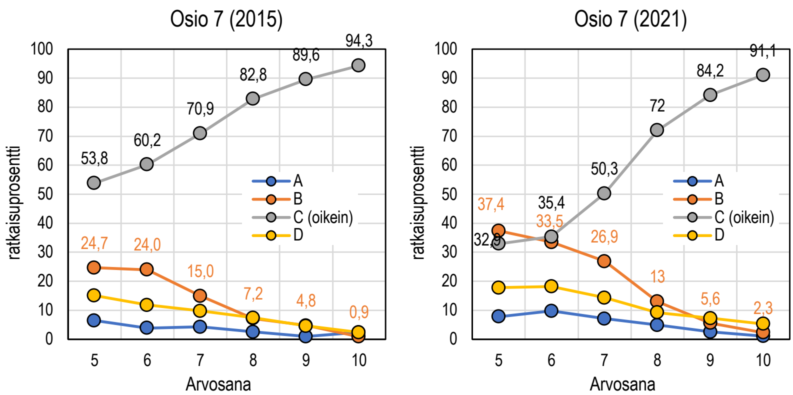
nimi	numero julkaistavassa sarjassa	sisältöalue	tavoitealueen tiivistetty kuvaus	tehtävä	ratkaisu-vaihtoehdot	virheellisen ratkaisun mahdollinen syy
M12S1T10_163	5.	Ajattelun taidot (S1)	laskee päässään, tekee päätelmiä (T10)	Minulta menee 24 minuuttia, kun kävelen koulusta kotiin nopeudella 5 km/h. Kauanko kotimatkani kestää, kun ajan sen pyörällä nopeudella 15 km/h?	Päässälasku	ei osata muuntaa arkielämän tilannetta matemaattisen ongelman muotoon
M15S4T15_174	7.	Funktiot (S4)	tulkitsee ja tuottaa funktion (T15)	Omenan hinta saadaan funktion f(x)=2x+1 arvosta, jossa x on omenoiden määrä (kg). Kuinka paljon maksaa 3,5 kg omenoita?	a) 24,5 € b) 9 € c) 8 € d) 7 €	a) 20 + 3,5 + 1 b) 2×4 + 1 c) oikea ratkaisu d) 2×3,5
M11S6T19_184	14.	Tietojenkäsittely ja tilastot sekä todennäköisyys (S6)	määrittää tilastollisia tunnuslukuja ja laskee todennäköisyyksiä (T19)	Katrilla on alla olevat pelimerkit pussissa ja pussia on ravistettu.  Katri nostaa pussista yhden pelimerkin katsomatta. Millä todennäköisyydellä pelimerkissä oleva luku on jaollinen kolmella?	a) 1/11 b) 1/3 c) 4/11 d) 1/4 e) 4/7	a) pelimerkki 3 kaikkiaan 11 vaihtoehdosta b) kolmella jaollisten määrä **c) oikea ratkaisu** d) kolmella jaollisia 4 ja niitä yksi e) kaikkien tapausten määrä ratkaistu väärin
M15S5T16_179	15.	Geometria (S5)	ymmärtää geometristen käsitteiden yhteyksiä (T16)	Oikokulmassa on kolme kulmaa, a, a ja 50 astetta. Kuinka suuri on kulma a?	a) 20 b) 55 c) 65 d) 75	a) oikokulma sekoittuu suorakulmaan b) yksinkertainen laskuvirhe **c) oikea ratkaisu** d) mekaaninen laskuvirhe
M12S2T11_170A	18.	Luvut ja laskutoimitukset (S2)	hallitsee peruslaskutoimituksia rationaaliluvuilla (T11)	Mitä tulee vastaukseksi laskulausekkeesta 2−5·2−5?	tuottamistehtävä	laskujärjestysvirhe

nimi	numero julkaistavassa sarjassa	sisältöalue	tavoitealueen tiivistetty kuvaus	tehtävä	ratkaisu-vaihtoehdot	virheellisen ratkaisun mahdollinen syy
M12S5T16_177	21.	Geometria (S5)	ymmärtää geometristen käsitteiden yhteyksiä (T16)	Mikä väitteistä on tosi? Ympyröi.  	a) Pyramidin pohja on suorakulmainen kolmio b) Pyramidin korkeus on 7 cm c) Pyramidin pohjaneliön sivun pituus on 4 cm d) Pyramidin pohjaneliön pinta-ala on 64 cm2 e) Pyramidin sivutahkon pinta-ala on 56 cm2	a) ei ymmärrä mikä on pohja/ kolmio b) ei ymmärrä korkeuden määritelmää c) kiinnittää huomiota vain yksittäiseen numeroon **d) oikea ratkaisu** e) unohdettu jakaa kahdella
M12S5T17_178	25.	Geometria (S5)	hyödyntää suorakulmaiseen kolmioon ja ympyrään liittyviä ominaisuuksia (T17)	Kuinka monta alla olevan kaltaista harmaata suorakulmaista kolmiota tarvitaan peittämään tarkalleen suorakulmion pinta?  	a) neljä b) kuusi c) kahdeksan d) kymmenen e) kaksitoista	a) ei hahmoteta kolmion kokoa tai kolmion koko laskettu väärin (2×3) b) 2×3; ymmärtänyt tehtävän väärin **c) oikea ratkaisu** d) 4+6; täysin väärä vastaus e) 4×6/2; kolmion laskukaava ehkä oikein, mutta muutoin väärin operoitu
M00S3T14_172B	32.	Algebra (S3)	ratkaisee yhtälöitä (T14)	Ratkaise yhtälö 5x−7 = 3x. Perustele	Tuottamistehtävä	ei hallita yhtälönratkaisua tai ei haluta perustella asiaa

Monivalintatehtävien ratkaisuissa tapahtuneita muutoksia

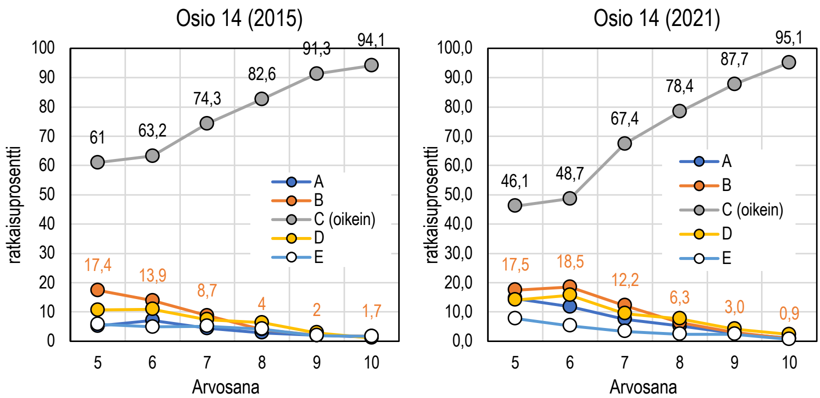
Tarkastellaan ensin yksittäisiä monivalintatehtäviä ja niiden vastauksissa tapahtuneita muutoksia niiden aikapisteiden välillä, jolloin tehtävä oli ensimmäisen kerran linkkitehtävänä ja vuoden 2021 aineistossa. Vuoden 2015 aineistossa vertailukohtana on paperi-kynä-testin tulos, jolloin saadaan tietoa myös välineen vaikutuksesta tulokseen. Koska aiemmilta vuosilta ei ole saatavissa tietoa eri populaatioista, tarkastellaan osaamisen muutosta arvosanaluokittain. Muistetaan, että arvosanan 5 tai 6 saaneet oppilaat kuuluivat pääsääntöisesti alapopulaatioon ja 9 tai 10 saaneet oppilaat pääsääntöisesti yläpopulaatioon.

Osiossa 7 tuli ratkaista yksinkertainen arkielämän tilanteeseen liittyvä funktiotehtävä sijoittamalla annettu luku yhtälöön (Kuvio 26). Vuoteen 2015 nähden kaikissa arvosanaluokissa oikean vastauksen löytymisen todennäköisyys on laskenut hieman, mutta erityisesti arvosanan 5, 6 ja 7 saaneet oppilaat valitsivat vaihtoehdon, jossa annettu arvo 3,5 näytetään pyöristetyn ylöspäin arvoon 4, jolloin funktio tuottaa omenoille liian korkean hinnan (9 €). Ero vastauksissa vuosien välillä on huomattava. On myös mahdollista, että tehtävässä käytetty sanasto on oppilaalle vierasta ja ongelma muodostuu siitä, että heikommat oppilaat eivät tiedä, mitä pitää laskea.



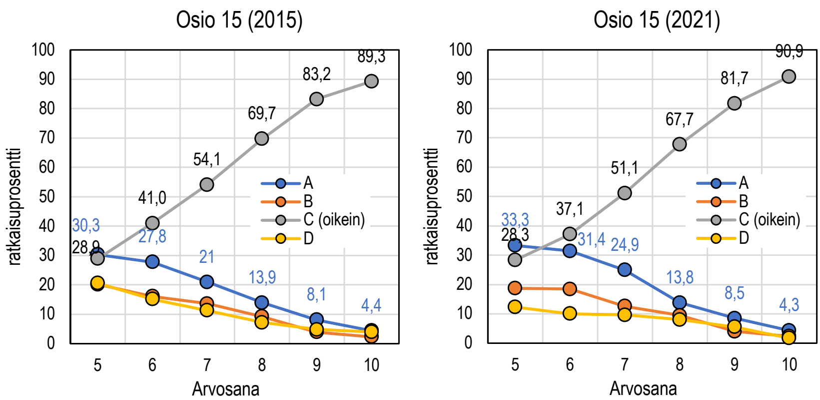
KUVIO 4.26. Tehtävässä 7 suoriutumisen muuttuminen vuosien 2015 paperi-kynä-testin ja 2021 digitaalisen testin välillä

Osiossa 14 tuli ratkaista todennäköisyys saada satunnaisesti kolmella jaollinen luku annettujen lukujen joukosta (Kuvio 27). Ensin siis piti ratkaista kolmella jaollisten määrä (4) ja sitten ratkaista näiden määrä suhteessa kaikkiin lukuihin (11). Tämän tehtävän osalta parhaat oppilaat osasivat ratkaista tehtävän yhtä todennäköisesti kuin aiempana vuonna, mutta matalammissa arvosanaluokissa virheellisten vaihtoehtojen valinta on kasvanut oleellisesti. Erityisesti vaihtoehdon A eli numeron 3 todennäköisyys kaikista luvuista (1/11) suosio on kasvanut aiempaan mittaukseen nähden. Näyttää siis siltä, että oppilaat ovat tunnistaneet numeron kolme, mutta joko jättäneet tunnistamatta sanan ”jaollinen” tai eivät ole jaksaneet lukea tarkasti koko tehtäväksi antoa. Myös vaihtoehto D, jossa ratkaisuksi ehdotetaan numeron 3 osuutta kolmella jaollisista luvuista (1/4), on kasvanut arvosanaluokissa 5–6.



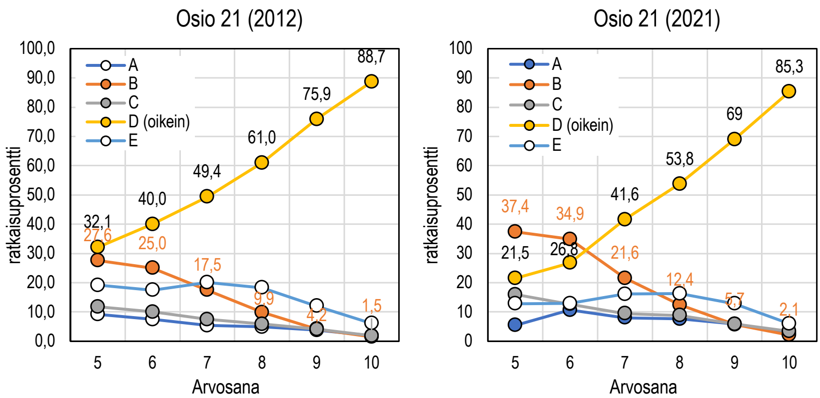
KUVIO 4.27. Tehtävässä 14 suoriutumisen muuttuminen vuosien 2015 paperi-kynä-testin ja 2021 digitaalisen testin välillä

Osiossa 15 tuli laskea oikokulman kulmia, kun kaksi niistä on identtisiä mutta tuntemattomia ja tunnettu kulma on 50 astetta (Kuvio 28). Kokonaisuutena ajatellen oikea ratkaisu löytyy vuonna 2021 kaikissa arvosanaluokissa lähes identtisellä todennäköisyydellä kuin vuonna 2015. Mekaanisen laskuvirheen osuus näyttää hieman *laskeneen* alimmissa arvosanluokissa (vaihtoehto D), mutta vastaavasti oikokulma on saattanut hieman aiempaa useammin sekoittua suoraan kulmaan. Kaikkiaan tehtävän ratkaisussa ei ole juuri tapahtunut muutosta mittauskertojen välillä.



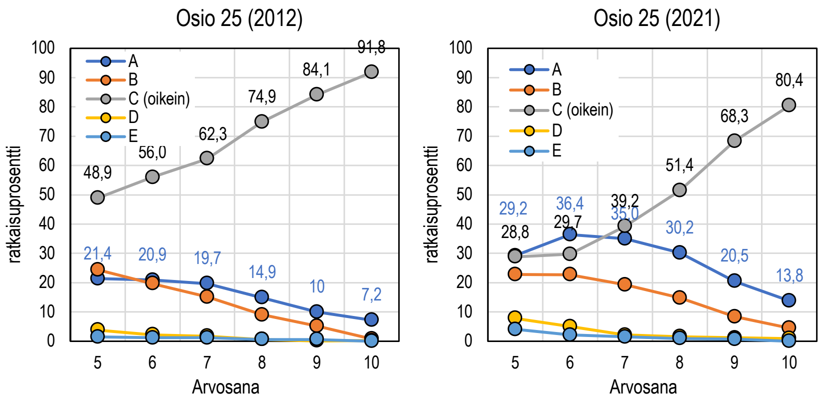
KUVIO 4.28. Tehtävässä 15 suoriutumisen muuttuminen vuosien 2015 paperi-kynä-testin ja 2021 digitaalisen testin välillä

Osiossa 21 tuli tunnistaa pyramidiin liittyviä käsitteitä ja laskea sen ominaisuuksia (Kuvio 29). Kaikissa arvosanaluokissa oikean ratkaisun löytymisestä on tullut hieman epätodennäköisempää, mutta erityisesti matalimmissa ryhmissä sekoitetaan pyramidin korkeus sivun pituuteen (vaihtoehto B). Sivun pituuteen liittyvä vaihtoehto B (7 cm) näkyy suoraan tehtävässä ja sen perusteella olisi pitänyt laskea korkeus, mutta aiempaa useampi arvosanaluokissa 5–7 ollut oppilas otti sivun pituuden korkeutena.



KUVIO 4.29. Tehtävässä 21 suoriutumisen muuttuminen vuosien 2012 paperi-kynä-testin ja 2021 digitaalisen testin välillä

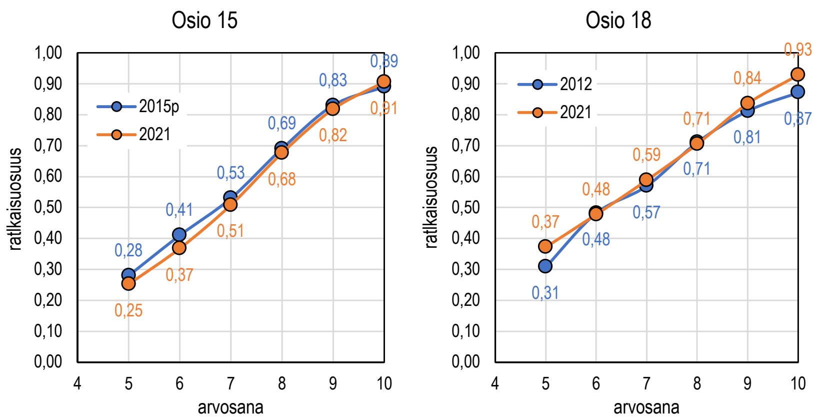
Viimeiseksi monivalintatehtävistä tehtävässä 25 tuli laskea tai arvioida, kuinka monta kertaa pienempi kolmio sisältyy suurempaan neliöön (Kuvio 30). Ratkaisun voi laskea vertaamalla kolmion ja neliön pinta-aloja tai karkeasti piirtämällä kolmioita neliön sisään ja arvioimalla määrän. Huomattavaa on, kaikissa arvosanaluokissa oikean vastauksen löytäminen on selvästi epätodennäköisempää kuin aiemmassa mittauksessa 2012, joskin ero on selkeintä erityisesti arvosanan 6 saaneilla oppilailla. Useimmiten valituksi vaihtoehdoksi on tullut vaihtoehto ”4 kertaa” kun oikea vastaus on ”8 kertaa” eli selkeästi suurempi luku. Koska ongelma esiintyy myös korkeimmissa arvosanaluokissa, kyseessä ei varmaankaan ole itse laskemiseen vaan välineeseen liittyvä haaste. Paperi-kynä-kokeissa oppilaat saattoivat helposti tarkistaa tuloksen mielekkyyden tai kolmioiden määrän hahmottelemalla niitä neliön sisään. Nykyinen digitaalinen alusta saattaa olla liian kankea tämän tyyppiseen karkeaan hahmotteluun. On mahdollista, että kun tekniikka ajan myös kehittyy ja mahdollisia piirto-ohjelmia lisätään testialustoille ja niitä opitaan käyttämään, tehtävän ratkaisun onnistuminen saattaa palata aiemmalle tasolle.



KUVIO 4.30. Tehtävässä 25 suoriutumisen muuttuminen vuosien 2011 paperi-kynä-testin ja 2021 digitaalisen testin välillä

Ratkaisuprosenteissa tapahtuneita muutoksia

Tarkentavaa analyysia varten käytettävissä oli 8 tehtävää, joilla vuoden 2021 julkaistava tehtäväsarja linkitettiin aiempien vuosien tehtäväsarjoihin. Tämän pienen otoksen perusteella näyttää siltä, että kansallisen osaamisen heikkenemisestä voi tehdä kolme keskeistä havaintoa. Ensinnäkin osassa tehtävistä osaamisen lasku on erittäin pientä, ellei olematonta kaikissa arvosanaluokissa. Tämän kaltaisia ovat tehtävät 15 ja 18 (Kuvio 31). Näistä laskujärjestystehtävässä (tehtävä 18) osaaminen on osittain tullut jopa hieman paremmaksi vuoteen 2012 nähden, ja kulmien laskemisen osalta (tehtävä 15) muutos on marginaalista alimmissa arvosanaluokissa.

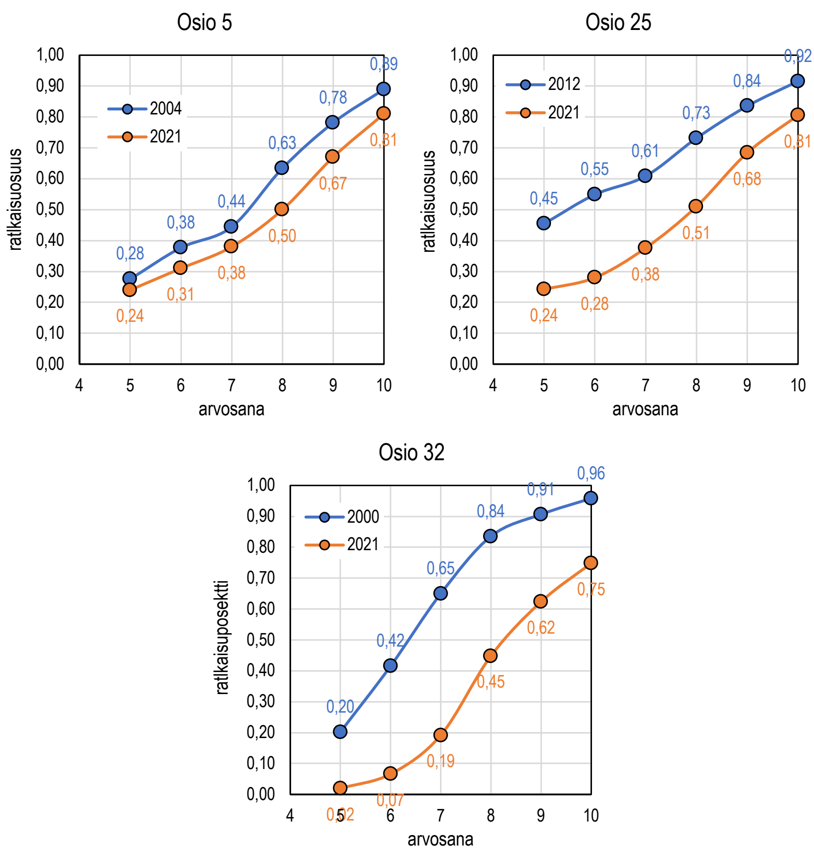


KUVIO 4.31. Suoriutumisen muuttuminen aiempaan nähden tehtävissä 15 ja 18

Näytteen perusteella ei tietenkään voida sanoa, yhdistääkö tämänkaltaisia tehtäviä jokin sisällöllinen elementti. Huomioitavaa on, että tutkituista tehtävistä tehtävä 18 on ainoa, joka ainoana puhtaasti edustaa sisältöaluetta luvut ja laskutoimitukset (S2). Tehtävä 15 myös voisi olla tässä sisältöalueessa, sillä geometrian osalta tässä ei oppilaan tarvitse tietää muuta, kuin kuinka paljon oikokulma on asteina. Tämän jälkeen tehtävä on varsin suoraviivainen päässälasku.

Toiseksi osassa tehtävistä osaaminen on laskenut melko tasaisesti *kaikissa* arvosanaluokissa. Tämän kaltaisia ovat tehtävät 5, 25 ja 32 (Kuvio 32). Näistä ainakin kahta yhdistää välineestä eli digitaalisesta testauksesta johtuvat tekijät.

Tehtävä 5 oli aikarajoitettu, komplisoitu päässälaskutehtävä, jossa piti ratkaista kävelyn nopeuden perusteella polkupyörällä koulumatkaan käytetty aika. Vuoden 2004 paperi-kynä-testissä tässä osuudessa testausta paperia ja kynää saattoi käyttää kuultujen päässälaskutehtävien apuna. Myös vuoden 2021 digitaalisessa testauksessa paperia ja kynää sai käyttää apuna tässä osuudessa. On mahdollista, että digitaalisessa testauksessa hahmotelmapapereita on ehkä alettu käyttää myöhemmin, jos lainkaan. Digitaalisen alustan käyttö ei siis ehkä rohkaissut muiden rinnakkaisten apuvälineiden käyttöön. Tämä näkyy selkeimmin tehtävässä 25.



KUVIO 4.32. Suoriutumisen muuttuminen aiempaan nähden tehtävissä 5, 25 ja 32

Tehtävässä 25 piti laskea neliön sisään mahtuvien kolmioiden määrä. Jopa parhaat oppilaat arvioivat kolmioiden määrän *puolet* pienemmäksi kuin mikä oikea vastaus oli. Huomattiin, että vastuksessa oli todennäköinen väline-efekti: tehtävä oli tyypiltään sellainen, että paperi-kynä-testaus helpotti vastauksen järkevyyden tarkistamista, ellei peräti mahdollistanut koko tehtävän ratkaisemisen pelkästään piirtämällä. Digitaalinen alusta ei ollut tähän hyvä työkalu.

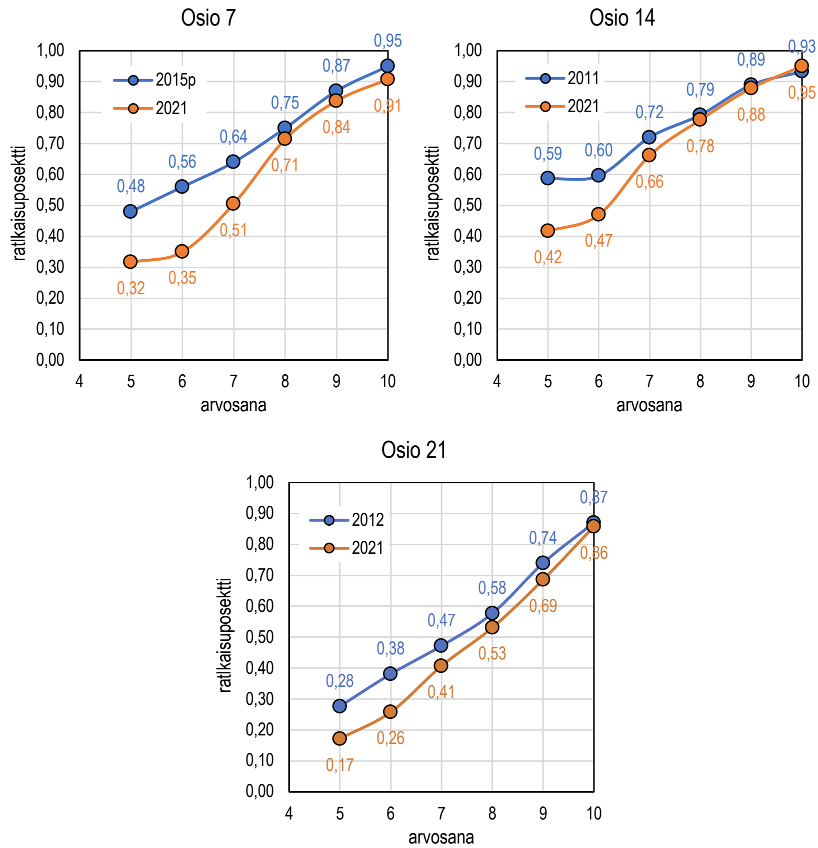
Tehtävässä 32 piti ratkaista yhtälö ja perustella ratkaisu eli kirjoittaa laskusuoritus matematiikkaeditorilla järjestelmään. Toisaalla tässä raportissa (Metsämuuronen & Nousiainen, 2023, luku 2) pohditaan tehtävää 32 tarkemmin. Tämä oli yksi niistä tehtävistä, joita ei voitu käyttää linkkitehtävänä, koska osiossa vastaamisen logiikka oli muuttunut. Tehtävästä saattoi saada 2 pistettä: yhden oikeasta vastauksesta ja toisen perustelusta eli oikeasta laskusuorituksesta. Kun vuonna 2000 paperi-kynä-testissä 66 prosenttia oppilaista antoi ratkaisulleen perustelun, vuoden 2021 aineistossa vain 29 prosenttia antoi perustelun. Lähes 40 prosenttiyksikön lasku perustelujen antamisessa ei voi johtua pelkästään osaamisen muutoksesta. Opettajat kommentoivatkin asiaa omassa palautteessaan. Oppilaat olisivat varmasti osanneet antaa perustelun, mutta eivät jostain syystä halunneet kirjoittaa sitä kankeilla matematiikkaeditoreilla. Tämän tyyppisiin mekaanisiin haasteisiin on syytä kiinnittää huomiota tulevissa arvioinneissa.

Kolmas osaamisen heikkenemiseen liittyvä seikka havainnollistuu tehtävissä 7, 14 ja 21 (Kuvio 33) sekä mahdollisesti myös edellä käsitellyssä tehtävässä 25. Näille yhteistä on se, että osaaminen on heikentynyt erityisesti alimmissa osaamisluokissa—selkeästi arvosanoja 5 tai 6 saaneilla, mutta osittain myös arvosanan 7 saaneilla.

Omenan hinnan määräytymistä funktion perusteella mittaavaa tehtävää 7 käsiteltiin jo edellä monivalintatehtävien yhteydessä. Huomattiin, että erityisesti alimmissa osaamisluokissa oppilailla näyttää olleen taipumusta pyöristää luku 3,5 liian aikaisin luvuksi 4 ja ratkaista lasku tämän perusteella. Tämä oli yleisempää arvosanoja 5, 6 tai 7 saaneilla oppilailla.

Todennäköisyystehtävässä 14 erityisesti arvosanan 5 tai 6 saaneet oppilaat valitsivat aiempaa useammin vaihtoehdon, jossa ei luettu tarkasti tehtävänantoa, vaan poimittiin tekstistä yksittäinen sana ”kolme” kun olisi pitänyt ymmärtää poimia sanat ”kolmella jaollinen”, mikä johti tietyn tyyppiseen virheratkaisuun (ks. myös edellä kuvio 8 ja siihen liittyvä teksti virhemekanismista). Tämä tehtävä oli haasteellinen erityisesti heikoimpia arvosanoja saaneille maahanmuuttotaustaisille oppilaille, joita sanallisesti pitkähkö tehtävänanto ja termi "jaollinen" ovat saattaneet hämätä.

Tehtävässä 21, jota myös käsiteltiin edellä monivalintatehtävien yhteydessä, piti tunnistaa pyramidiin liittyviä termejä ja laskea sen ominaisuuksia. Tehtävä oli työläs, sillä sen yhteydessä piti laskea neljä tai viisi laskuoperaatiota, koska vaihtoehdot oli ilmaistu erilaisten laskutoimitusten tuloksina. Heikoilla oppilailla oli aiempaa useammin taipumuksena ottaa tehtävänannossa mainittu luku ja luulla tätä oikeaksi vastaukseksi. Kyse voi olla myös matemaattisten tehtävien kielellistämisestä, jolloin myös kielitaitoon liittyvät puutteet korostuvat. Osiossa 25 (ks. edellä kuvio 32) voi myös havaita, että sen lisäksi että osaaminen on heikentynyt kaikissa arvosanaluokissa selvästi, alimmissa arvosanaluokissa osaaminen on laskenut enemmän kuin muissa luokissa. Edellä monivalintatehtävien yhteydessä havaittiin, että erityisesti arvosanan 6 saaneilla oppilailla on aiempaa useammin taipumusta valita tietty virhevaihtoehto, mikä saattoi johtua digitaalisesta testiympäristöstä.



KUVIO 4.33. Suoriutumisen muuttuminen aiempaan nähden tehtävissä 7, 14 ja 21

Sisältöalueiden eroista ei yksittäisten tehtävien perusteella voida vetää pitkälle meneviä johtopäätöksiä. Tässä käsitellyissä geometrian tehtävissä (15, 21, 25) ja tehtävissä 21 ja 25 osaaminen on kuitenkin selvästi aiempaa heikompaa erityisesti arvosanoja 5–7 saaneilla oppilailla. Näissä tehtävien ratkaisemisen oleellisena osana on hyvä luetun ymmärtäminen. Myös julkaisemattomissa linkkitehtävissä havaittiin samanlaista ilmiötä. Pitkien sanallisten tehtävien osaaminen oli hyvin heikkoa, vaikka lähtökohtaisesti tehtävän ratkaisemiseen olisivat riittäneet peruslaskutoimitukset. Oppilaiden kyky ratkaista tehtäviä ei välttämättä olekaan heikentynyt, vaan syynä näyttää olevan ainakin osittain heikko taito ymmärtää sanallisia tehtävänantoja.

Ymmärrettävästi edellä kuvatut osiot ovat vain näytteitä niistä mekanismeista, joilla viimeaikainen kansallinen osaamisen taso on ollut laskujohteinen. Yhteenvetona voidaan todeta, että osa tason heikentymisestä voidaan suoraan liittää välineisiin eli digitaalisen testaamisen alkukankeuteen ja uudessa ympäristössä toimimisen lyhyeen historiaan. On tietenkin oleellista, että oppilaille annetaan riittävästi aikaa tutustua digitaaliseen testiympäristöön ennen varsinaista testausta. Toisaalta näyttää siltä, että osaaminen on aidosti heikentynyt erityisesti osaamisen heikommassa ääripäässä. Aiempaa useammin heikot oppilaat eivät tunnista termejä tai sanastoa ja pidempien tekstien lukeminen ja ymmärtäminen saattaa olla haasteellista. Huomataan lisäksi, että kansallinen osaamisen tason muutos on moniulotteinen ilmiö, johon vaikuttavat monet tekijät lähtien testeissä vastaamisen muuttuneista tekniikoista yhteiskunnassa tapahtuviin radikaaleihin muutoksiin esimerkiksi ajankäytössä.

4.5 Pohdintaa ja suosituksia

4.5.1 Keskeisiä osaamisen muutokseen huomioita

Lähtökohtina tässä artikkelissa tehdyille analyyseille oli kolme huomiota matematiikan osaamisen muutoksista kansallisissa aineistoissa. Ensiksi kansallinen osaamisen taso on ollut laskujohteinen ainakin vuodesta 2000 lähtien, toiseksi helpoissa tehtävissä erittäin hyvin suoriutuneiden oppilaiden osuus on vähentynyt systemaattisesti vuosien varrella ja kolmanneksi viimeisimmässä matematiikan oppimistulosarvioinnissa osaamisen jakaumissa on selkeä leventynyt muoto, mikä viittaa siihen, että ääripäissä osaaminen on eriytynyt toisistaan. Analyysissa mallitettiin nämä populaatiot ja pyrittiin löytämään tekijöitä, joiden perusteella heikoimpaan osaamisryhmään kuuluvia oppilaita voitaisiin paikallistaa ja tarjota aiempaa tehostetummin tukea.

Vaikka matematiikan osaamisen laskun alkua voidaan suomalaisissa armeija-aineistoissa jäljittää 1990-luvun alkupuolelle, ja vaikka tiedetään että vuonna 2010 osaamisen taso vastasi vuoden 1985 tasoa, on pidettävä mielessä, että faktisesti nykylapset saattavat tietää oleellisesti enemmän monista asioista kuin 1980-luvulla eläneet nuoret puhumattakaan 1960-luvulla nuoruuttaan eläneistä oppilaista. Jos saisimme aikakoneella 15-vuotiaan vuodesta 1962 ja vuodesta 1982 ja laittaisimme nämä keskustelemaan 15-vuotiaan kanssa vuonna 2021, saattaisimme huomata, että nuorten ymmärrys ja tietämys maailmasta ja maailman politiikasta, eriarvoisuudesta, luonnosta, luonnonlaeista, elokuvista, mainonnasta, DNA:sta, avaruusmatkailusta puhumattakaan tietokoneista ja digitaalisesta maailmasta on kyllä kasvanut selvästi. Samoin kielitaito on monilta osin lisääntynyt oleellisesti, vaikka tietenkin jotain on myös menetetty. Tämän tyyppiseen positiiviseen muutokseen viittaa ns. Flynnin ilmiö: älykkyystesteillä mitatulla ”älykkyydellä” on ollut taipumusta lisääntyä populaatiossa ajan myötä. Toisaalta tutkijat ovat huomauttaneet käänteisestä Flynnin ilmiöstä: mitattu osaaminen näyttää populaatiotasolla laskevan, kuten on käynyt kansallisessa aineistossamme. Keskeinen kysymys on, onko tämä ”menetetty” kognitiivisia kykyjä vai jotain muuta? Onko ”perustaitoja” menetetty—vai ovatko taidot vaihtuneet sellaisiksi, joita ei ole ollut (vielä) tarpeen tai mahdollista mitata? Osa uusista taidoista saattaa olla niin itsestään selvyyksiä nykyisessä maailmassa, että niihin ei ole ollut tarpeenkaan kiinnittää huomiota, kuten esimerkiksi tietotekninen osaaminen ja tähän linkittyvät taidot.

Vainikainen ja Hautamäki (2022) esittävät oppimaan oppimisen aineistojen perusteella, että Suomessa osaamisen laskussa ei ehkä kuitenkaan ole kyse aidosta anti-Flynn ilmiöstä eli älykkyydessä tapahtuvasta muutoksesta, vaan pikemminkin motivaation puutteesta ja vähentyneestä halusta panostaa ns. *low-stake* kokeissa ja koulutyössä. Kyse voisi siis olla ennemmin jonkinlaisesta motivaation puutteen seurauksena syntyneestä osaamisen heikkenemisestä kuin aidosti kognitiivisessa kyvykkyydessä tapahtuneista muutoksista. Tämä johtaa ajatukset kahteen suuntaan.

Yhtäältä voi olla järkevää pohtia, kuinka voisimme motivoida oppilaita tekemään kansallisia ja kansainvälisiä *low-stake* testejä vakavasti ja parasta osaamistaan näyttäen. Toisaalta mahdolliseen motivaation puutteeseen saattavat johtaa monenlaiset käytännöt koulussa ja kotona alkaen mahdollisesta vaatimustason laskemisesta ulkoa opettelun väheksymiseen. Heikoimmilla oppilailla saattaa aidosti olla sellainen harhakäsitys, että juuri mitään asioita ei tarvitse opetella ulkoa, koska kaikki tieto löytyy internetistä. Ulkoa opetteleminen kuitenkin automatisoi matemaattisten operaatioiden hallintaa ja nopeuttaa prosesseja. Haasteellista onkin tietää, mitkä olisivat ne asiat, jotka arkisessa koulutyössä on järkevää edellyttää heikonkin oppilaan ulkoa opittaviksi (kuten esimerkiksi kertotaulu) ja mitkä olisivat ne tekijät, operaatiot ja käsitteet, joiden ulkoa opettelemisen ei toisi suurtakaan lisäarvoa oppilaille. Parhaat oppilaat tietenkin ymmärtävät, että *kaikki* koulukontekstissa esiin tulevat käsitteet, kaavat ja operaatiot on järkevää hallita, jotta menestyisi myöhemmissä opinnoissa.

Joitain systemaattisia huomioita tehtiin julkaistavien tehtävien perusteella. Teknisissä ja puhtaissa laskutoimituksissa ei näkynyt huomattavia eroja heikompien ja parempien oppilaiden osaamisessa. Keskeinen ero heikompien ja parempien oppilaiden välillä näyttää tulevan erityisesti sanallissa tehtävissä. Kun tehtävään sisältyy laskemisen lisäksi sanallinen ongelma, niin varsinkin arvosanoja 5–7 saaneet oppilaat osaavat näitä heikommin. Epäselvää on, onko syynä oppilaiden haluttomuus keskittyä ratkaisemaan hankalampia tehtäviä vai kyky ymmärtää niitä.

Osassa algebran sisältöalueen tehtävistä osaamisen lasku voi yhtäältä johtua käytetystä välineestä, sillä oppilaille näytti olevan taitotasosta riippumatta haasteellista tai vaivalloista esimerkiksi kirjoittaa laskutoimituksia testauksessa käytetyllä matematiikkaeditorilla. Toisaalta julkaistujen helppojen tehtävien tekeminen on pääosin mahdollista kuudesluokkalaiselle oppilaalle. Niinpä suurin osa tehtävistä ei välttämättä edellytä uutta teoreettista ainesta näiden ratkaisemiseen, vaan ainoastaan muutamien termien selittämistä ja aiemmin opittujen taitojen soveltamista. Analyysissä käytetyt tehtävät olivat verrattain helppoja perustehtäviä, ja näyttääkin siltä, että varsinkin heikoilla oppilailla oli välineen käytön tuoman haasteen lisäksi ongelmia myös perusyhtälön ratkaisussa.

4.5.2 Suosituksia

Heikosti menestyville oppilaille on syytä kohdentaa aiempaa intensiivisempää tukea mahdollisimman varhaisessa vaiheessa, jotta voidaan estää mahdollinen matemaattinen syrjäytyminen. Aiemmassa pitkittäistutkimuksessa (Metsämuuronen, 2017) havaittiin, että muutos matematiikan osaamisessa on suurempaa alakoulussa kuin yläkoulussa. Voidaankin arvioida, että oppilas, joka jää jälkeen matematiikan oppimisessa luokkatasoilla 3–6, vaatii huomattavan paljon tukea yläkoulun aikana, jotta hän pystyisi kirimään muodostuneen eron kiinni. Onkin siis järkevää suunnata tukiresursseja alkuopetukseen ja tämän jälkeiseen aikaan riittävän varhain ja riittävän tehokkaasti, ennen kuin osaamisessa ilmeneviä puutteita on vaikea korjata ylempien luokkien aikana.

On järkevää tietoisesti ja systemaattisesti kartoittaa niitä oppilaita, jotka potentiaalisesti hyötyisivät aiempaa intensiivisemmästä tuesta, vaikka he eivät sitä aktiivisesti ehkä kysykään. Näitä ryhmiä saattaisivat olla esimerkiksi arvosanan 5 tai 6 saaneet oppilaat, maahanmuuttotaustaiset oppilaat, heikommin menestyneet tytöt, erityisesti maahanmuuttotaustaiset tytöt (ks. Metsämuuronen

& Nousiainen, 2021), ja ne oppilaat, jotka eivät ehkä viihdy koulussa. Aiemmasta koulun alkuvaiheen mittauksesta (Ukkola ym., 2020) tiedetään, että keskimääräistä heikompaa koulutulokkaan lähtötasoa ennustivat jo varhain tehty tehostetun tai erityisen tuen päätös, suomi toisena kielenä -status eli käytännössä maahanmuuttotausta, lähisuvussa esiintyneet oppimisvaikeudet, sekä jossain määrin myös se, että lapsi on syntynyt loppuvuodesta ja huoltajien matala koulutustausta. Näihin oppilasryhmiin voi olla järkevää kiinnittää huomiota myös alkuopetuksen jälkeen.

Osalla heikosti menestyvistä oppilaista saattaa myös olla numerotaidottomuutta eli dyskalkuliaa. On järkevää varhaisessa vaiheessa testata heikosti menestyviä oppilaita tämän varalta, jotta riittävät tukitoimet olisivat oikea-aikaisia.

Näyttää siltä, että oppilaiden osaaminen erityisesti matematiikan erityissanastossa, ongelmanratkaisutehtävissä ja sanallistetuissa tehtävissä on oleellisesti heikentynyt vuosien varrella. On siis järkevää kiinnittää huomiota yhtäältä oppilaiden sanavaraston kehittymiseen, tekstien ymmärtämisen ja ajattelun taitoihin ja toisaalta taitoon työskennellä pitkäjänteisesti ongelmaratkaisutehtävien parissa. Näiden taitojen kehittäminen jo alakoulusta alkaen helpottaisi monen oppilaan kykyä ratkaista tehtäviä, joissa pitää soveltaa jo opittuja tietoja.

4.6 Lähteet

Altinok, N., Diebolt, C., & Demeulemeester, J.-L. (2014). A new international database on education quality: 1965–2010. *Applied Economics, 46*(11), 1212–1247. 10.1080/00036846.2013.868592

Altinok, N., Angrist, N., Patrinos, H. A. (2018). *Global data set on education quality (1965–2015)*. Policy Research Working Paper; No. 8314. World Bank. http://hdl.handle.net/10986/29281

Flynn, J. R. (1984). The mean of IQ of Americans: Massive gains 1932 to 1978. *Psychological Bulletin, 95*(1), 29–51. https://doi.org/10.1037/0033-2909.95.1.29

Flynn, J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin, 101*, 171–191. https://doi.org/10.1037/0033-2909.101.2.171

Flynn, J. R. (2009a). *What is intelligence? Beyond the Flynn effect*. 2. laitos. Cambridge University Press.

Flynn, J. R. (2009b). Requiem for nutrition as the cause of IQ gains: Raven's gains in Britain 1938–2008. *Economics and Human Biology,* 7(1), 18–27. https://doi.org/10.1016/j.ehb.2009.01.009

Flynn, J. R., & Shayer, M. (2018). IQ decline and Piaget: Does the rot start at the top? *Intelligence, 66*, 112–121. https://doi.org/10.1016/j.intell.2017.11.010

Hautamäki, J., Kupiainen, S., Marjanen, J., Vainikainen, M-P., & Hotulainen, R. (2013). *Oppimaan oppiminen peruskoulun päättövaiheessa: Tilanne vuonna 2012 ja muutos vuodesta 2001*. Tutkimuksia 347. Helsingin yliopisto, opettajankoulutuslaitos.

Kalenius, A. (2023). *Sivistyskatsaus 2023*. Opetus- ja kulttuuriministeriön julkaisuja 2023:3. https://julkaisut.valtioneuvosto.fi/handle/10024/164564

Metsämuuronen, J. (2013). A new method to setting standard for the wide range of language proficiency levels. *International Education Research*, 1(1), 1–21. https://doi.org/10.12735/IER.V1I1P01.

Metsämuuronen, J. (2018). Common framework for mathematics—Discussions of possibilities to develop a set of general standards for assessing proficiency in mathematics. *International Electronic Journal of Mathematics Education*, *13*(2), 13–39. https://doi.org/10.12973/iejme/2693

Metsämuuronen, J. (2019). Educational assessment and some related indicators of educational equality and equity. *Education Quarterly Reviews*, 2(4), 770–788. https://doi.org/10.31014/aior.1993.02.04.105.

Metsämuuronen, J. (2022). Directional nature of the product-moment correlation coefficient and some consequences. *Frontiers in Psychology*, *13*:988660. https://doi.org/10.3389/fpsyg.2022.988660

Metsämuuronen, J. & Nousiainen, S. (2021). *Matematiikkaa COVID-19-pandemian varjossa. Matematiikan osaaminen 9. luokan lopussa keväällä 2021*. Julkaisut 27:2021. Kansallinen koulutuksen arviointikeskus. https://karvi.fi/wp-content/uploads/2021/12/KARVI_2721.pdf

Metsämuuronen, J. & Nousiainen, S. (2023).Yleiset menetelmäratkaisut matematiikan oppimistulosten arvioinnissa vuonna 2021. Teoksessa J. Metsämuuronen & S. Nousiainen (toim.), *Matematiikkaa COVID-19-pandemian varjossa II. Menetelmälliset ratkaisut matematiikan 9. luokan arvioinnissa keväällä 2021* (ss. 21–82). Julkaisut 5:2023. Kansallinen koulutuksen arviointikeskus.

Metsämuuronen, J. & Ukkola, A. (2022). Rudimentary stages of the mathematical thinking and proficiency. Mathematical skills of low-performing pupils at the beginning of the first grade. *LUMAT: International Journal on Math, Science and Technology Education, 10*(2), 56–83. https://doi.org/10.31129/10.1.1632

OECD (2019). *Country note*. OECD Publications.

Suomi: https://www.oecd.org/pisa/publications/PISA2018_CN_FIN.pdf;

Ruotsi: https://www.oecd.org/pisa/publications/PISA2018_CN_SWE.pdf;

Norja: https://www.oecd.org/pisa/ publications/PISA2018_CN_NOR.pdf;

Tanska: https://www.oecd.org/pisa/publications/ PISA2018_CN_DNK.pdf;

Viro: https://www.oecd.org/pisa/publications/PISA2018_CN_EST.pdf.

Shayer, M., & Ginsburg, D. (2009). Thirty years on—A large anti- Flynn effect? (II): 13- and 14-year-olds. Piagetian tests of formal operations norms 1976–2006/7. *British Journal of Educational Psychology, 79*(3), 409–418. https://doi.org/10.1348/978185408x383123

Shayer, M., Ginsburg, D., & Coe, R. (2007). Thirty years on—a large anti-Flynn effect? The Piagetian test Volume & Heaviness norms 1975–2003. *British Journal of Educational Psychology,* 77(1), 25–41. https://doi.org/10.1348/000709906x96987

Ukkola, A., Metsämuuronen, J. & Paananen, M. (2020). *Alkumittauksen syventäviä kysymyksiä*. Julkaisut 10:2020. Kansallinen koulutuksen arviointikeskus. https://karvi.fi/wp-content/uploads/2020/08/KARVI_Alkumittaus.pdf

Vainikainen, M.-P. & Hautamäki, J. (2022). Three Studies on Learning to Learn in Finland: Anti-Flynn Effects 2001–2017. *Scandinavian Journal of Educational Research, 66*(1), 43–58 . https://doi.org/10.1080/00313831.2020.1833240

Van der Schoot, F. (2009). Cito variation of the Bookmark method. Reference supplement, Section I. Teoksessa S. Takala (toim.), *Relating language examinations to the Common European Framework of Reference for Languages: Learning, teaching, assessment (CEFR). A Manual*. Language Policy Division, Strasbourg. https://rm.coe.int/1680667a24

VATT (2018). *Economic Policy Council Report 2017*. Economic Policy Council. VATT. https://www.talouspolitiikanarviointineuvosto.fi/wordpress/wp-content/uploads/2018/01/Report2017.pdf

Woodley, M. A., & Meisenberg, G. (2013). In the Netherlands the anti-Flynn effect is a Jensen effect. *Personality and Individual Differences, 54*(8), 871–876. http://dx.doi.org/10.1016/j.paid.2012.12.02

4.7 Liitteet

LIITE 4.1. Kolmen populaation erottelussa käytettyjä muuttujia

	alapopulaatio	keskipopulaatio	yläpopulaatio	
otoskoko korkeimmillaan	**n_1 = 1488**	**n_2 = 2017**	**n_3 =1812**	
tehostettu tuki + erityinen tuki	tehostettu tai erityinen tuki (75,1–83,4 %)	yleinen tuki (41,8 %)	yleinen tuki (39,8 %)	n
yleinen tuki	18,1	41,9	40,0	4351
tehostettu tuki	75,2	19,9	4,8	557
erityinen tuki	84,3	11,8	3,9	280
mat arvosana	4,5,6 (82,9 %)	7,8 (68,2 %)	9, 10 (72,9 %)	
mat arvosana 4–6	82,9	26,0	1,0	1235
mat arvosana 7 tai 8	16,4	68,2	26,1	2130
mat arvosana 9 tai 10	0,8	5,8	72,9	1895
Sukupuoli				
Tyttö	27,2	41,2	31,6	2590
Poika	28,6	34,9	36,5	2723
K S2C status	S2 status (59 %)	ei S2 statusta (39,0 %)	ei S2 statusta (36 %)	
S2-status	59,0	26,5	14,5	442
Ei S2-statusta	25,0	39,0	36,0	4834
T4BF Minkä ikäisenä aloitti koulun käynnin suomessa	>7 (61 %)	5–7 (39%)	5–7 (35	
5–7-vuotiaana	16,3	38,7	35,0	5087
yli 7 vuotiaana	61,3	21,5	17,3	191
T7. erikois- tai erityisluokka	joustava perusopetus, pienryhmä, erityisluokka (87,0 %)	ei erikoisluokalla tai liikunta-, ilmaisu-, viestintä-, tai taideluokka (39,3–44,0 %)	stem-, mat-, science-, teknologia-, musiikki- tai kieliluokka (52,9–71,3 %)	
T7DC0 ei erityisluokalla	27,6	39,3	33	4288
T7DC1 JOPO, PIENR., ERIT.	87,0	11,6	1,4	207
T7DC2 LIIK ILMAISU TAIDE	16,2	44,0	39,8	259
T7DC3 MUS KIELI	11,6	35,5	52,9	242
T7DC4 MAT STEM LUMA YRITT	22,6	6,1	71,3	115

	alapopulaatio	keskipopulaatio	yläpopulaatio	
otoskoko korkeimmillaan	**n_1 = 1488**	**n_2 = 2017**	**n_3 =1812**	
T8 Mitä aiot tehdä peruskoulun jälkeen	ammatillinen koulutus, Töihin; Lisävalmiudet (55,0–70,2 %)	lukio, lyhyt matematiikka (65,3 %)	lukio, pitkä matematiikka (68,5 %)	
T8DLPitkä. Lukio, Pitkä mat.	3,6	27,9	68,5	2176
T8DLLyhyt. Lukio, Lyhyt mat.	24,5	65,0	10,5	894
T8DAmm. Ammatillinen	55,0	7,3	7,7	1750
T8DLisäTyöVäli. Muut	70,2	21,0	8,9	124
T1011B SES äidin ja isän korkein koulutus yhdessä	peruskoulu, ammatillinen, en tiedä (36,8–37,8 %)	peruskoulu, ammatillinen (42,5–43,6)	yliopisto (57,6)	
T1011B0 ei tiedä kumpaakaan ja puuttuva	37,8	37,2	25	1223
T1011B1 molemmilla peruskoulu tai ammatillinen ja muu	36,8	43,6	19,5	779
T1011B2 toisella peruskoulu tai ammatillinen ja muu ja toisella tätä ylempi	29,3	42,5	28,2	964
T1011B3 molemmilla lukio tai/ja ammattikorkeakoulu (tai toisella yliopisto)	23,7	38,1	38,1	1407
T1011B4 molemmilla yliopistokoulutus	13,1	29,2	57,6	944
T12 SES SUM minulla on… (max 5)	0–1 (58,3 %)	2–5 (36,4–39,4 %)	4–5 (37,4 %)	
T12D 0–1 viidestä vaihtoehdosta	58,3	25,4	16,3	331
T12D 2–3 viidestä vaihtoehdosta	37,8	36,4	25,8	889
T12D 4–5 viidestä vaihtoehdosta	22,2	39,4	38,4	3663
T12.1 pöytä opiskelua varten (ei ole)	52,1	31,5	16,4	499
T12.2 oma huone (ei ole)	42,6	25,9	31,5	359
T12.3 rauhallinen paikka opiskelulle (ei ole)	45,9	30,9	23,2	825
T12.4 tietokone, jolla voit tehdä koulutehtäviä (ei ole)	45,9	33,8	20,4	737
T12.5 tukevia tietokoneohjelmistoja (ei ole)	32,3	37,4	30,3	2708

	alapopulaatio	keskipopulaatio	yläpopulaatio	
otoskoko korkeimmillaan	**n_1 = 1488**	**n_2 = 2017**	**n_3 =1812**	
T13 SES kotonani on SUM (max 5)	0–1 (53,7%)	2–3 (41,6%)	4–5 (50,0 %)	
TD13C1. 0–1 viidestä	53,7	33	13,3	1224
TD13C2. 2–3 viidestä	24,8	41,6	33,6	1855
TD13C3. 4–5 viidestä	12,6	37,4	50	1804
T13.1 E-kirjojen lukulaitteita esim. Kindle	29,3	38,3	32,5	4017
T13.2 kulttuuri esineitä esim. taulut, kaunokirjallisuus (ei ole)	42,2	37,8	20	2060
T13.3 musiikki-instrumentteja esim. kitara, piano (ei ole)	40,6	37,8	21,6	2117
T13.4 kaikilla lapsilla oma huone (ei ole)	36,9	34,2	28,9	1136
T13.5 minulle tarkoitettuja kirjoja muita kuin koulukirjoja (ei ole)	47,4	35,3	17,3	1755
T14 miten viihdyt koulussa	erittäin huonosti (53,1 %)	melko hyvin (40,9 %)	erittäin hyvin (45,6 %)	
T14B1 erittäin huonosti	53,1	34,4	12,5	224
T14B2 melko huonosti	34,4	38,2	27,4	620
T14B3 melko hyvin	24	40,9	35	3052
T14B4 erittäin hyvin	26,9	27,6	45,6	834
T15 Poissaoloja	>20 pv (47,0 %)	6–20 pv. (41,6–43,4 %)	0–5 pv. (41,5 %)	
T15 poissaoloja 0–5 pv	22	36,5	41,5	2260
T15 poissaoloja 6–10 pv	22,5	43,4	34,1	1132
T15 poissaoloja 11–20 pv	29,7	41,6	28,7	697
T15 poissaoloja yli 20 pv	47	32,3	20,3	561
T18A Positiivinen Tunnetila	ei koskaan tai hyvin harvoin positiivinen (42–67 %)	joskus tai harvoin positiivinen tunne (46–49 %)	usein tai lähes aina positiivinen tunne (60–72 %)	
T18A1 ei koskaan	67,4	27,5	5,1	316
T18A2 harvoin	42,1	45,8	12,1	779
T18A3 joskus	24,9	48,7	26,4	1719
T18A4 usein	10,9	29,7	59,7	1346
T18A5 lähes aina	9,4	18,8	71,9	288

	alapopulaatio	**keskipopulaatio**	**yläpopulaatio**	
otoskoko korkeimmillaan	**n_1 = 1488**	**n_2 = 2017**	**n_3 =1812**	
T18 Negatiivinen Tunnetila	lähes aina negatiivinen tunne (55 %)	Joskus tai usein negatiivinen tunne (46−49 %)	ei lainkaan tai harvoin negatiivinen tunne (44−56 %)	
T18B1 ei koskaan	20	24,3	55,7	805
T18B2 harvoin	18,8	37,5	43,8	1610
T18B3 joskus	29,2	46,2	24,7	1286
T18B4 usein	35,4	49,2	15,4	545
T18B5 lähes aina	54,5	36,6	8,9	191
T22 tukiopetuksen määrä luokkien 789 aikana	Tukiopetusta >9 kertaa (72%)	Tukiopetusta 1−8 kertaa (47 %)	Tukiopetusta ei lainkaan (53%)	
T221 Ei kertaakaan	13,2	34,2	52,6	2573
T222 1−3 kertaa	35,5	47,1	17,4	1185
T223 4−8 kertaa	44,1	47,4	8,5	424
T224 yli 8 kertaa	71,5	26,7	1,8	330
Puuttuvien vastausten määrä imputoinnin jälkeen	puuttuvia vastauksia > 17,7 % (50 %)	puuttuvia vastauksia 0−4,8 % tai > 17,7 % (43 %)	puuttuvia vastauksia 1−17,7 % (45−55%)	
Missing 0%	32,1	37,6	30,3	3155
Missing 0−4,8 %	18,3	32	49,7	557
Missing 4,8−9,7 %	7,9	6,8	55,2	543
Missing 9,7−17,7 %	12,2	42,7	45	531
Missing > 17,7 %	49,7	42,6	7,7	531
Hierarkkinen taso H3 (soveltaminen)	<376)	376−559	> 559	
Tehtävätyyppi: monivalinta	<399	399−561	> 561	
Tavoitealue T10 ajatteluntaidot	<361	361−551	> 551	
Sisältöalue S5	<374	374−566	> 566	

LIITE 4.2 Tasojen A1.3, A2.1 ja A2.2 kuvaukset Matematiikan yleisessä viitekehyksessä (CFM; Metsämuuronen, 2018)

Osaamisen taso	**Osaamisen tiivistetty kuvaus** ***"Tällä tasolla oppilas..."***	**Käsitteiden hallinta** ***"Tällä tasolla oppilas..."***	**Operaatioiden hallinta** ***"Tällä tasolla oppilas..."***	**Abstraktin ja matemaattisen ajattelun hallinta** ***"Tällä tasolla oppilas..."***
A1.3 Toiminnallinen alkeistaso	• osaa käyttää sujuvasti luonnollisia numeroita alueella $-\infty - +\infty$ • tuntee ja ymmärtää desimaalijärjestelmän paikkajärjestelmänä ja osaa sitä käyttää • ymmärtää yhteen-, vähennys-, kerto- ja jakolaskun ja osaa käyttää niitä arkielämän tilanteissa • ymmärtää rationaalilukujen käsitteen • ei osaa nostaa lukua luonnollisen luvun potenssiin eikä osaa jakaa lukua alkutekijöihin • ei osaa käyttää suhteellista osuutta, prosenttilaskua ja muita laskentamenettelyjä päivittäisessä elämässä esiin tulevien ongelmien ratkaisemisessa	**Luvut, laskutoimitukset ja algebra** • ymmärtää sujuvasti luonnollisia numeroita alueella $-\infty - +\infty$ • ymmärtää negatiivisen luvun ja murtoluvun käsitteet ja osaa esittää ne eri menetelmin • ymmärtää desimaalijärjestelmän desimaalimurtolukuina • ymmärtää negatiiviset luvut • ymmärtää yhteen-, vähennys-, kerto- ja jakolaskua • ymmärtää sulkumerkinnän käsitteen • on perusymmärrys rationaalilukujen käsitteestä • ei hallitse käsitettä potenssi tai juuri • ei hallitse käsitettä alkutekijä	**Luvut, laskutoimitukset ja algebra** • osaa soveltaa yhteen-, vähennys-, kerto- ja jakolaskua arjen tilanteissa • osaa käyttää desimaalijärjestelmää desimaalimurtolukuina; osaa käyttää negatiivisia lukuja ja murtolukuja • osaa arvioida etukäteen tuloksen suuruuden ja ongelman ratkaistua tarkistaa laskennan vaiheet ja arvioida ratkaisun järkevyyttä • osaa muotoilla ja jatkaa numerosarjoja tai esittää korrelaatioita • osaa tehdä peruslaskelmia pisteillä, janalla, vaakaviivalla, säteellä, viivalla ja kulmalla yksinkertaisilla tasokuvioilla • ymmärtää järjestyksen merkityksen laskelmissa, mutta voi tehdä virheitä varsinaisissa laskelmissa • ei osaa arvioida mahdollista tulosta ja laatia suunnitelmaa ongelman ratkaisemiseksi • ei osaa nostaa lukua luonnollisen luvun potenssiin eikä osaa jakaa lukua alkutekijöihin • ei pysty ratkaisemaan ongelmia, joissa tarvitaan neliöjuurta • ei voi käyttää suhteellista osuutta, prosenttilaskua ja muita laskentamenettelyjä päivittäisessä elämässä esiin tulevien ongelmien ratkaisemisessa	• osaa esittää matemaattisia käsitteitä monin eri tavoin—välineillä, kuvilla, symboleilla, sanoilla, numeroilla tai kaavioilla • yrittää tietoisesti keskittää huomiota havaintoja tehdessään; osaa välittää havaintoja ja ajatuksia monin eri tavoin: toimimalla, puhumalla, kirjoittamalla ja käyttämällä symboleja • osaa kuvata yksinkertaisia tosielämän tilanteita ja ilmiöitä matemaattisesti vertaamalla, luokittelemalla, järjestämällä, rakentamalla ja mallintamalla • osaa ryhmitellä tai luokitella tietyn tai valitun kriteerin perusteella, osaa etsiä yhteistä ominaisuuksia, erottaa laadullisen ja kvantitatiivisen ominaisuuden ja osaa kuvata esine- ja esineryhmiä esittäen niistä oikeita ja epätosia väitteitä • osaa esittää matemaattisia ongelmia uudessa muodossa; osaa tulkita yksinkertaista tekstiä, kuvitusta tai tapahtumaa ja tehdä suunnitelman ongelman ratkaisemiseksi • osaa käyttää sääntöjä • ei pysty havaitsemaan yhtäläisyyksiä ja säännönmukaisuuksia eri tapahtumien välillä

Osaamisen taso	Osaamisen tiivistetty kuvaus *”Tällä tasolla oppilas…”*	Käsitteiden hallinta *”Tällä tasolla oppilas…”*	Operaatioiden hallinta *”Tällä tasolla oppilas…”*	Abstraktin ja matemaattisen ajattelun hallinta *”Tällä tasolla oppilas…”*
A1.3 Toiminnallinen alkeistaso		**Algebra** • Ymmärtää, että kirjain voi symboloida numeroa • Ei hallitse funktion *f(x)* käsitettä **Geometria** • tuntee geometrian peruskäsitteet: pisteet, jana, vaakaviiva, säde, suora ja kulma ja niiden yhteyden yksinkertaisiin tasokuvioihin • ymmärtää mittauksen periaatteen • ei hallitse muita kuin geometrisia peruskuvioita	**Algebra** • osaa suorittaa yksinkertaisia laskutoimituksia ensimmäisen asteen yhtälöllä piiloluvun muunnelmana; esim. 2 + a = 3, mikä on a? a = 1 • ei pysty ratkaisemaan yksinkertaista ensimmäisen asteen yhtälöä • ei osaa sieventää yksinkertaista algebrallista lauseketta • ei osaa laskea potensseja • ei osaa muotoilla yksinkertaista yhtälöä jokapäiväiseen elämään liittyvästä ongelmasta eikä ratkaista sitä algebrallisesti eikä päättelyllä **Geometria** • osaa muodostaa kuvioita annettujen ohjeiden mukaisesti, huo-maa yksinkertaisten geometristen kuvioiden ominaisuudet ja tuntee tasokuvioiden käsitteiden muodostaman rakenteen • tunnistaa samankaltaisuuden; voi heijastaa kuvan viivan yli ja laajentaa ja pienentää lukuja tietyllä suhteella; osaa tunnistaa kuvioita, jotka ovat symmetrisiä viivan suhteen • osaa arvioida mitattavan kohteen koon ja mittaustuloksen järkevyyden ja miten tulos ilmaistaan sopivilla mittayksiköillä • osaa laskea suunnikkaan ja kolmion pinta-alan ja piirin • osaa käyttää yksinkertaisia heijastuksia ja laajennuksia	• ei osaa käyttää puheessaan loogisia elementtejä, kuten ja, tai, jos. • ei voi arvioida yksinkertaisten väitteiden totuutta • ei pysty muuttamaan yksinkertaista ongelmaa tekstissä matemaattiseksi esitysmuodoksi, laatimaan suunnitelmaa ongelman ratkaisemiseksi, ratkaista se eikä tarkistamaan tuloksen oikeellisuutta • ei osaa käyttää luokittelua matemaattisten ongelmien ratkaisemisessa • ei osaa esittää mahdollisia vaihtoehtoisia ratkaisuja systemaattisesti käyttäen taulukkoa, jalavaa, polkukaaviota tai muuta kaaviota

Osaamisen taso	Osaamisen tiivistetty kuvaus *”Tällä tasolla oppilas…”*	Käsitteiden hallinta *”Tällä tasolla oppilas…”*	Operaatioiden hallinta *”Tällä tasolla oppilas…”*	Abstraktin ja matemaattisen ajattelun hallinta *”Tällä tasolla oppilas…”*
A1.3 Toiminnallinen alkeistaso		**Tiedon käsittely, tilastot ja todennäköisyys** • ymmärtää mahdottoman ja varman käsitteet • ymmärtää todennäköisyyden käsitteen	• osaa havaita geometristen peruskuvioiden ominaisuudet • tietää piirin, pinta-alan ja tilavuuden käsitteet, mutta ei osaa laskea niitä • ei tunnista muita kuin geometrisia perusmuotoja eikä tiedä niiden ominaisuuksia • ei osaa käyttää harppia ja viivainta yksinkertaisten geometristen rakenteiden tekemiseen • ei löydä samanlaisia, yhteneviä ja symmetrisiä kuvioita eikä osaa soveltaa tätä taitoa kahden kulman välisten ominaisuuksien tutkimiseen yksinkertaisissa tilanteissa • ei osaa soveltaa kahden kulman välisiä suhteita yksinkertaisissakaan tilanteissa • ei osaa käyttää Pythagoraan lausetta ja trigonometriaa suorakulmaisen kolmion osien ratkaisemiseen • ei osaa suorittaa mittausta ja siihen liittyviä laskelmia ja muuntaa yleisimpiäkään mittayksiköitä **Tiedon käsittely, tilastot ja todennäköisyys** • osaa kerätä tietoja ja järjestää, luokitella ja esittää niitä tilastolukuina • osaa lukea yksinkertaisia taulukoita ja kaavioita • osaa selventää erilaisten tapahtumien ja vaihtoehtojen määrää ja arvioida, mikä on mahdoton tai varma tapahtuma	

Osaamisen taso	Osaamisen tiivistetty kuvaus *”Tällä tasolla oppilas…”*	Käsitteiden hallinta *”Tällä tasolla oppilas…”*	Operaatioiden hallinta *”Tällä tasolla oppilas…”*	Abstraktin ja matemaattisen ajattelun hallinta *”Tällä tasolla oppilas…”*
A1.3 Toiminnallinen alkeistaso		**Funktiot** • ymmärtää koordinaatiston käsitteen • ymmärtää numerosarjan takana olevien sääntöjen käsitteen • ei hallitse lineaarifunktion, vakion tai kulmakertoimen käsitteitä	• ymmärtää todennäköisyyden ja satunnaisuuden merkityksen päivittäisissä tilanteissa, mutta ei pysty määrittämään mahdollisten tapahtumien määrää ja tekemään yksinkertaista empiiristä todennäköisyyteen liittyvää tutkimusta • ei osaa lukea erilaisia taulukoita ja kaavioita eikä määrittää frekvenssejä, keskiarvoa, mediaania ja moodia annetusta materiaalista **Funktiot** • osaa havainnollistaa kahta muuttujaa koordinaatiossa • voi jatkaa yksinkertaista numerosarjaa annetun säännön mukaan, mutta ei osaa kuvata tietyn numerosarjan yleistä sääntöä suullisesti • ei osaa määrittää pisteen koordinaatteja koordinaattijärjestelmässä • ei pysty laatimaan taulukkoa numeropareista annetun säännön mukaan • ei osaa etsiä lineaarifunktion nollapistettä • ei tiedä lineaarisen yhtälön vakion ja kulmakertoimen merkitystä; ei osaa määrittää kahden suoran leikkauspistettä piirtämällä niitä	

Osaamisen taso	Osaamisen tiivistetty kuvaus *”Tällä tasolla oppilas…”*	Käsitteiden hallinta *”Tällä tasolla oppilas…”*	Operaatioiden hallinta *”Tällä tasolla oppilas…”*	Abstraktin ja matemaattisen ajattelun hallinta *”Tällä tasolla oppilas…”*
A2.1 Kehittyvä perustaso	• käyttää suhdetta, prosentin laskemista ja muita laskentamenettelyjä ratkaistakseen ongelmia arkielämässä vastaan tulevissa tilanteissa • pystyy muodostamaan yksinkertaisen yhtälön ja ratkaisemaan sen joko algebrallisesti tai päättelemällä ratkaistakseen ongelmia arkielämässä vastaan tulevissa tilanteissa • osaa laskea suorakulmion, ympyrän ja yksinkertaisen monikulmion piirin, alueen ja yksinkertaisten kappaleiden tilavuuden • ymmärtää todennäköisyyden ja satunnaisuuden merkityksen arkielämän tilanteissa • tietää, kuinka koordinaattipisteet määrätään koordinaatistoon • ei hallitse Pythagoraan lauseeseen ja trigonometriaan liittyviä käsitteitä eikä osaa käyttää näitä suorakulmaisen kolmien osien ratkaisemisessa • ei hallitse frekvenssejä, keskiarvoa, mediaania ja moodia eikä juuren ja potenssin käsitteitä eikä osaa laskea näihin liittyviä laskuja	**Luvut, laskutoimitukset ja algebra** • ymmärtää suhteellisen osuuden ja prosentin käsitteet • ymmärtää rationaalilukujen käsitteen • ymmärtää sieventämisen käsitteen • ymmärtää jakojäännöksen käsitteen • on hyvä ymmärrys sulkujen käsitteestä ja osaa käyttää sitä ilman virheitä • ei hallitse juuren ja potenssin käsitteitä **Algebra** • ymmärtää ensimmäisen asteen yhtälön • ymmärtää suhteiden, suhteellisten osuuksien ja prosenttiosuuksien käsitteet • ei hallitse yhtälöparin käsitettä **Geometria** • ymmärtää hyvin ympärysmitan, pinta-alan ja tilavuuden käsitteet • ei hallitse yhtenevien ja symmetristen kuvioiden käsitteitä • ei hallitse Pythagoraan lauseeseen ja trigonometriaan liittyviä käsitteitä	**Luvut, laskutoimitukset ja algebra** • osaa arvioida mahdollisen tuloksen ja laatia suunnitelman ongelman ratkaisemiseksi; omaa luotettavat peruslaskentataidot • käyttää suhde- ja prosenttilaskentaa ja muita laskentamenettelyjä päivittäisessä elämässä esiin tulevien ongelmien ratkaisemisessa • osaa jakaa rationaalisilla luvuilla ja käsitellä jakojäännöstä • käyttää varmasti laskujärjestystä • ei osaa käyttää potensseja ja juuria **Algebra** • osaa ratkaista yksinkertaisen ensimmäisen asteen yhtälön • voi sieventää yksinkertaisen algebrallisen lausekkeen • osaa muotoilla yksinkertaisen yhtälön jokapäiväiseen elämään liittyvästä ongelmasta ja ratkaista sen joko algebrallisesti tai päättelyllä • Ei osaa laskea potensseja ja juuria **Geometria** • tunnistaa muita kuin geometrisia perusmuotoja ja tietää niiden ominaisuuksia • osaa laskea suorakulmion, ympyrän ja yksinkertaisen monikulmion piirin, alueen ja yksinkertaisten kappaleiden tilavuuden • käyttää harppia ja viivainta yksinkertaisten geometristen rakenteiden tekemiseen • ei löydä samanlaisia, yhteneviä ja symmetrisiä kuvioita	• havaitsee yhtäläisyyksiä ja säännönmukaisuuksia eri tapahtumien välillä • osaa käyttää puheessaan loogisia elementtejä kuten ja, tai, jos on (olemassa), ei ole olemassa • osaa arvioida yksinkertaisten väitteiden totuuden

Osaamisen taso	Osaamisen tiivistetty kuvaus *”Tällä tasolla oppilas…”*	Käsitteiden hallinta *”Tällä tasolla oppilas…”*	Operaatioiden hallinta *”Tällä tasolla oppilas…”*	Abstraktin ja matemaattisen ajattelun hallinta *”Tällä tasolla oppilas…”*
A2.1 Kehittyvä perustaso		 **Tiedon käsittely, tilastot ja todennäköisyys** • ymmärtää empiirisen tiedonhankinnan proseduurin • ymmärtää todennäköisyyden ja satunnaisuuden merkitykset • ei hallitse frekvensseihin, keskiarvoihin, mediaaniin ja moodiin liittyviä käsitteitä **Funktiot** • ymmärtää koordinaatisto-järjestelmän ja siihen liittyviä käsitteitä • ei hallitse lineaarisen yhtälön vakion ja kulmakertoimen käsitteitä	• ei osaa käyttää Pythagoraan lausetta ja trigonometriaa suorakulmaisen kolmion osien ratkaisemiseen • ei voi suorittaa mittausta ja siihen liittyviä laskelmia ja muuntaa yleisimpiä mittayksiköitä **Tiedon käsittely, tilastot ja todennäköisyys** • osaa määrittää mahdollisten tapahtumien lukumäärä ja järjestää yksinkertainen empiirinen tutkimus todennäköisyydestä • ymmärtää todennäköisyyden ja satunnaisuuden merkityksen päivittäisissä tilanteissa • ei osaa määrittää frekvenssejä, keskiarvoa, mediaania ja moodia annetusta materiaalista **Funktiot** • osaa määrittää pisteen koordinaatit koordinaattijärjestelmässä • osaa laatia lukupareista taulukon annetun säännön mukaisesti • osaa jatkaa numerosarjaa annetun säännön mukaisesti ja osaa kuvailla tietyn numerosarjan yleissääntöä suullisesti • ei osaa etsiä lineaarifunktion nollapistettä • ei osaa käyttää lineaarisen yhtälön vakiota ja kulmakerrointa ja määrittää kahden suoran leikkauspistettä piirtämällä niitä	

Osaamisen taso	Osaamisen tiivistetty kuvaus *”Tällä tasolla oppilas…”*	Käsitteiden hallinta *”Tällä tasolla oppilas…”*	Operaatioiden hallinta *”Tällä tasolla oppilas…”*	Abstraktin ja matemaattisen ajattelun hallinta *”Tällä tasolla oppilas…”*
A2.2 Toiminnallinen perustaso	• hallitsee potenssilaskun perusteet ja pystyy kytkemään sen kertolaskuun • hallitsee neliön (toisen potenssin) ja sen yhteyden käytännön tilanteisiin • löytää samanlaiset, yhteneväiset (kongruentit) ja symmetriset muodot ja pystyy näiden avulla tutkimaan kahden kulman ominaisuuksia yksinketaisessa tilanteessa • lukee erilaisia taulukoita ja diagrammeja ja pystyy määrittämään frekvenssit, keskiarvot, mediaanin ja moodin annetusta aineistosta • tietää, kuinka etsiä lineaarisen funktion nollakohta • Ei hallitse potenssi- tai eksponenttifunktioita eikä analyyttistä geometriaa	**Luvut, laskutoimitukset ja algebra** • ymmärtää juuren ja potenssin käsitteet • ymmärtää alkutekijän käsitteen • ei ymmärrä lukujoukkojen käsitettä • ei ymmärrä potenssifunktion tai eksponenttifunktion käsitettä **Algebra** • ymmärtää yhtälöparin käsitteen • ei hallitse potenssiyhtälöiden käsitteitä eikä potenssifunktioita eikä murtolukupotenssia **Geometria** • ymmärtää yhtenevien ja symmetristen kuvioiden käsitteet • ymmärtää Pythagoraan lauseen ja trigonometrian käsitteet • ei hallitse polynomi- tai eksponenttifunktioihin eikä analyyttiseen geometriaan liittyviä käsitteitä	**Luvut, laskutoimitukset ja algebra** • osaa nostaa luvun luonnollisen luvun potenssiin • osaa jakaa luvun alkutekijöihin • ei hallitse lukuteoriaa eikä aritmeettisia ja geometrisia lukujoukkoja • ei osaa käyttää potenssien ja neliöjuurien laskentasääntöjä • ei osaa tutkia teho- ja eksponenttifunktioita **Algebra** • osaa laskea potensseja luonnollisilla luvuilla • osaa ratkaista ongelmia, joissa tarvitaan neliöjuuria • osaa käyttää yhtälöpareja yksinkertaisten ongelmien ratkaisemiseen • osaa arvioida tuloksen logiikkaa ja tarkastella ratkaisunsa eri vaiheita • ei pysty ratkaisemaan potenssi-yhtälöitä tai potenssifunktioita • ei osaa käyttää murtopotensseja **Geometria** • löytää samanlaisia, yhteneviä ja symmetrisiä kuvioita ja osaa soveltaa tätä taitoa kahden kulman välisten ominaisuuksien tutkimiseen yksinkertaisissa tilanteissa • soveltaa kahden kulman välisiä suhteita yksinkertaisissa tilanteissa • käyttää Pythagoraan lausetta ja trigonometriaa suorakulmaisen kolmion osien ratkaisemiseen • suorittaa mittauksia ja siihen liittyviä laskelmia sekä muuntaa yleisimmät mittayksiköt	• osaa muuttaa yksinkertaisen tekstin tehtävän matemaattiseksi esitysmuodoksi, tehdä suunnitelman ongelman ratkaisemiseksi, ratkaista sen ja tarkistaa tuloksen oikeellisuuden • osaa käyttää luokittelua matemaattisten ongelmien ratkaisemisessa • osaa esittää mahdollisia vaihtoehtoisia ratkaisuja systemaattisesti taulukon, puukaavion, polkukaavion tai muun kaavion avulla

Osaamisen taso	Osaamisen tiivistetty kuvaus *”Tällä tasolla oppilas…”*	Käsitteiden hallinta *”Tällä tasolla oppilas…”*	Operaatioiden hallinta *”Tällä tasolla oppilas…”*	Abstraktin ja matemaattisen ajattelun hallinta *”Tällä tasolla oppilas…”*
A2.2 Toiminnallinen perustaso		**Tiedon käsittely, tilastot ja todenäköisyys** • ymmärtää frekvenssin, keskiarvon, mediaanin ja moodin käsitteet **Funktiot** • ymmärtää lineaarisen yhtälön käsitteen • ymmärtää lineaarisen yhtälön vakion ja kulmakertoimen käsitteet • ymmärtää leikkauspisteen käsitteet • ei hallitse polynomifunktioihin ja analyyttisen geometriaan liittyviä käsitteitä	• ei hallitse ympyrän geometriaa, sini- ja kosinisääntöjä eikä analyyttistä geometriaa • ei hallitse vektorilaskentaa **Tiedon käsittely, tilastot ja todenäköisyys** • lukee erilaisia taulukoita ja kaavioita sekä määrittää annetusta materiaalista taajuuksia, keskiarvoa, mediaania ja moodia • ei hallitse diskreettejä ja jatkuvia tilastojakaumia, jakeluparametreja, matemaattista ja tilastollista todennäköisyyttä tai kombinatoriikkaa • ei hallitse todennäköisyyksien laskemista koskevia sääntöjä, diskreettejä ja jatkuvia todennäköisyysjakaumia, diskreettien jakaumien odotusarvoja eikä normaalijakaumaa **Funktiot** • osaa etsiä lineaarifunktion nollapisteen • osaa jatkaa numerosarjaa annetun säännön mukaan ja osaa kuvata tietyn numerosarjan yleissääntöä suullisesti • tietää lineaarisen yhtälön vakion ja kulmakertoimen merkityksen; osaa määrittää kahden suoran leikkauspisteen piirtämällä ne • ei hallitse polynomifunktioita-differentiaalilaskennan alkeita, rationaali- ja logaritmifunktioita tai trigonometrisiä funktioita	

5

Tunteiden mittaaminen matematiikan arvioinnissa— Tunnemittari, uskomukset ja kontrolli–arvo-teoria

Reito Visajaani Salonen, Helsingin yliopiston humanistis-yhteiskuntatieteellinen instituutti

5

- Vuoden 2021 matematiikan osaamisen arvioinnin yhteydessä kerättiin aineistoa oppilaskyselyä varten kehitetyllä tunnemittarilla. Tunnemittaria tutkittiin sisäisen rakenteen ja toiminnan osalta eksploratiivisella ja konfirmatorisella faktorianalyysillä. Lisäksi tunnemittaria mallinnettiin konfirmatorisella faktorianalyysillä yhdessä Karvin oppimistulosarvioinneissa vakiintuneen uskomusmittarin kanssa. Tunteiden ja uskomusten toimintaa peilattiin Pekrunin kontrolli–arvo-teoriaan.

- Tunnemittarista pystyttiin erottamaan positiivinen ja negatiivinen ulottuvuus. Myös aktivaatioon, aktivoiva ja passivoiva -jakoon, perustuva luokittelu pystyttiin erottamaan sekä positiivisesta että negatiivisesta ulottuvuudesta.

- Tunnemittari toimii itsenäisenä mittarina hyvin tai vähintäänkin kohtuullisesti jokaisessa tutkitussa mallissa.

- Mittari vaatii vielä kehitystyötä, mikäli halutaan erotella tunteita aktivaatioluokittelun avulla.

5.1 Johdanto

Matematiikan arviointien yhteydessä tunteiden tutkiminen ja tulkitseminen on ollut harvinaista laajoilla kyselyotannoilla. Tunteet ja niiden voimakkuus vaihtuvat nopeatempoisemmin kuin uskomukset (Hannula, 2012), minkä vuoksi kansallisissa oppimistulosarvioinneissa on keskitytty tarkastelemaan uskomusten ja osaamisen välistä yhteyttä (Hannula ym., 2014; Tuohilampi & Giaconi, 2013). Tunteiden ajatellaan sisältävän fyysisiä reaktioita ja vaikuttavan kognitiivisiin prosesseihin, kun uskomukset taas liittyvät vahvasti päätöksentekoon ja metakognitiivisiin piirteisiin (DiMartio & Zan, 2011). Tunteita on yleensä tutkittu tehtävänratkaisun yhteydessä (De Corte ym., 2011; DeBellis ym., 2006; Goldin, 2000), mutta tässä arvioinnissa matematiikkaan liittyviä tunteita on tarkasteltu kysymällä eri tunteiden kokemisen yleisyyttä viimeisen vuoden aikana.

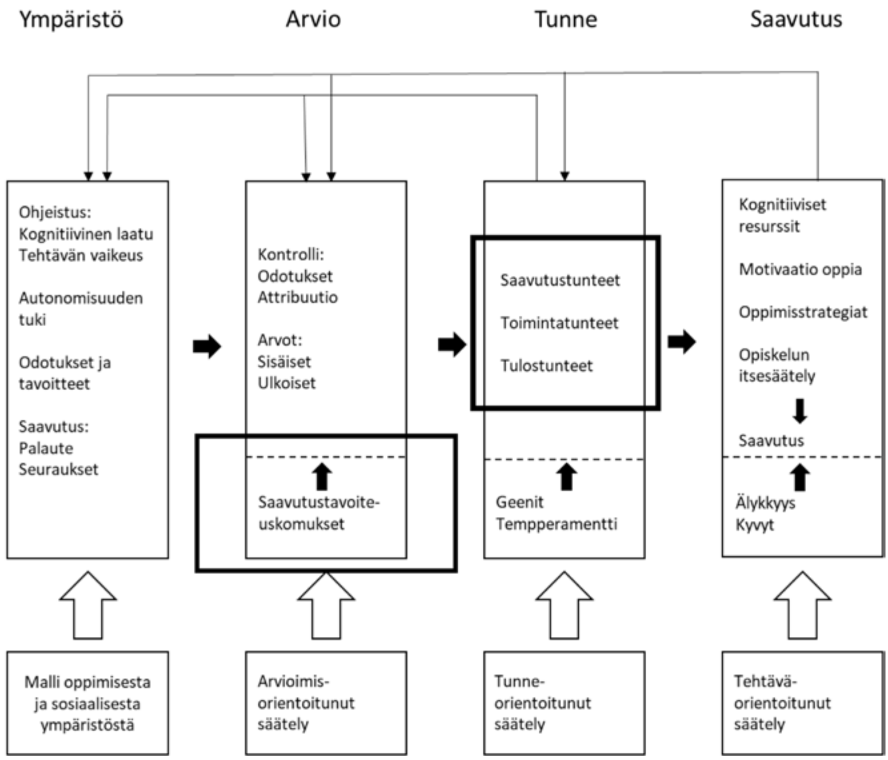
Tunteiden tutkimuksessa on oltava varovainen tulkintojen kanssa, koska varsinkin kyselytutkimuksissa kyselyajankohdalle läheisten tapahtumien vaikutus voi korostua vastaamisessa. Koska tässä tutkimuksessa matematiikkaan liittyviä tunteita mitattiin niiden kokemisen yleisyyden mukaan, voidaan niiden ajatella olevan luonteeltaan samalla tavalla pysyviä kuin uskomusten.

Useat matematiikkaan liittyviin tunteisiin, uskomuksiin ja osaamiseen keskittyvät tutkimukset lähestyvät aihetta negatiivisten tunteiden, kuten ahdistuksen, kautta (Villavicencio & Bernardo, 2016). Pekrun ja kollegat (2002) osoittivat meta-analyysissään ahdistuksen olleen useimmiten tutkittu tunne matematiikan kontekstissa 2000-luvun alkuun mennessä, jonka jälkeen Huang (2011) omassa meta-analyysissään esitti positiivisiin tunteisiin liittyvän tutkimuksen lisääntyneen ja tutkimuksen osoittaneen, että positiiviset tunteet kumoavat negatiivisten tunnetilojen vaikutuksia. Villavicencio ja Bernardo (2016) osoittivat trigonometrian kurssin ja kurssiarvioinnin kontekstissa positiivisten tunteiden yhteyden minäpystyvyyteen ja itsesäätelyyn.

Karvin 9. luokan matematiikan arvioinnissa sovellettiin mittaria, joka sisälsi kymmenen eri tunnetta. Näitä tunteita oli selvennetty vastaajalle tunteita kuvaavilla emojeilla. Mittari sisälsi viisi positiivista ja viisi negatiivista tunnetta. Jotta tunteisiin liittyviä teorioita voitaisiin tässä yhteydessä tulkita mielekkäästi, on selvitettävä, miten tunnemittari reagoi oppimistuloksiin (ks. Salonen ym., 2023). Tässä luvussa käsittelenkin tunnemittarin vaihtoehtoisia käyttötapoja ja arvioin sen yleistä toimivuutta uskomusten ja tunteiden näkökulmasta. Tutkittavat uskomukset (minäkäsitys, matematiikasta pitäminen ja matematiikan hyödyllisyys) ovat Karvin vakiomittareita, jotka pohjautuvat Fennema–Sherman-mittaristoon, josta Metsämuuronen (2012) on tehnyt tarkemman toimivuustarkastelun ja redusoinnin (ks. vuoden 2021 aineiston osalta Metsämuuronen & Nousiainen, 2023; luku 2). Tässä tutkimuksessa mielenkiinto ei ole yksistään tunteiden suunnassa vaan myös siinä, miten nyt käytetty tunnemittari kykenee erottelemaan nämä tunteet ja kuinka käyttökelpoisia ne ovat jatkotutkimuksissa. Lisäksi tunteiden aktivaation roolia pyritään selvittämään suhteessa uskomuksiin sekä tutkitaan, erotteleeko tunnemittari myös aktivaatiota.

5.2 Teoriataustaa

Uskomusten, tunteiden ja osaamisen välisten yhteyksien tutkimiseen on useita teoreettisia lähestymistapoja. Tämän artikkelin kannalta soveltuvimmaksi arvioitiin Pekrunin (2006) kontrolli–arvo-teoria (*control-value theory*), josta pääkohdat on esitetty kuviossa 34. Tämä malli kuuluu osaamistavoiteteorioiden laajaan kenttään (Elliot & Dweck, 1988; Elliot & Pekrun, 2007; Pekrun ym., 2006, 2009). Pekrunin ja kollegoiden (2002) kirjallisuuskatsauksen perusteella tutkituin tunne on ahdistuneisuus, kun tarkastellaan tunteiden sekä oppimisen ja osaamisen välisiä yhteyksiä. Uudemmissa tutkimuksissa positiivisten tunteiden tutkimuksen on havaittu lisääntyneen (mm. Huang, 2011), kuten myös tutkimus positiivisten sekä negatiivisten tunteiden vaikutuksesta saavutuksiin (mm. Camacho-Morles ym., 2021).



KUVIO 5.34. Pekrunin (2006) malli kontrolli–arvo-teoriasta. Tunne- ja uskomusosat korostettuina

Tässä tutkimuksessa käsitellään tunnemittarin toimintaa itsenäisenä mittarina sekä suhteessa Pekrunin (2006) kontrolli–arvo-teoriaan. Mittarin soveltumista tutkimuskäyttöön arvioidaan tulosten ymmärrettävyyden ja sovellettavuuden osalta. Mittarin ymmärrettävyyttä käsitellään suhteessa vallitsevaan teoriaan ja soveltuvuutta tutkitaan tilastollisesti muun muassa konfirmatorisella faktorianalyysillä. Tunnemittarin sisäisen toiminnan tarkastelun jälkeen peilataan sen toimivuutta suhteessa uskomusmittareihin. Tällöin käsitellään kontrolli–arvo-teorian mukaisesti arvio- ja tunnetasojen välistä yhteyttä. Tunteiden ja uskomusten välillä katsotaan olevan yhteys, kun niiden välillä on tilastollisesti merkitsevä korrelaatio. Mittarin käytettävyyttä arvioidaan muodostuttujen latenttien muuttujien määrän ja tilastollisten tunnuslukujen perusteella. Koska käytettävissä on kaiken kaikkiaan kymmenen tunnetta, on oletettavaa, että sellainen useamman faktorin ratkaisu, jossa perinteisesti odotetaan vähintään kolmen väittämän sijoittuvan samalle faktorille, ei ole mahdollinen. Analyysit toteutetaan lähtökohtaisesti mittarin jatkokehittämiseksi.

Pekrunin (2006) kontrolli–arvo-teoriaa hyödynnetään uskomusten ja tunteiden välisen yhteyden mallintamisessa. Pekrun (2006) esittää, että yksilön opittavan asian tärkeyteen ja arvokkuuteen liittämät uskomukset johtavat nautintoon. Arvioidessaan tilanteen itselleen arvokkaaksi tai tärkeäksi voi yksilö siis kokea nautintoa opittavasta asiasta. Tunnereaktio vaikuttaa arvion lisäksi päätökseen jatkaa opittavan asian käsittelyä ja kognitiivista prosessointia. Lopulta prosessointi sulautuu nautinnon tunteeseen. Uskomusten ajatellaan olevan määrittävässä roolissa tunnereaktion muodostumiselle (Hannula, 2012). On kuitenkin osoitettu, ettei suunta ole vain uskomuksista tunteisiin ja osaamistasoon, vaan myös tunteilla ja osaamistasolla on vaikutus uskomuksiin (Villavicencio & Bernardo, 2016).

Kontrolli–arvo-teoria ei täysin vastaa Karvin mittaristojen asetelmaa, mutta on hyödyllistä ymmärtää teorioiden sovellettavuus kansallisissa oppimistulosarvioinneissa. Esimerkiksi tunnemittarin kehitystyössä ”akateemisten tunteiden” (*academic emotions*) malli (Pekrun, 2017; Pekrun ym. 2002) voi edesauttaa monipuolisesti tulosten tulkintaa arviointien yhteydessä. Pekrunin viitekehyksessä akateemisilla tunteilla tarkoitetaan erilaisia tunnekokemuksia, joita oppijat kokevat oppimis- ja opettamistilanteissa (Pekrun ym. 2002).

Tunnemittareissa tunteet on perinteisesti jaettu kirjallisuudessa tunnearvoiltaan positiivisiin ja negatiivisiin tunteisiin. Pekrunin ja kollegoiden myöhemmissä akateemisista tunteista tekemissä tutkimuksissa (mm. Pekrun, 2017; Pekrun & Linnenbrink-Garcia, 2012; Pekrun & Perry, 2014; ks. myös Barrett ym., 2018) kaksinapaisen tunnearvotulkinnan lisäksi tarkastelua on laajennettu myös tunteen aktivaatioon. Pekrun ja Perry (2014) esittävät taulukon 23 mukaisesti, että positiiviset ja negatiiviset tunteet voidaan yhtäältä erotella aktivoiviin ja passivoiviin, ja toisaalta niiden ilmentymät voidaan luokitella tuleviin (*prospective*) ja takautuviin (*retropsective*). Tämä jaottelu tunnearvon ja aktivaation mukaan mahdollistaa tunteiden ymmärtämisen monitasoisempana ilmiönä, jossa saman tunnearvon, kuten positiivisten tunteiden, ei oleteta tuottavan automaattisesti samanlaista käyttäytymistä (Barrett ym., 2018). Tämän artikkelin näkökulmasta aktivaation tarkastelun voi nähdä laajentavan ymmärrystä siitä, miten oppilaan tunnetilat ja oppimistulokset liittyvät toisiinsa, ja se on siten pohja tulevien oppimistulosarviointien kehittämiselle.

TAULUKKO 5.23. Akateemisten tunteiden kolmiulotteinen taksonomia (Pekrun, 2017; Pekrun & Perry, 2014)

	Positiivinen		Negatiivinen	
Tavoitteen kohde	Aktivoiva	Passivoiva	Aktivoiva	Passivoiva
Toiminto	Nautinto	Rentoutuminen	Viha	Tylsyys
			Turhautuminen	
Ilmentymä / Tuleva	Toivo	Helpotus (ennakoiva)	Ahdistus	Toivottomuus
	Ilo (ennakoiva)			
Ilmentymä / Takautuva	Ilo	Tyytyväisyys	Häpeä	Surullisuus
	Ylpeys	Helpotus	Viha	Pettymys
	Kiitollisuus			

Tunnemittarin osalta ei tarkastella mallin laajempaa toimivuutta suhteessa uskomuksiin ja osaamistasoon, vaan ainoastaan yleisen tunnemittarin ja uskomusten yhteyden toimivuutta mallintamisessa. Mittarin toimivuus määrittyy aineistolähtöisesti sen mukaan, kuinka hyvin eri ulottuvuudet ovat keskenään yhteydessä. Koska kehitetyn tunnemittarin taustalla ei ole mikään vallitseva teoria, on mahdollista, ettei tunnemittari ole tarkasteluissa teorian näkökulmasta toimivuudeltaan parhain. Monen teorian soveltaminen tunnemittariin ei olisi yksittäisen tutkimuksen näkökulmasta mielekästä, niinpä Pekrunin (2006) kontrolli–arvo-teoria valittiin tämän artikkelin lähtökohdaksi tarkasteltaessa tunnemittarin soveltuvuutta osana oppimistulosarviointeja.

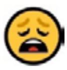
5.3 Tunne- ja uskomusmittarit

5.3.1 Uskomusmittarit

Uskomusmittaristo on ollut käytössä Karvin matematiikan oppimistulosarvioinneissa vakiomittarina (mm. Julin & Rautopuro, 2016; Metsämuuronen 2017; Niemi & Metsämuuronen, 2010). Mittaristo käsittää matematiikasta pitämisen (esimerkiksi ”*Matematiikka on yksi lempiaineistani.*”), minäkäsityksen itsestä matematiikan osaajana (esimerkiksi ”*Mielestäni olen hyvä matematiikassa.*”) sekä hyödyllisyysuskomukset (esimerkiksi ”*Uskon tarvitsevani työelämässä matematiikan tietoja ja taitoja.*”), joista jokaisesta on esitetty viisi väittämää. Käytetyt väittämät ja niiden väliset suhteet on esitetty tarkemmin Metsämuurosen (2012) mittarin validointia koskevassa artikkelissa. Kysymyksiin vastattiin asteikolla yhdestä viiteen (*1 = Olen täysin eri mieltä, 2 = Olen jonkin verran eri mieltä, 3 = Kantani on epävarma tai minulla ei ole selvää käsitystä, 4 = Olen jonkin verran samaa mieltä, 5 = Olen täysin samaa mieltä*). Perinteisesti näistä kysymyksistä on muodostettu summamuuttujat, joiden arvoja on tulkittu. Tässä artikkelissa käytetään latenttien muuttujien eli faktorianalyysin kautta löydettyjen faktoreiden mallinnuksen perustana kuitenkin uskomusmittarien yksittäisiä kysymyksiä (ks. myös luku 2) ja dimensiot kytketään tunnemittarin dimensioihin.

5.3.2 Tunnemittarit

Emootio- eli tunnemittari oli nyt ensimmäistä kertaa käytössä matematiikan osaamisen kansallisessa arvioinnissa. Tunnemittaristo (kuvio 35) sisälsi kymmenen tunnetilaa. Tunteet pyydettiin arvioimaan vastaamalla kysymykseen: ”Mieti matematiikan opiskeluasi kokonaisuutena. Missä määrin alla esitetty tunnetila yhdistyy matematiikan opiskeluusi?”

	ei lainkaan	harvoin	joskus	usein	lähes aina
 innostunut					
 kiinnostunut					
 onnistunut					
 tyytyväinen					
 varma					
 vihainen					
 avuton					
 ahdistunut					
 pettynyt					
 epävarma					

KUVIO 5.35. Tunnemittaristo ja vastausvaihtoehdot

Emojien eli hymiöiden käytön ajateltiin helpottavan vastaamista sekä tekevän siitä oppilaille mielenkiintoisempaa. Mittari koostuu viidestä positiivisesta ja viidestä negatiivisesta tunteesta. Koska mittaristoa ei ollut käytetty aiemmin, mittariston tutkiminen ja käyttökelpoisuuden tarkastelu toteutettiin pitkälti aineistolähtöisesti. Mittarin perinteisiä luotettavuustarkasteluja käsittelevät toisaalla Metsämuuronen ja Nousiainen (2023; ks. luku 2).

5.4 Menetelmät

Aineistoon sovelletaan eksploratiivista sekä konfirmatorista faktorianalyysia. Eksploratiivista faktorianalyysia hyödynnetään etsittäessä artikkelin aineistosta tunnemittarin eri ulottuvuuksia ja sisäistä rakennetta. Konfirmatorisella faktorianalyysillä puolestaan varmistetaan Karvin aikaisemmin julkaisuissaan käyttämän Fenneman–Shermanin mittariin (Metsämuuronen, 2012) pohjautuvan uskomusmittariston toimivuus tunnemittarin kanssa.

5.4.1 Aineisto

Tunne- ja uskomusmittarien osalta käytettiin Karvin vuoden 2021 yhdeksännen luokan matematiikan arvioinnin aineistosta ne vastaajat, joilla ei ollut matematiikan osalta yksilöllistettyä opetussuunnitelmaa (HOJKS) (N = 12 356). HOJKS-päätöksen saaneet oppilaat (N = 347) olivat arvioinnin näkökulmasta erityisryhmä, jonka opetus ja tavoitteet eivät vastaa laajuudeltaan muita osallistujia.

5.4.2 Aineiston käsittely

Aineistosta poistettiin erityisryhmänä erityisen tuen piiriin kuuluvat oppijat, koska heidän vaatimuksensa ja lähtökohtansa opetukseen ovat perusjoukosta poikkeavat. Lisäksi muodostettiin valmiiksi summamuuttujat eli keskiarvot mittarin samaan ulottuvuuksiin kuuluvista väittämistä. Uskomusmittarin väittämät samaan summamuuttujaan olivat aikaisempien tutkimusten (Metsämuuronen, 2012, 2017; Hannula ym., 2014) mukaiset, ja tunnemittarin väittämät yhdistettiin summamuuttujaksi eksploratiivisen faktorianalyysin tulosten mukaisesti. Analyyseissa käytetäänkin pääsääntöisesti malleja, joissa summamuuttujien pohjalla oleva latentti ilmiö on mallinnettu siihen kuuluvien osakysymysten avulla. Summamuuttujien avulla kuvataan aineistoa keskiarvon ja keskihajontojen näkökulmasta.

5.4.3 Analyysimenetelmät

Analyyseissä tutkittiin, tukeeko aineisto oppilaskyselyn tunnemittarin jakamista myönteisiin ja kielteisiin ulottuvuuksiin sekä tarkemmin vielä aktivoiviin ja passivoiviin komponentteihin. Oppilaskyselyssä käytettiin myös Karvin arvioinneissa vakiintunutta matematiikasta pitämisen, minäkäsityksen ja hyödyllisyysuskomuksen mittaristoa. Tunnemittariston tutkimisen kannalta oleellista oli selvittää aineistolähtöisesti, miten eri tunteet ryhmittyvät keskenään, ja lopuksi tarkistaa, kuinka mittarit toimivat teorian näkökulmasta.

5.4.4 Faktorianalyysi

Eksploratiivisen faktorianalyysin avulla etsittiin tunnemittarin kysymysten muodostamia yhtenäisiä kokonaisuuksia eli faktoreita, joita voidaan kutsua myös latenteiksi ilmiöiksi. Faktoreiden lopullisen rakenteen tulee perustua arvioon väittämien yhtenäisyydestä ja yhteydestä taustateoriaan. Tunnemittariston osalta tutkittiin eksploratiivisella faktorianalyysillä kahden, kolmen ja neljän faktorin ratkaisuja käyttäen suurimman uskottavuuden (*maximun likelihood*) estimointia ja *direct oblimin* -vinokulmarotaatiota. Analyysit suoritettiin aluksi perustuen ominaisarvon $\lambda = 1$ ylittäviin faktoreihin, ja muut vaihtoehtoiset faktorianalyysit toteutettiin asettamalla analyysille vaatimukseksi haluttu määrä faktoreita. Konfirmatorisella faktorianalyysillä tutkittiin eksploratiivisen faktorianalyysin pohjalta löydettyjen tunnemittarien latenttien muuttujien vastaavuutta vakiintuneemmin käytössä olleisiin uskomusmittarin ulottuvuuksiin sekä tutkittiin tunnemittarin ja uskomusmittarin toimivuutta. Eksploratiivinen faktorianalyysi toteutettiin IMB SPSS -ohjelmiston versiolla 27 ja konfirmatorinen faktorianalyysi MPLUS 8.6 -ohjelmistolla.

Faktorianalyysien tilastollisten mallien tarkastelussa käytettyjen tunnuslukujen hyväksytyt, kirjallisuudessa usein esiintyvät raja-arvot ovat CFI:n (*comparative fit in index*) / TLI:n (*Tucker-Lewis index*) osalta ⩾ 0,95/0,90 (Leskinen, 1987; Schreiber ym., 2006), SRMR:n (*standardized root mean square residual*) eli keskimääräisen jäännöskorrelaation osalta ⩽ 0.08 (Leskinen, 1987; Schreiber ym., 2006) ja keskineliövirheen neliöjuuren RMSEA:n (*the root mean square error of approximation*) osalta ⩽ 0,07 (Halme ym., 2014; Steiger, 2007; Hooper ym., 2008).

5.5 Tulokset

Ensimmäiseksi esitetään eksploratiivisen faktorianalyysin tulokset tunnemittarin osalta. Konfirmatorisen faktorianalyysin tuloksissa erotetaan tunnemittarin, uskomusmittarin ja molemmat mittarit yhdistävän analyysin tulokset omiin osioihinsa.

5.5.1 Tunnemittari: Eksploratiivinen faktorianalyysi

Tunnemittariin valittiin alun perin teoreettisesti viisi positiivista (innostunut, kiinnostunut, onnistunut, tyytyväinen ja varma, F2.1 alfa = 0,910) ja viisi negatiivista tunnetta (vihainen, avuton, ahdistunut, pettynyt ja epävarma, F2.2 alfa = 0,879). Mittarit jakautuivat eksploratiivisessa faktorianalyysissä selkeästi negatiiviseen ja positiiviseen faktoriin (Taulukko 24). Kolmen faktorin malli osoittautui haastavaksi, sillä positiiviset tunteet jakaantuivat selkeästi kahdelle keskenään voimakkaan negatiivisesti korreloivalle faktorille eikä positiivisten ja negatiivisten tunteiden väliltä löytynyt yhteistä faktoria. Teoreettisessa mielessä tämä ei siis yksinään tuo mielekästä lisäarvoa verrattuna kahden faktorin ratkaisuun.

Tunnemittaristosta erotettiin lisäksi neljän faktorin ratkaisu. Kahden faktorin ratkaisussa löydetty positiivinen ulottuvuus jakaantui selkeästi kahteen faktoriin, kun taas negatiivinen ulottuvuus ei yksiselitteisesti jakaantunut kahteen faktoriin. Negatiivisten tunteiden osalta avuttomuuden tunne voidaan nähdä aktivoivana tunteena, joten sen sijoittaminen analyyseissä yhteen vihaisuuden tunteen kanssa oli perusteltua. Positiiviset tunteet jakautuivat teoreettisesti passivoiviin (tyytyväinen, onnistunut, varma, alpha = 0,904) ja aktivoiviin (innostunut, kiinnostunut, alpha = 0,880). Negatiivisten tunteiden osalta passivoivat tunteet (ahdistunut, pettynyt, epävarma, alpha = 0,862) eivät erottuneet aktivoivista tunteista (avuton, vihainen, alpha = 0,735) selkeästi avuttomuuden tunteen latauduttua tasaisesti aktivoivien ja passivoivien tunteiden välille. Aktivoivien ja passivoivien tunteiden välillä sekä positiivisella että negatiivisella ulottuvuudella oli suhteellisen vahvat korrelaatiot keskenään.

TAULUKKO 5.24. Faktoriratkaisut tunnemuuttujista ja niiden korrelaatiomatriisit.

	Kaksi faktoria		Kolme faktoria			Neljä faktoria			
	F2.1	F2.2	F3.1	F3.2	F3.3	F4.1	F4.2	F4.3	F4.4
kiinnostunut	0,813	0,061	0,093	-0,002	-0,874	0,197	0,105	-0,123	-0,705
innostunut	0,767	0,077	0,153	-0,013	-0,719	-0,037	-0,052	0,016	-0,940
tyytyväinen	0,856	-0,055	0,889	0,010	-0,028	0,913	0,042	-0,035	0,003
onnistunut	0,850	-0,048	0,853	0,009	-0,051	0,879	0,043	-0,037	-0,016
varma	0,767	-0,126	0,760	-0,078	-0,050	0,721	-0,189	0,132	-0,110
pettynyt	0,041	0,861	-0,105	0,826	-0,136	-0,111	0,729	0,118	-0,08
ahdistunut	0,022	0,824	-0,034	0,811	-0,043	-0,070	0,613	0,234	-0,029
epävarma	0,047	0,818	-0,034	0,800	-0,072	0,043	0,898	-0,075	0,075
avuton	-0,052	0,761	-0,019	0,768	0,049	-0,087	0,438	0,416	0,004
vihainen	-0,084	0,575	0,100	0,615	0,208	0,006	0,084	0,685	0,091
Faktorien väliset korrelaatiomatriisit									
Faktori	**F2.1**	**F2.2**	**Faktori**	**F3.1**	**F3.2**	**Faktori**	**F4.1**	**F4.2**	**F4.3**
F2.1	1		F3.1	1		F4.1	1		
F2.2	-0,428	1	F3.2	-0,450	1	F4.2	-0,396	1	
			F3.3	-0,687	0,270	F4.3	-0,339	0,636	1
						F4.4	-0,721	0,191	0,260

Koska tunnemittarin neljän ratkaisun malli on tulkinnallisesti haastava negatiivisia tunteita sisältävien faktoreiden osalta, jatkettiin tunnemittarin toimivuuden tarkastelua jättämällä analyysista pois varmuuden ja epävarmuuden tunteet. Ne poistettiin, koska nämä tunteet eivät sisälly Pekrunin akateemisten tunteiden viitekehykseen, ja koska avuttomuuden tunteen sijoittaminen aktivaation mukaan oli jatkokehityksen näkökulmasta mielenkiintoinen ongelma. Käyttämällä suurimman uskottavuuden estimointia (Taulukko 25) ei malliin muodostunut neljättä faktoria ollenkaan (yksikään väittämän latausarvo ei ylitä hyväksyttäväksi latausarvoksi asetettua raja-arvoa 0,3), vaan negatiiviset väittämät latautuivat kaikki samalle faktorille. Lisäksi analyysissä onnistuneisuuden tunteen kohdalla estimaatti yli 1 viittaa Heywood-tapaukseen (Cooperman & Waller, 2022), jolloin aineisto ei sovellu käytettäväksi analyysissä johtuen joko aineiston riittämättömyydestä tai autokorrelaatiosta muuttujien välillä. Tämän arvon ylittymä voi myös liittyä käytettyyn ortogonaaliseen rotaatioon (*direct oblimin*) tai faktorit on muodostettu liian vähillä muuttujilla (Watkins, 2018). Pääakseliratkaisu (*principal axis*) sen sijaan tuotti neljä faktoria, joille kullekin sijoittui kaksi tunnetilaa kuvaavaa väitettä (Taulukko 25).

TAULUKKO 5.25. Neljän faktorin ratkaisut faktorianalyysissä ilman varmuuden ja epävarmuuden tunnetta

	Maximum likelihood				*Principal axis*			
	F1	**F2**	**F3**	**F4**	**F1**	**F2**	**F3**	**F4**
kiinnostunut	0,022	0,014	0,912	-0,014	0,060	0,060	-0,851	-0,070
innostunut	0,013	-0,001	0,849	0,026	-0,016	-0,057	-0,879	0,038
tyytyväinen	0,585	-0,093	0,261	0,065	0,934	-0,055	-0,068	0,036
onnistunut	1,0141	0,018	-0,048	-0,027	0,824	0,043	0,028	-0,031
pettynyt	-0,018	0,876	0,030	-0,242	-0,025	0,872	0,003	-0,039
ahdistunut	-0,002	0,81	-0,001	-0,067	0,017	0,670	0,032	0,167
avuton	-0,090	0,753	0,024	0,169	-0,099	0,204	-0,070	0,630
vihainen	0,025	0,601	-0,114	0,247	0,013	-0,021	0,053	0,723
Faktorien väliset korrelaatiomatriisit								
Faktori	**1**	**2**	**3**	**4**	**Faktori**	**1**	**2**	**3**
1	1				1	1		
2	-0,398	1			2	-0,415	1	
3	0,704	-0,304	1		3	-0,756	0,207	1
4	0,011	0,111	-0,170	1	4	-0,428	0,830	0,371

Kuten taulukosta 25 huomataan, on avuttomuuden tunteen sijoittaminen aktivoivien negatiivisten tunteiden joukkoon perusteltua. Tätä sijoitusta tukee myös teoreettisesti Farchin ja kollegoiden (2018) näkemys, että avuttomuuden tunne liittyy tekemiseen tai laukaisevana tekijänä on akuutti stressireaktio. Toisaalta Filippello ja kollegat (2018) tuovat esiin avuttomuuden voivan olla opittua. Tämä voi johtaa koulutehtävien välttämiseen ja näkyä heikompina oppimistuloksina. Neljän faktorin mallista käytetään jatkossa kuitenkin kaikki väittämät sisältävää versiota, koska väittämien vähentämisestä seurasi analyyseissa tilastollisia ongelmia.

Keskiarvotarkastelujen perusteella voidaan todeta, että passivoivia tunteita esiintyy useammin kuin aktivoivia (Taulukko 26). Kaikki jakaumat ovat normaalijakautuneita (arvo välillä $-1 \leq x \leq +1$) vinouman ja huipukkuuden tarkastelun perusteella.

TAULUKKO 5.26. Tunnefaktorien kuvailevat tiedot

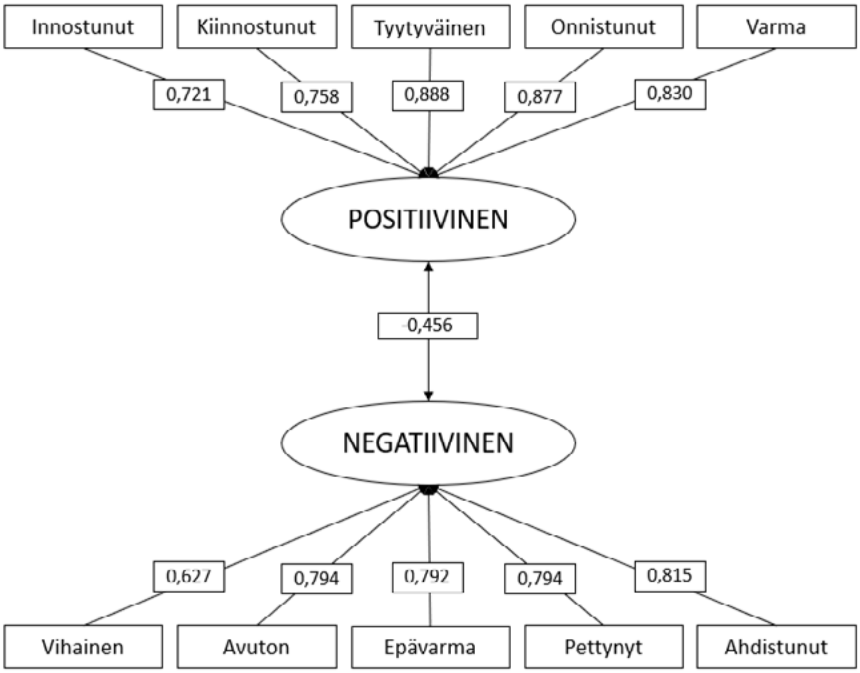
	Positiiviset tunteet	Positiiviset aktivoivat tunteet	Positiiviset passivoivat tunteet	Negatiiviset tunteet	Negatiiviset aktivoivat tunteet	Negatiiviset passivoivat tunteet
Keskiarvo	3,04	2,89	3,14	2,57	2,49	2,63
N	10805	10797	10793	10785	10776	10779
Osiot	5	2	3	5	2	3
Keskihajonta	0,91	1,03	0,95	0,99	1,06	1,08
Vinouma	-0,24	-0,05	-0,35	0,43	0,44	0,38
Huipukkuus	-0,21	-0,52	-0,22	-0,39	-0,49	-0,55
Minimi	1	1	1	1	1	1
Maksimi	5	5	5	5	5	5
Varianssi	0,82	1,05	0,90	0,98	1,12	1,16
Alfa	0,910	0,880	0,904	0,879	0,735	0,862

5.5.2 Tunnemittari: Konfirmatorisen faktorianalyysin tulokset

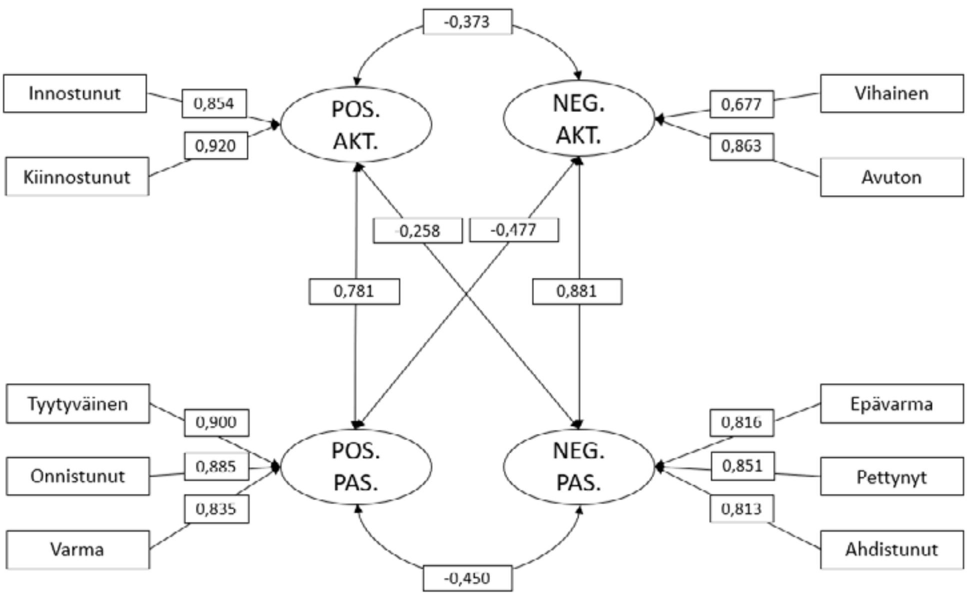
Tunnemittareista luotuja eksploratiivisia faktoriratkaisuja tutkittiin tarkemmin konfirmatorisella faktorianalyysillä. Ratkaisuista tilastollisten tunnuslukujen perusteella neljän faktorin ratkaisu toimii parhaiten (Taulukko 27). Kahden faktorin ratkaisu ei kummassakin faktorianalyysimenetelmässä osoittautunut riittäväksi, koska sekä TLI että RMSEA jäivät alle kirjallisuudessa suositeltujen raja-arvojen. Kuvissa 35a ja 35b on esitetty kuvaajat kahden ja neljän faktorin faktorimalleista, joissa faktorien välillä esitetään niiden keskinäiset korrelaatiot sekä väittämien latausarvot faktorille.

TAULUKKO 5.27. Tunnemittarin faktoriratkaisujen tilastolliset tunnusluvut

	LOGL	df	AIC	BIC	SBIC	CFI	TLI	RMSEA	SRMR	ChiSq	pChi
EFA 2F	-128958,051	39	257994,102	258276,985	258153,048	0,921	0,864	0,141	0,044	68237,395	<0,001
EFA 3F	-126833,536	47	253761,072	254101,981	253952,622	0,984	0,959	0,077	0,020	68237,395	<0,001
EFA 4F	-126354,128	54	252816,257	253207,940	253036,336	0,998	0,990	0,038	0,006	68237,395	<0,001
CFA 2F	-129287,219	31	258636,439	258861,294	258762,781	0,912	0,883	0,130	0,049	68237,395	<0,001
CFA 3F	-127239,734	33	254545,468	254784,830	254679,961	0,972	0,960	0,076	0,033	68237,395	<0,001
CFA 4F	-126804,685	37	253683,370	253951,745	253834,164	0,985	0,975	0,060	0,021	68237,395	<0,001



KUVIO 5.36a. Kahden faktorin konfirmatorisen faktorimallin standardoidut estimaatit



KUVIO 5.36b. Neljän faktorin konfirmatorisen faktorimallin standardoidut estimaatit

5.5.3 Uskomusmittari: Konfirmatorisen faktorianalyysin tulokset

Kuvailevat tiedot

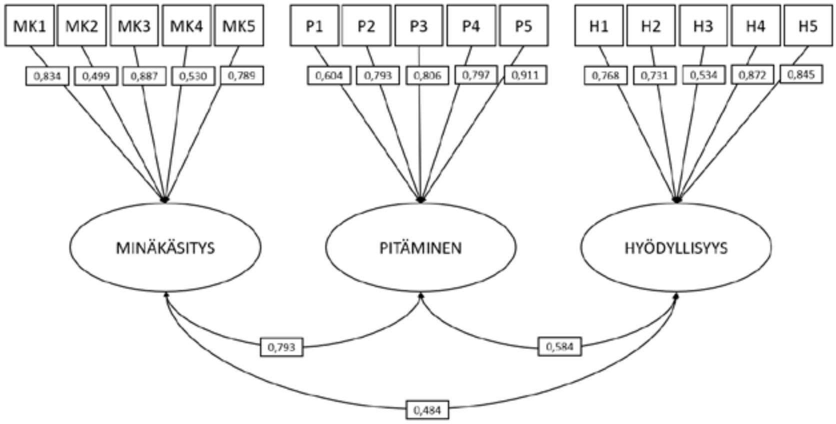
Kaikki uskomusmittarit ovat keskiarvotarkastelujen perusteella normaalijakautuneita. Pitämisuskomuksessa varianssin määrä on hieman muita uskomusmittareita suurempi. Uskomus matematiikan hyödyllisyydestä on vahvin, kun taas matematiikasta pitäminen ja minäkäsitys matematiikan osaamisesta ovat hieman heikompia, vaikkei niiden taso olekaan kokonaisuudessaan kovin matala (Taulukko 28).

TAULUKKO 5.28. Uskomusmittarien kuvailevat tiedot

	Minäkäsitys	Pitäminen	Hyöty
Keskiarvo	3,03	2,97	3,55
N	10957	10940	10951
Osiot	5	5	4
Keskihajonta	0,93	1,02	0,93
Vinouma	-0,10	0,02	-0,40
Huipukkuus	-0,50	-0,71	-0,29
Minimi	1	1	1
Maksimi	5	5	5
Varianssi	0,86	1,04	0,86
Alfa	0,830	0,855	0,810

Konfirmatorinen faktorianalyysi

Uskomusmittarin eksploratiivista ratkaisua tutkittiin/tarkasteltiin (lähemmin) konfirmatorisella faktorianalyysillä. Malli toimii kohtuullisesti, vaikkakaan RMSEA ei alita sille kirjallisuudessa hyväksytyiksi esitettyjä raja-arvoja (0,07) ja TLI jää hieman alle 0,90 raja-arvon. Tunnusluvut on esitetty kuvan 37 selitteessä. Uskomusmittarin eri faktorit korreloivat vahvasti keskenään, mikä on myös peruste analysoida niitä samanaikaisesti.



LOGL = -189744,269; df = 74; AIC = 386831,283; BIC = 387157,225; SBIC = 387014,221; CFI = 0,914; TLI = 0,894; RMSEA = 0,097; SEMR = 0,055; ChiSq = 7252,747; pChi < 0,001.

KUVIO 5.37. Uskomusmittarin konfirmatorinen faktorianalyysi

5.5.4 Tunnemittari ja uskomusmittari: Konfirmatorisen faktorianalyysin tulokset

Konfirmatorisessa faktorianalyysissä uskomukset ja tunnefaktorit pidettiin yhtenä kokonaisuutena. Tulokset on esitetty taulukossa 29, jossa tunnemittarin faktoriratkaisu on kuvattu numerolla ja tunnuksella F, kun uskomusmittarit ovat samat jokaisessa mallissa.

TAULUKKO 5.29. Uskomus- ja tunnemittareiden tilastolliset tunnusluvut konfirmatorisissa faktorianalyyseissä

	LOGL	df	AIC	BIC	SBIC	CFI	TLI	RMSEA	SRMR	ChiSq	pChi
CFA 2F	-308534,116	82	617232,233	617823,808	617563,224	0,897	0,883	0,082	0,056	158215,592	<0,001
CFA 3F	-305609,482	87	611392,965	612020,612	611744,138	0,934	0,923	0,066	0,049	158215,592	<0,001
CFA 4F	-305156,012	94	610500,024	611178,171	610879,453	0,940	0,928	0,064	0,045	158215,592	<0,001

Uskomusmittarien ja tunnemittarin yhdistävistä faktorimalleista parhaiten toimii tunnemittarin neljän faktorin faktoriratkaisu. Mikään malli ei ole huono, vaikkakin kahden faktorin ratkaisujen kohdalla pitää olla varovaisempi tulkittaessa tuloksia. Kahden faktorin malliin liittyvät tunnusluvut osoittavat, että malli ei nyt esitetyssä muodossa sovi kovin hyvin aineistossa olevaan korrelaatiorakenteeseen. Taulukossa 30 esitetään kaikkien taulukossa 29 esitettyjen konfirmatoristen faktorianalyysien mallien faktorien väliset korrelaatiot. Pääsääntöisesti korrelaatiot ovat yli 0,30 paitsi hyödyllisyysuskomuksen ja negatiivisten tunteiden osalta.

TAULUKKO 5.30. Tunnemittarin ja uskomusmittarin faktorien väliset korrelaatiot eri ratkaisuissa

2 Tunnefaktoria						
	1.	2.	3.	4.		
1. Positiivinen	1					
2. Negatiivinen	-0,461	1				
3. Minäkäsitys	0,709	-0,484	1			
4. Pitäminen	0,686	-0,356	0,792	1		
5. Hyödyllisyys	0,464	-0,180	0,484	0,583		
3 Tunnefaktoria						
	1.	2.	3.	4.	5.	
1. Positiivinen: Aktivoiva	1					
2. Positiivinen: Passivoiva	0,783	1				
3. Negatiivinen	-0,310	-0,480	1			
4. Minäkäsitys	0,577	0,707	-0,484	1		
5. Pitäminen	0,737	0,618	-0,356	0,792	1	
6. Hyödyllisyys	0,479	0,428	-0,180	0,484	0,582	
4 Tunnefaktoria						
	1.	2.	3.	4.	5.	6.
1. Positiivinen: Aktivoiva	1					
2. Positiivinen: Passivoiva	0,785	1				
3. Negatiivinen: Aktivoiva	-0,378	-0,485	1			
4. Negatiivinen: Passivoiva	-0,264	-0,456	0,882	1		
5. Minäkäsitys	0,577	0,707	-0,500	-0,456	1	
6. Pitäminen	0,737	0,620	-0,406	-0,316	0,792	1
7. Hyödyllisyys	0,479	0,430	-0,237	-0,145	0,484	0,582

5.6 Yhteenveto

Tunnemittari toimii itsenäisenä mittarina hyvin tai vähintäänkin kohtuullisesti jokaisessa tutkitussa mallissa (Taulukko 27). Eksploratiivisen faktorianalyysin (Taulukko 24) perusteella neljän faktorin malli osoittautuu tilastollisesti parhaimmaksi (Taulukko 27) mutta on kaikista eri ratkaisuista tulkinnallisesti haastavin. Avuttomuuden tunne voitaisiin sijoittaa sekä aktivoiviin että passivoiviin negatiivisiin tunteisiin. Ongelma ratkaistiin sijoittamalla avuttomuuden ja vihaisuuden tunteet samalle faktorille. Myös pääakselifaktoriratkaisu osoittaa, että valittu ratkaisu on järkevä (Taulukko 25). Ratkaisulle on lisäksi teoreettinen tuki, jonka Farchi ja kollegat (2018) ovat esittäneet. Sekä kolmen että neljän faktorin ratkaisussa esiintyy faktoreiden välillä korkeita korrelaatioita ($r > 0{,}50$), mikä on syytä huomioida niiden myöhemmässä käytössä.

Uskomusten ja tunnemittarin muodostama malli mukailee kohtuullisesti kontrolli–arvo-teoriaa (Pekrun, 2006). Vaikka uskomusten ja tunteiden väliset korrelaatiot ovat tilastollisesti merkitseviä, eivät ne ole kaikilta osiltaan vahvoja. Pekrunin viitekehyksen näkökannalta tunnemittari vaatii kehittämistyötä, koska varsinkin negatiivinen tunnemittari koostuu hyvin vahvasti passivoivista tunteista. Erityisesti tämä korostuu kolmen faktorin mallissa, jossa negatiivinen faktori on oma kokonaisuutensa, kun taas neljän faktorin mallissa vihaisuuden tunne latautuu selkeästi aktivoivien tunteiden faktorille (Taulukko 24). Skinner ja kollegat (2008) ovat esittäneet negatiivisten tunteiden jakautuvan selkeimmin aktivoiviin ja passivoiviin tunteisiin opiskeluun sitouduttaessa, joten negatiivisen tunnemittarin jatkokehitys on suositeltavaa. Positiivinen tunnemittari toimii sisäisesti johdonmukaisesti, jolloin aktivoiva ja passivoiva ulottuvuus ovat yhtenevät faktorianalyysin eri versioissa.

Analyysien myötä on ilmennyt, että arvioinnissa käytetty tunnemittari ei kovin hyvin paikallistu sen taustalla oleviin tunneteorioihin. Sitä voi osaltaan selittää sekä mittarin uutuus että sen toteutustapa. Lisäksi tunteet on käsitelty voimakkuuden sijaan määrällisenä muuttujana, jossa pohja on esiintyvyyden yleisyydessä. Tämä erilainen lähestymistapa saattaa vaikuttaa teoreettisten mallien soveltamista. Tästä seuraa myös tulkinnallisia ongelmia suhteessa eri teorioihin ja aikaisempaan tunnetutkimukseen. Tunnemittarissa heikkoutena on myös kysyttyjen tunteiden määrä. Tunnemittaria tulisi kehittää esimerkiksi Pekrunin (2002, 2017) mallin mukaisesti. Mittarin toimivuuden kannalta jokaisen ulottuvuuden eli faktorin tulisi koostua vähintään kolmesta väittämästä. Mittaria tulisi siis tarvittaessa täydentää tunnearvon ja aktivoinnin muodostaman nelikentän jokaisesta osiosta. Koska muuttujat perustuvat kyselyihin, on muistettava myös, että niiden tulkintaan vaikuttaa vastaajien ikään tai vastausmotivaatioon liittyviä tekijöitä. Tunteen yleisyyttä koko kouluajan osalta voi olla vaikea mitata, koska tunnetta ylipäätään määrittää vahvasti oppilaan viimeisin kokemus. Tästä seuraa, että päätelmiin on suhtauduttava varovaisuudella. Tunnemittaria voi käyttää tällaisenaan, mutta tutkija joutuu tekemään menetelmällisiä ratkaisuja käsitellessään tunnemittaria rakenneyhtälömallinnuksessa ja/tai muissa tilastollisissa analyyseissä.

Kokonaisuutena tunnemittari on kansallisiin arviointeihin arvokas lisä, joka mahdollistaa entistä laajempien analyysien tekemisen ja tunteiden vaikutuksen peilaamisen oppimiseen, opetukseen, opetussuunnitelmaan ja kouluun. Mittari vaatii vielä kehitystyötä, mutta jo nykyiselläänkin sitä voidaan käyttää tarkasteltaessa tunteiden, uskomusten ja osaamisen välisiä yhteyksiä.

5.7 Lähteet

Barrett, L. F., Lewis, M., & Haviland-Jones, J. M. (toim.). (2018). *Handbook of Emotions*. 4. laitos. Guilford Publications.

Camacho-Morles, J., Slemp, G. R., Pekrun, R., Loderer, K., Hou, H., & Oades, L. G. (2021). Activity achievement emotions and academic performance: A meta-analysis. *Educational Psychology Review, 33*(3), 1051–1095. https://doi.org/10.1007/s10648-020-09585-3

Cooperman, A. W., & Waller, N. G. (2022). Heywood you go away! Examining causes, effects, and treatments for Heywood cases in exploratory factor analysis. *Psychological Methods, 27*(2), 156–176. https://doi.org/10.1037/met0000384

De Corte, E., Depaepe, F., Op ,t Eynde, P., & Verschaffel, L. (2011). Students' self-regulation of emotions in mathematics: An analysis of meta-emotional knowledge and skills. *ZDM-The International Journal on Mathematics Education*, *43*, 483–495. https://doi.org/10.1007/s11858-011-0333-6

DeBellis, V. A., & Goldin, G. A. (2006). Affect and meta-affect in mathematical problem solving: A representational perspective. *Educational Studies in Mathematics*, *63*(2), 131–147. https://doi.org/10.1007/s10649-006-9026-4

Di Martino, P. & Zan, R. (2011). Attitude towards mathematics: a bridge between beliefs and emotions. *ZDM Mathematics Education, 43*, 471–482. https://doi.org/10.1007/s11858-011-0309-6

Elliot, A. J., & Pekrun, R. (2007). Emotion in the hierarchical model of approach-avoidance achievement motivation. Teoksessa P. A. Schutz & R. Pekrun (toim.), *Emotion in Education* (ss. 57–73). Elsevier Academic. 10.1016/B978-012372545-5/50005-8.

Elliott, E. S., & Dweck, C. S. (1988). Goals: An approach to motivation and achievement. *Journal of Personality and Social Psychology, 54*(1), 5–12. https://doi.org/10.1037/0022-3514.54.1.5

Farchi, M., Levy, T. B., Gershon, B. B., Hirsch-Gornemann, M. B., Whiteson, A., & Gidron, Y. (2018). The SIX Cs model for immediate cognitive psychological first aid: From helplessness to active efficient coping. *International Journal of Emergency Mental Health and Human Resilience, 20*(2), 1–12. https://doi.org/10.4172/1522-4821.1000395

Filippello, P., Harrington, N., Costa, S., Buzzai, C., & Sorrenti, L. (2018). Perceived parental psychological control and school learned helplessness: The role of frustration intolerance as a mediator factor. *School Psychology International, 39*(4), 360–377. https://doi.org/10.1177/0143034318775140

Goldin, G. A. (2000). Affective pathways and representation in mathematical problem solving. *Mathematical Thinking and Learning*, *2*(3), 209–219. https://doi.org/10.1207/S15327833MTL0203_3

Halme, N., Kanste, O., Nummi, T., & Perälä, M. L. (2014). Rakenneyhtälömallin kehittäminen ja arviointi: tutkimuksen kohteena avun antaminen lasten ja perheiden palveluissa. *Sosiaalilääketieteellinen aikakauslehti, 51*(4), 272–288. http://ojs.tsv.fi/index.php/SA/article/view/48474

Hannula, M. S. (2012). Exploring new dimensions of mathematics-related affect: Embodied and social theories. *Research in Mathematics Education, 14*(2), 137–161. https://doi.org/10.1080/14794802.2012.694281

Hannula, M. S., Bofah, E., Tuohilampi, L., & Metsämuuronen, J. (2014). A longitudinal analysis of the relationship between mathematics-related affect and achievement in Finland. Teoksessa S. Oesterle, P. Liljedahl, C. Nicol, & D. Allan (toim.), *Proceedings of the Joint Meeting 3—249 of PME 38 and PME-NA 36*, Vol. 3, 249–256.

Hooper, D., Coughlan, J., Mullen, M. R. (2008). Structural equation modelling. Guidelines for determining model fit. *Electronic Journal of Business Research Methods, 6*(1),53–60. https://academic-publishing.org/index.php/ejbrm/article/download/1224/1187

Huang, C. (2011). Achievement Goals and Achievement Emotions: A Meta-analysis. *Educational Psychology Review, 23*, 359. https://doi.org/10.1007/s10648-011-9155-x

Julin, S., & Rautopuro, J. (2016). *Läksyt tekijäänsä neuvovat. Perusopetuksen matematiikan oppimistulosten arviointi, 9. vuosiluokalla 2015.* Julkaisut 20:2016. Kansallinen koulutuksen arviointikeskus.

Leskinen, E. (1987). *Faktorianalyysi. Konfirmatoristen faktorimallien teoria ja rakentaminen*. Jyväskylän yliopiston tilastotieteen laitoksen julkaisuja 1.

Metsämuuronen, J. (2012). Challenges of the Fennema-Sherman test in the international comparisons. *International Journal of Psychological Studies*, *4*(3), 1–22. https://doi.org/10.5539/ijps.v4n3p1

Metsämuuronen, J. (2017). *Oppia ikä kaikki. Matemaattinen osaaminen toisen asteen koulutuksen lopussa 2015.* Julkaisut 1:2017. Kansallinen koulutuksen arviointikeskus. https://karvi.fi/wp-content/uploads/2017/03/KARVI_0117-1.pdf

Metsämuuronen, J. & Nousiainen, S. (2023).Yleiset menetelmäratkaisut matematiikan oppimistulosten arvioinnissa vuonna 2021. Teoksessa J. Metsämuuronen & S. Nousiainen (toim.), *Matematiikkaa COVID-19-pandemian varjossa II. Menetelmälliset ratkaisut matematiikan 9. luokan arvioinnissa keväällä 2021* (ss. 21–82). Julkaisut 5:2023. Kansallinen koulutuksen arviointikeskus.

Niemi, E. K., & Metsämuuronen, J. (toim.) (2010). *Miten matematiikan taidot kehittyvät? Matematiikan oppimistulokset peruskoulun viidennen vuosiluokan jälkeen vuonna 2008*. Koulutuksen seurantaraportit 2010:2. Opetushallitus. https://karvi.fi/wp-content/uploads/2014/09/OPH_0410.pdf

Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational psychology review*, *18*(4), 315–341. https://doi.org/10.1007/s10648-006-9029-9

Pekrun, R. (2017), Emotion and achievement during adolescence. *Child Development Perspectives, 11*(3), 215–221. https://doi.org/10.1111/cdep.12237

Pekrun, R., Elliot, A. J., & Maier, M. A. (2009). Achievement goals and achievement emotions: Testing a model of their joint relations with academic performance. *Journal of Educational Psychology, 101*, 115–135. https://doi.org/10.1037/a0013383

Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational psychologist*, *37*(2), 91–105. https://doi.org/10.1207/S15326985EP3702_4

Pekrun, R., & Linnenbrink-Garcia, L. (2012). Academic emotions and student engagement. Teoksessa S. L. Christenson, A. L. Reschly, & C. Wylie (toim.), *Handbook of research on student engagement* (ss. 259–282). Springer Science + Business Media. https://doi.org/10.1007/978-1-4614-2018-7_12

Pekrun, R., & Perry, R. P. (2014). Control-value theory of achievement emotions. Teoksessa R. Pekrun & L. Linnenbrink-Garcia, *International Handbook of Emotions in Education* (ss. 130–151). Routledge.

Salonen, R. V., Hannula, M. S. & Haataja, E. (2023). Tunteiden rooli yhdeksäsluokkalaisten matematiikan osaamisessa ja kokemuksissa matematiikan opetuksesta: Kansallisen arviointitutkimuksen näkökulma. Kansallinen koulutuksen arviointikeskus. (ilmestyy myöhemmin Julkaisuja sarjassa)

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of educational research*, *99*(6), 323–338. http://dx.doi.org/10.3200/JOER.99.6.323-3

Skinner, E., Furrer, C., Marchand, G., & Kindermann, T. (2008). Engagement and disaffection in the classroom: Part of a larger motivational dynamic? *Journal of educational psychology*, *100*(4), 765–781. https://doi.org/10.1037/a0012840

Steiger, J. H. (2007). Understanding the limitations of global fit assessment in structural equation modeling. *Personality and Individual differences*, *42*(5), 893–898. https://doi.org/10.1016/j.paid.2006.09.017

Tuohilampi, L., & Giaconi, V. (2013). Minäkäsitys, motivaatio sekä tunteet matematiikkaan liittyen: kolmasluokkalaisten vertailua Suomessa ja Chilessä. Teoksessa M. Hähkiöniemi, H. Leppäaho, P. Nieminen, & V. Viiri, *Proceedings of the 2012 Annual Conference of Finnish Mathematics and Science Education Research Association: Matematiikan ja luonnontieteiden opetuksen tutkimusseuran konferenssijulkaisu 2012* (ss. 117–128). Jyväskylän yliopisto.

Villavicencio, F. T., Bernardo, A. B. I. (2016). Beyond math anxiety: Positive emotions predict mathematics achievement, self-regulation, and self-efficacy. *The Asia-Pacific Education Researcher 25*, 415–422. https://doi.org/10.1007/s40299-015-0251-4

Watkins, M. W. (2018). Exploratory factor analysis: A guide to best practice. *Journal of Black Psychology, 44*(3), 219–246. https://doi.org/10.1177/0095798418771807

6

Yhteenvetoa ja pohdintaa vuoden 2021 matematiikan osaamisen arvioinnin menetelmällisistä ratkaisuista

Jari Metsämuuronen, Karvi
Saara Nousiainen, Karvi

6

6.1 Yhteenvetoa vuoden 2021 matematiikan arvioinnin menetelmäratkaisuista

Vuoden 2021 oppimistulosten arviointi oli järjestyksessään kahdeksas perusopetuksen päättövaiheen matematiikan osaamista kartoittava tiedonkeruu. Aiemmassa raportissa (Metsämuuronen & Nousiainen, 2021) on kuvattu keskeisiä tasa-arvokysymyksiä. Tässä raportissa kuvataan keskeisiä menetelmällisiä haasteita ja ratkaisuja.

Tiedonkeruu oli monella tavalla aiemmista matematiikan arvioinneista poikkeava:

- taustalla oli vasta valmistunut matematiikan oppimistulosarvioinnin viitekehys, jossa matematiikka jaetaan Opetussuunnitelman perusteiden mukaisesti 6 sisältöalueeseen ja 20 tavoitteeseen,
- uutena tavoitteena oli ohjelmointi ja algoritminen ajattelu, jota ei ole aiemmissa päättövaiheen arvioinneissa mitattu,
- arviointia varten valmisteltiin poikkeuksellisen paljon uusia tehtäviä, joita voidaan käyttää linkkitehtävinä myös tulevissa 3. ja 6. luokan arvioinneissa,
- käytettävissä oli syksyllä 2021 käyttöön tulleiden uusien päättöarvioinnin kriteerien valmistelutyön välituloksia ja myöhemmin julkaistu luonnos,
- arviointi oli mahdollista suorittaa aiempaa vapaammin yhden viikon aikana, koska käytössä oli useita versioita,
- ensimmäisen kerran matematiikan 9. luokan oppimistulosarviointi toteutettiin kokonaan digitaalisesti, ja noin 90 % tehtävistä oli automaattisesti pisteitettäviä,
- opettajille annettiin oppilaiden vastausten perusteella kaksi arvosanaehdotusta, joita opettaja saattoi käyttää oppilasarvioinnin tukena,

- tehtäväsarjojen rakenne ja toteutus poikkesivat aiemmasta:
  - kun aiemmin tehtäväsarjoja on ollut yksi, vuoden 2021 tiedonkeruussa käytettiin yhdeksää tehtäväsarjaa,
  - samassa koulussa ja luokassa oli käytössä useita erilaisia tehtäväsarjoja,
  - yksi tehtäväsarjoista oli käytössä kaikissa kouluissa, ja tähän sarjaan liittyvät tehtävät julkaistaan tämän raportin yhteydessä vapaaseen käyttöön,
  - kuudessa tehtäväsarjassa oli yhteisten tehtävien lisäksi painotettuna jokin matematiikan kuudesta sisältöalueesta,
  - osa opiskelijoista osallistui arvioinnin toiseen vaiheeseen, jossa kartoitettiin, minkälaisia matemaattisiin oppimisvaikeuksiin, numerotaidottomuuteen eli dyskalkuliaan liittyviä piirteitä heikosti menestyneillä oppilailla oli,
  - osa oppilaista osallistui ylöspäin eriytettyyn testiin, jossa kartoitettiin, kuinka pitkälle matematiikan taidoiltaan parhaat osaajat ovat edenneet.

Kaikkiaan tiedonkeruu on monipuolisempi ja laajempi kuin aiemmat matematiikan päättövaiheen arvioinnit. Myös otoskooltaan aineisto on selvästi laajempi kuin aiemmin. Tämä johtui otantatavasta, jossa mukaan valitun koulutuksen järjestäjän kouluista valittiin edustava otos aineistoon.

Käytetyt osaamisen mittarit ovat osuvia (valideja) ja tarkkoja (reliaabeleita) tuottamaan oppilaita erottelevaa tietoa matematiikan osaamisesta ja sen muuttumisesta viimeisen 23 vuoden aikana.

6.2 Pohdintaa tulevia arviointeja silmällä pitäen

Tulevissa matematiikan 9. luokan arvioinneissa ei ole välttämättä tarpeen tehdä yhtä kattavaa tiedonkeruuta kuin tässä arvioinnissa. Nyt saatu tieto toimii hyvänä perustana seuraaville mittauksille. Arviointia varten valmisteltiin poikkeuksellisen paljon tehtäviä, jotka ovat käytettävissä Karvin matematiikan ja äidinkielen pitkittäisarvioinnissa 6. ja 9. luokan mittauksissa vuosina 2024 ja 2027.

Mittaus tehtiin Turun yliopiston Oppimisanalytiikan tutkimusinstituutin ViLLE-järjestelmällä, koska Karvin oma arviointialusta ei tukenut kaikki matematiikan tehtävätyyppejä ja alusta oli varattu muihin, samaan aikaan tehtäviin mittauksiin. Samanaikaisesti tämän raportin kirjoittamisen kanssa on kehitteillä uusi digitaalinen arviointijärjestelmä, jonka 1. versio on tarkoitus olla käytössä ensimmäisen kerran vuonna 2024 matematiikan ja äidinkielen 6. luokan tiedonkeruussa. Vaikka uusi järjestelmä edustaa alan viimeisintä kehitystä, sen ensimmäisissä versioissa ei todennäköisesti ole vielä kaikkia matematiikan arvioinnissa tarvittavia, digitaaliseen testaukseen parhaimmillaan tai luovimmillaan kuuluvia osiotyyppejä tai ominaisuuksia. Näitä kehitellään tulevina vuosina, mikäli kehitystyöhön saadaan rahoitus. Tämänkaltaisia tulevaisuuden ominaisuuksia saattavat olla diagnostisiin tarkasteluihin ja tekoälyyn liittyvät tehtävien jatkoanalyysit tai adaptiivisen testauksen mahdollisuudet.

6.3 Suositukset kootusti

Lukujen 3 ja 4 perusteella annettavat suositukset on koottu tähän tiivistetysti.

1. Matematiikan taidoiltaan heikoille oppilaille tulee kohdentaa aiempaa intensiivisempää oppimisen tukea mahdollisimman varhaisessa vaiheessa, jotta voidaan varmistaa, ettei oppilaalle jää puutteita taitoihin, mikä puolestaan edelleen vaikuttaa uuden oppimiseen. Tukiopetusresursseja tulee suunnata jo alkuopetuksesta lähtien riittävän tehokkaasti, ennen kuin osaamisessa ilmeneviä puutteita on vaikea korjata ylempien luokkien aikana.
2. Sellaisia oppilaita, jotka potentiaalisesti hyötyisivät intensiivisemmästä yleisestä tuesta, on syytä tietoisesti ja systemaattisesti kartoittaa, vaikka he eivät apua aktiivisesti kysykään. Oppilasryhmiä, jotka todennäköisesti hyötyisivät intensiivisemmästä yleisestä tuesta ovat esimerkiksi arvosanan 5 ja 6 saaneet oppilaat ja ne oppilaat, jotka eivät viihdy koulussa.
3. Matematiikan taidoiltaan heikkojen oppilaiden osalta on tarpeen kartoittaa mahdollinen numerotaidottomuus eli dyskalkulia, jotta heille voidaan tarjota oikea-aikaista ja kohdennettua tukea.
4. Ensimmäisillä vuosiluokilla on syytä tukea oppilaiden sanavaraston kehittymistä ja taitoa työskennellä pitkäjänteisesti ongelmaratkaisutehtävien parissa.
5. Kansallista rahoitusta tarvitaan kansallisten arviointialustojen pitkäjänteiseen kehittämiseen. Uudet teknologiat kuten tekoäly ovat arvioinneissa mahdollisuus. Niiden ymmärtäminen ja käyttöönotto vaatii syventymistä ja pitkäaikaista rahoitusta. Arvioinnin kansallisessa digitalisoinnissa tarvitaan myös ICT-asiantuntijoita.

7

Lähteet koko raporttiin

7

AERA, APA, & NCME (2014). *Standards for Educational and Psychological Testing*. American Educational Research Association, American Psychological Association, & National Council of Measurement in Education.

Altinok, N., Diebolt, C., & Demeulemeester, J.-L. (2014). A new international database on education quality: 1965–2010. *Applied Economics, 46*(11), 1212–1247. 10.1080/00036846.2013.868592

Altinok, N., Angrist, N., Patrinos, H. A. (2018). *Global data set on education quality (1965–2015)*. Policy Research Working Paper; No. 8314. World Bank. http://hdl.handle.net/10986/29281

Anderson, L. W., & Krathwohl, D. R. (toim.) (2000). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Complete edition. Pearson.

Armor, D. (1973). Theta reliability and factor scaling. *Sociological Methodology, 5*, 17–50. https://doi.org/10.2307/270831

AVI (2022). *Perusopetuksen ja varhaiskasvatuksen tilannekuva*. Aluehallintovirasto. https://avi.fi/tietoa-meista/toimintamme/tuotamme-tietoa/perusopetuksen-ja-varhaiskasvatuksen-tilannekuva

Barrett, L. F., Lewis, M., & Haviland-Jones, J. M. (toim.). (2018). *Handbook of Emotions*. 4. laitos. Guilford Publications.

Bentler, P. M. (2009). Alpha, dimension-free, and model-based internal consistency reliability. *Psychometrika, 74*(1), 137–143. https://doi.org/10.1007/s11336-008-9100-1

Blikstad-Balas, M., Roe, A., Pedersen Dalland, C., & Klette, K. (2022). Homeschooling in Norway during the pandemic. Digital learning with unequal access to qualified help at home and unequal learning opportunities provided by the school. Teoksessa F. M. Reimers (toim.), *Primary and secondary education during Covid-19. Disruptions to educational opportunity during a pandemic* (ss. 177–212). Springer. https://link.springer.com/content/pdf/10.1007%2F978-3-030-81500-4.pdf

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain*. David McKay Company.

Bravais, A. (1844). Analyse Mathematique. Sur les probabilités des erreurs de situation d'un point. (Mathematical analysis. Of the probabilities of the point errors). *Mémoires présentés par divers savants à l'Académie Royale des Siences de l'Institut de France* (*Memoirs presented by various scholars to the Royal Academy of Sciences of the Institute of France*), *9*, 255–332. https://books.google.fi/books?id=7g_hAQAACAAJ&redir_esc=y

Brown, A. & Croudace, T. (2015). Scoring and estimating score precision using multidimensional IRT. Teoksessa S. P. Reise, & D. A. Revicki (toim.), *Handbook of Item Response Theory Modeling: Applications to Typical Performance Assessment* (ss. 307–333). Routledge/Taylor & Francis Group.

Brown, G.T.L. (2019). Technologies and Infrastructure: Costs and Obstacles in Developing Large-Scale Computer-Based Testing. *Education Inquiry, 10*(1), 4–20. https://doi.org/10.1080/20004508.2018.1529528

Bryant, M. (2021). Is Facebook leading us on a journey to the metaverse? *The Guardian*. 26.9.2021. https://www.theguardian.com/technology/2021/sep/26/is-facebook-leading-us-on-a-journey-to-the-metaverse

Butler, A.C., Karpicke, J.D. & Roediger, H.L. 3rd (2007). The effect of type and timing of feedback on learning from multiple-choice tests. *Journal of Experimental Psychology, Applied, 13*(4), 273–281. http://www.ncbi.nlm.nih.gov/pubmed/18194050

Butler, A.C., Karpicke, J.D. & Roediger, H.L. 3rd (2008). Correcting a metacognitive error: feedback increases retention of low-confidence correct responses. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(4), 918–928. https://doi.org/10.1037/0278-7393.34.4.918

Butler, A.C. & Roediger, H.L. 3rd (2008). Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. *Memory & Cognition 36*, 604–616. http://www.ncbi.nlm.nih.gov/pubmed/18491500

Camacho-Morles, J., Slemp, G. R., Pekrun, R., Loderer, K., Hou, H., & Oades, L. G. (2021). Activity achievement emotions and academic performance: A meta-analysis. *Educational Psychology Review, 33*(3), 1051–1095. https://doi.org/10.1007/s10648-020-09585-3

Cheng, Y., Yuan, K.-H., & Liu, C. (2012). Comparison of reliability measures under factor analysis and item response theory. *Educational and Psychological Measurement, 72*(1), 52–67. https://doi.org/10.1177/0013164411407315

Clariana, R. & Wallace, P. (2002). Paper–based versus computer–based assessment: key factors associated with the test mode effect. *British Journal of Educational Technology, 33*(5), 593–602. https://doi.org/10.1111/1467-8535.00294

Cohen, J. (1969). *Statistical power analysis for the behavioral sciences*. 1. laitos. Academic press.

Cohen, J. (1973). Eta-squared and partial eta-squared in fixed factor ANOVA designs. *Educational and Psychological Measurement, 33*(1), 107–112. https://doi.org/10.1177/001316447303300111

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. 2. laitos. Erlbaum.

Cooperman, A. W., & Waller, N. G. (2022). Heywood you go away! Examining causes, effects, and treatments for Heywood cases in exploratory factor analysis. *Psychological Methods, 27*(2), 156–176. https://doi.org/10.1037/met0000384

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3) 297–334. https://doi.org/10.1007/BF02310555

Cummings, J. J., & Bailenson, J. N. (2016). How immersive is enough? A meta-analysis of the effect of immersive technology on user presence. *Media Psychology, 19*(2), 272–309. https://doi.org/10.1080/15213269.2015.1015740

De Corte, E., Depaepe, F., Op ,t Eynde, P., & Verschaffel, L. (2011). Students' self-regulation of emotions in mathematics: An analysis of meta-emotional knowledge and skills. *ZDM-The International Journal on Mathematics Education*, *43*, 483–495. https://doi.org/10.1007/s11858-011-0333-6

DeBellis, V. A., & Goldin, G. A. (2006). Affect and meta-affect in mathematical problem solving: A representational perspective. *Educational Studies in Mathematics*, *63*(2), 131–147. https://doi.org/10.1007/s10649-006-9026-4

Dikli, S. (2006). An overview of automated scoring of essays. *Journal of Technology, Learning, and Assessment, 5*(1). https://ejournals.bc.edu/index.php/jtla/article/view/1640

Di Martino, P. & Zan, R. (2011). Attitude towards mathematics: a bridge between beliefs and emotions. *ZDM Mathematics Education, 43*, 471–482. https://doi.org/10.1007/s11858-011-0309-6

Dorans, N. J., Moses T. P., & Eignor, D. R. (2010). *Principles and practices of test score equating*. ETS RR-10-29. ETS, Princeton, New Jersey. https://onlinelibrary.wiley.com/doi/pdf/10.1002/j.2333-8504.2010.tb02236.x

Eglash, R., Bennet, A., Babbitt, W., Lachney, M., Reinhardt, M. & Hammond-Sowah, D. (2020). Decolonizing posthumanism: Indigenous material agency in generative STEM. *British Journal of Educational Technology, 51*(4), 1334–1353. https://doi.org/10.1111/bjet.12963

Elliot, A. J., & Pekrun, R. (2007). Emotion in the hierarchical model of approach-avoidance achievement motivation. Teoksessa P. A. Schutz & R. Pekrun (toim.), *Emotion in Education* (ss. 57–73). Elsevier Academic. 10.1016/B978-012372545-5/50005-8.

Elliott, E. S., & Dweck, C. S. (1988). Goals: An approach to motivation and achievement. *Journal of Personality and Social Psychology, 54*(1), 5–12. https://doi.org/10.1037/0022-3514.54.1.5

Falk, C. F., & Savalei, V. (2011). The relationship between unstandardized and standardized alpha, true reliability, and the underlying measurement model. *Journal of Personality Assessment, 93*(5), 445–53. https://doi.org/10.1080/00223891.2011.594129

Farchi, M., Levy, T. B., Gershon, B. B., Hirsch-Gornemann, M. B., Whiteson, A., & Gidron, Y. (2018). The SIX Cs model for immediate cognitive psychological first aid: From helplessness to active efficient coping. *International Journal of Emergency Mental Health and Human Resilience, 20*(2), 1–12. https://doi.org/10.4172/1522-4821.1000395

Fennema, E., & Sherman, J. A. (1976). Fennema-Sherman mathematics attitude scales: Instruments designed to measure attitudes toward the learning of mathematics by females and males. *Journal for Research in Mathematics Education, 7(*5), 324–326. https://doi.org/10.2307/748467

Feurstein, R., Rand, Y., Hoffman, M.B. (1979). *The dynamic assessment of retarded performers: The learning potential assessment device: Theory, instruments, and techniques*. University Park Press.

Filippello, P., Harrington, N., Costa, S., Buzzai, C., & Sorrenti, L. (2018). Perceived parental psychological control and school learned helplessness: The role of frustration intolerance as a mediator factor. *School Psychology International, 39*(4), 360–377. https://doi.org/10.1177/0143034318775140

Fishbein, B., Martin, M.O., Mullis, I.V.S, & Foy, P. (2018). The TIMSS 2019 Item Equivalence Study: Examining Mode Effects for Computer-Based Assessment and Implications for Measuring Trends. *Large-scale Assessments in Education, 6*, Article 11. https://doi.org/10.1186/s40536-018-0064-z

Flora, D. B. (2020). Your coefficient alpha is probably wrong, but which coefficient omega is right? A tutorial on using R to obtain better reliability estimates. *Advances in Methods and Practices in Psychological Science, 3*(4), 484–501. https://doi.org/10.1177/2515245920951747

Flynn, J. R. (1984). The mean of IQ of Americans: Massive gains 1932 to 1978. *Psychological Bulletin, 95*(1), 29–51. https://doi.org/10.1037/0033-2909.95.1.29

Flynn, J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin, 101*, 171–191. https://doi.org/10.1037/0033-2909.101.2.171

Flynn, J. R. (2009a). *What is intelligence? Beyond the Flynn effect*. 2. laitos. Cambridge University Press.

Flynn, J. R. (2009b). Requiem for nutrition as the cause of IQ gains: Raven's gains in Britain 1938–2008. *Economics and Human Biology,* 7(1), 18–27. https://doi.org/10.1016/j.ehb.2009.01.009

Flynn, J. R., & Shayer, M. (2018). IQ decline and Piaget: Does the rot start at the top? *Intelligence, 66*, 112–121. https://doi.org/10.1016/j.intell.2017.11.010

FUNA (2019). *FUNA—Toiminnallisen laskutaidon arviointi*. Oppimisanalytiikan keskus, Turun yliopisto. http://oppimisanalytiikka.fi/funa

Gadermann A. M., Guhn, M., & Zumbo, B. D. (2012) Estimating ordinal reliability for Likert-type and ordinal item response data: A conceptual, empirical, and practical guide. *Practical Assessment, Research, and Evaluation, 17*(3), 1–13. https://doi.org/10.7275/n560-j767

Goldin, G. A. (2000). Affective pathways and representation in mathematical problem solving. *Mathematical Thinking and Learning, 2*(3), 209–219. https://doi.org/10.1207/S15327833MTL0203_3

Goman, J., Huusko, M., Isoaho, K., Lehikko, A., Metsämuuronen, J., Rumpu N., Seppälä, H., Venäläinen, S., & Åkerlund, C. (2021). *Poikkeuksellisten opetusjärjestelyjen vaikutukset tasa-arvon ja yhden vertaisuuden toteutumiseen eri koulutusaloilla. Osa III: Kansallisen arvioinnin yhteenveto ja suositukset*. Julkaisut 8:2021. Kansallinen koulutuksen arviointikeskus. https://karvi.fi/wp-content/uploads/2021/04/KARVI_0821.pdf

Goodman, L. A., & Kruskal, W. H. (1954). Measures of association for cross classifications. *Journal of the American Statistical Association, 49*(268), 732–764. https://doi.org/10.1080/01621459.1954.10501231

Govindrajan, V. & Srivastava, A. (2020). What the shift to virtual learning could mean for the future of higher Ed. *Harvard Business Review*, March 31, 2020. https://hbr.org/2020/03/what-the-shift-to-virtual-learning-could-mean-for-the-future-of-higher-ed

Greving, S. & Richter, T. (2018). Examining the Testing Effect in University Teaching: Retrievability and Question Format Matter. *Frontiers in Psychology,* 9:2412. https://doi.org/10.3389/fpsyg.2018.02412

Grigorenko E.L. & Sternberg R.J. (1998). Dynamic testing. *Psychological Bulletin, 124*(1), 75–111. https://doi.org/10.1037/0033-2909.124.1.75

Halme, N., Kanste, O., Nummi, T., & Perälä, M. L. (2014). Rakenneyhtälömallin kehittäminen ja arviointi: tutkimuksen kohteena avun antaminen lasten ja perheiden palveluissa. *Sosiaalilääketieteellinen aikakauslehti, 51*(4), 272–288. http://ojs.tsv.fi/index.php/SA/article/view/48474

Hamhuis, E., Glas, C., & Meelissen, M. (2020). Tablet assessment in primary education: Are there performance differences between TIMSS' paper-and-pencil test and tablet test among Dutch grade-four students? *British Journal of Educational Technology, 51*(6), 2340–2358. https://doi.org/10.1111/bjet.12914

Hannu, J., Saikkonen, T., Häkkinen, J. Karttunen, J., & Moilanen, M. (2010). Enabling remote testing: Embedded test controller and mixed-signal test architecture. *Journal of Electronic Testing, 26*, 641–658. https://doi.org/10.1007/s10836-010-5175-6

Hannula, M. S. (2012). Exploring new dimensions of mathematics-related affect: Embodied and social theories. *Research in Mathematics Education, 14*(2), 137–161. https://doi.org/10.1080/14794802.2012.694281

Hannula, M. S., Bofah, E., Tuohilampi, L., & Metsämuuronen, J. (2014). A longitudinal analysis of the relationship between mathematics-related affect and achievement in Finland. Teoksessa S. Oesterle, P. Liljedahl, C. Nicol, & D. Allan (toim.), *Proceedings of the Joint Meeting 3—249 of PME 38 and PME-NA 36*, Vol. 3, 249–256.

Hansen, A. (2011). *What does dynamic testing entail?* Translated from Norwegian for the Daffodil project, 2011. Lifelong Learning Programme. EU. https://static1.squarespace.com/static/5c2c66dd297114a6e7b41163/t/5c4b79fc562fa7de1aa49e18/1548450304544/Article+on+Dynamic+Assessment.pdf

Harjunen, E. & Rautopuro J. (2015). *Kielenkäytön ajattelua ja ajattelun kielentämistä. Äidinkielen ja kirjallisuuden oppimistulokset perusopetuksen päättövaiheessa 2014: keskiössä kielentuntemus ja kirjoittaminen.* Julkaisut 2015:8. Kansallinen koulutuksen arviointikeskus. https://karvi.fi/wp-content/uploads/2015/04/KARVI_08151.pdf

Hautamäki, J., Kupiainen, S., Marjanen, J., Vainikainen, M-P., & Hotulainen, R. (2013). *Oppimaan oppiminen peruskoulun päättövaiheessa: Tilanne vuonna 2012 ja muutos vuodesta 2001.* Tutkimuksia 347.). Helsingin yliopisto, opettajankoulutuslaitos.

Haywood, H.C. (1992). Interactive assessment: A special issue. *Journal of Special Education, 26*, 233–234. https://doi.org/10.1177/002246699202600301

Haywood, H.C. (1997). Interactive assessment. Teoksessa R. L. Tayler (toim.), *Assessment of individuals with mental retardation*. Singular Publishing Group.

Heikkilä, T. (2021). Vuoden 2021 tärkeimmät megatrendit. *Sijoittaja.fi*. https://www.sijoittaja.fi/258979/vuoden-2021-tarkeimmat-megatrendit/

Heikkurinen, P. (2014). Kestävyyden käsitteen ulottuvuudet. *Tieteessä tapahtuu, 32*(4), 10–16. http://ojs.tsv.fi/index.php/tt/article/view/46149

Heise, D., & Bohrnstedt, G. (1970). Validity, invalidity, and reliability. *Sociological Methodology, 2*, 104–129. https://doi.org/10.2307/270785

Hellgren, J. & Marjanen, J. (2020). *Svenska och litteratur i slutet av årkurs 9. Resultat av en utvärdering av lärresultat våren 2019*. Publikationer 18:2020. Nationella Centret för Utbildningsutvärdering. https://karvi.fi/wp-content/uploads/2020/10/KARVI_1820.pdf

Holstein, K. (2017). *The Lumilo Project.* Haettu osoitteesta https://medium.com/lumilo/the-lumilo-project-cda82726f9ec (24.2.2023).

Hooper, D., Coughlan, J., Mullen, M. R. (2008). Structural equation modelling. Guidelines for determining model fit. *Electronic Journal of Business Research Methods, 6*(1),53–60. https://academic-publishing.org/index.php/ejbrm/article/download/1224/1187

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

Huang, C. (2011). Achievement Goals and Achievement Emotions: A Meta-analysis. *Educational Psychology Review, 23*, 359. https://doi.org/10.1007/s10648-011-9155-x

Huisman T (2006). *Luen, kirjoitan ja ratkaisen. Peruskoulun kolmasluokkalaisten oppimistulokset äidinkielessä ja kirjallisuudessa sekä matematiikassa.* Oppimistulosten arviointi 7/2006. Opetushallitus. https://karvi.fi/app/uploads/2014/09/OPH_0906.pdf

Huisman T & Silverström C (2006). *Läsa, skriva, räkna. Utvärdering av inlärningsresultat i modersmål och litteratur samt matematik i årskurs 3. Utvärdering av inlärningsresultat 8/2006.* Utbildningsstyrelsen. https://karvi.fi/app/uploads/2014/10/OPH_1006.pdf

Härmälä, M., Huhtanen, M., Puukko, M. & Marjanen, J. (2019). *A-englannin oppimistulokset 7. vuosiluokan alussa 2018.* Julkaisut 13:2019. Kansallinen koulutuksen arviointikeskus. https://karvi.fi/app/uploads/2019/05/KARVI_1319.pdf

Härmälä, M. & Marjanen, J. (2022). *Englantia koronapandemian aikaan. A-englannin oppimistulokset 9. vuosiluokan lopussa 2021.* Julkaisut 22:2022. Kansallinen koulutuksen arviointikeskus. https://karvi.fi/wp-content/uploads/2022/09/KARVI_2222.pdf

IELTS (2007). Weir, C., O'Sullivan, B., Yan, J. & Bax, S. (2007). Does the computer make a difference? The reaction of candidates to a computer-based versus a traditional hand-written form of the IELTS Writing component: effects and impact. *IELTS Research Reports, Volume* 7. British Council and IELTS Australia Pvt Limited. https://www.ielts.org/-/media/research-reports/ielts_rr_volume07_report6.ashx

IELTS (2013). *Biometric systems further tighten IELTS test security*. https://www.ielts.org/news/2013/biometric-systems-further-tighten-ielts-test-security

IELTS (2022). *IELTS Identity Verification*. https://www.idp.com/hongkong/ielts-hk/ielts-registration/identity-verification/?lang=en

Jacobsen, J (2020). *How to conduct remote user tests successfully.* TestingTime. https://www.testingtime.com/en/blog/how-to-conduct-remote-user-tests-successfully/

Jakku-Sihvonen, R. 1997. Johdanto. Teoksessa R. Jakku-Sihvonen (toim.), *Onnistuuko oppiminen—oppimistuloksien ja opetuksen laadun arviointiperusteita peruskoulussa ja lukiossa*. Arviointi 3/1997. Opetushallitus.

Jauhiainen, T., Lennes, M., & Marttila, T. (toim.) (2019), *Suomenkielisen tekoälyn kehittämisohjelma—esiselvitys*. Valtion kehitysyhtiö. https://vake.fi/wp-content/uploads/Vake-suomenkielisen-teko%C3%A4lyn-kehitt%C3%A4minen-esiselvitys-2019.pdf

Jerrim, J., Micklewright, J., Heine, J.-H., Salzer, C., & McKeown, C. (2018). PISA 2015: how big is the 'mode effect' and what has been done about it? *Oxford Review of Education, 44*(4), 476–493. 10.1080/03054985.2018.1430025

Johns Hopkins (2021). Johns Hopkins performs its first augmented reality surgeries in patients. *NeuroLogic*, Winter 2021. https://www.hopkinsmedicine.org/news/articles/johns-hopkins-performs-its-first-augmented-reality-surgeries-in-patients

Julin, S. & Rautopuro, J (2016). *Läksyt tekijäänsä neuvovat. Perusopetuksen matematiikan oppimistulosten arviointi 9. vuosiluokalla 2015*. Julkaisut 20:2016. Kansallinen koulutuksen arviointikeskus. https://karvi.fi/app/uploads/2016/04/KARVI_2016.pdf

Kaarakainen, M.-T., Kivinen, O., & Hutri, H. (2015). Pelit ja pelaaminen sosiaalisena oppimisympäristönä. Teoksessa R. Koskimaa, J. Suominen, F. Mäyrä, J. T. Harviainen, U. Friman, & J. Arjoranta (toim.), *Vuosikirja2015*. Suomen Pelitutkimuksen Seura. https://www.pelitutkimus.fi/vuosikirja2015/artikkeli-pelit-ja-pelaaminen-sosiaalisena-oppimisymparistona

Kaiser, H. F., & Caffrey, J. (1965). Alpha factor analysis. *Psychometrika*, *30*, 1–14. https://doi.org/10.1007/BF02289743

Kalenius, A. (2023). *Sivistyskatsaus 2023*. Opetus- ja kulttuuriministeriön julkaisuja 2023:3. https://julkaisut.valtioneuvosto.fi/handle/10024/164564

Karvi. (2020). *Koulutuksen arviointisuunnitelma 2020–2023*. Kansallinen koulutuksen arviointikeskus. https://karvi.fi/app/uploads/2020/04/Koulutuksen_arviointisuunnitelma_2020-2023.pdf

Kass, G. (1980). An exploratory technique for investigating large quantities of categorical data. *Applied Statistics, 29*(2), 119–127. https://doi.org/10.2307/2986296

Katajamäki, H. 2011. Maailmasta on kysymys. Sanomalehtiyliopisto kevät 2011. Vaasan yliopisto. Julkaisu n.o 34. *Kestävä kehitys*. http://www.uva.fi/materiaali/pdf/isbn_978-952-476-400-1.pdf

Kauppinen, M. & Marjanen, J. (2020). *Millaista on yhdeksäsluokkalaisten kielellinen osaaminen?—Suomen kielen ja kirjallisuuden oppimistulokset perusopetuksen päättövaiheessa 2019.* Julkaisuja 13:2020. Kansallinen koulutuksen arviointikeskus. https://karvi.fi/wp-content/uploads/2020/08/KARVI_1320.pdf

Kerner, S. M., & Gillis, A. S. (2022). *Web 3.0 (Web3)*. Definition. https://www.techtarget.com/whatis/definition/Web-30

Kim, S., & Feldt, L. S. (2010). The estimation of the IRT reliability coefficient and its lower and upper bounds, with comparisons to CTT reliability statistics. *Asia Pacific Education Review, 11*, 179–188. https://doi.org/10.1007/s12564-009-9062-8

Kitchin, R. (2013). Big data and human geography: Opportunities, challenges and risks. *Dialogues in Human Geography, 3*(3), 262–267. https://doi.org/10.1177/2043820613513388

Kolen, M. J., & Brennan, R. L. (2004). *Test equating, linking, and scaling: Methods and practices*. 2. laitos. Springer-Verlag.

Kreiner, S. (2021). Haastattelu 6.9.2021. *Folkeskolen*. Riise, A.B., "Svend Kreiner: Derfor var de adaptive test ikke så god en ide, som vi troede". https://www.folkeskolen.dk/evaluering-ledelse-matematik/svend-kreiner-derfor-var-de-adaptive-test-ikke-sa-god-en-ide-som-vi-troede/1361560

Krishnan, K. (2020). *Our education system is losing relevance. Here's how to unleash its potential.* World Economic Forum. Apr. 13, 2020. https://www.weforum.org/agenda/2020/04/our-education-system-is-losing-relevance-heres-how-to-update-it/?utm_source=sfmc&utm_medium=email&utm_campaign=2716680_Agenda_weekly-17April2020&utm_term=&emailType=Newsletter

Kuuluvainen, V., Virtanen, I., Rikkonen, L., & Isotalus, P. (2021). Testing an Immersive Virtual Environment for Decreasing Intergroup Anxiety among University Students: An Interpersonal Perspective. *International Journal of Emerging Technologies in Learning (iJET)*, *16*(16), 202–220. https://doi.org/10.3991/ijet.v16i16.19673

Kuusi, O. (1999). Teknologien kehityksen heikot signaalit. *FUTURA* 2/1999, 8–15.

Lavonen, J. & Salmera-Aro, K. (2022). Experiences of moving quickly to distance teaching and learning at all levels of education in Finland. Teoksessa F. M. Reimers (toim.), *Primary and secondary education during Covid-19. Disruptions to educational opportunity during a pandemic* (ss. 105–124). Springer. https://link.springer.com/content/pdf/10.1007%2F978-3-030-81500-4.pdf

Leskinen, E. (1987). *Faktorianalyysi. Konfirmatoristen faktorimallien teoria ja rakentaminen*. Jyväskylän yliopiston tilastotieteen laitoksen julkaisuja 1.

Linacre, J. M. (2000). Computer-adaptive testing: A methodology whose time has come. Teoksessa S. Chae, U. Kang, E. Jeon, & J.M. Linacre, *Development of Computerized Middle School Achievement Tests*. MESA Research Memorandum. Komesa Press. https://www.rasch.org/memo69.pdf

Linn, R. L. (1993). Linking results of distinct assessment. *Applied Measurement in Education*, *6*(1), 83–102. http://dx.doi.org/10.1207/s15324818ame0601_5

Litts, B.K., Searle, K.A., Brayboy, B.M.J. & Kafai, Y.B. (2020). Computing for all?: Examining critical biases in computational tools for learning. *British Journal of Educational Technology, 52*(2), 842–857. https://doi.org/10.1111/bjet.13059

Livingston, S. A., & Dorans, N. J. (2004). *A graphical approach to item analysis*. (Research Report No. RR-04-10). Educational Testing Service. https://doi.org/10.1002/j.2333-8504.2004.tb01937.x

Lord, F. M. (1952). The relationship of the reliability of multiple-choice test to the distribution of item difficulties. *Psychometrika, 17*(2), 181–194. http://dx.doi.org/10.1007/BF02288781

Lord, F. M. (1958). Some relations between Guttman's principal component scale analysis and other psychometric theory. *Psychometrika*, *23*(4), 291–296. http://dx.doi.org/10.1002/j.2333-8504.1957.tb00073.x

Lord, F. M., Novick, M. R., & Birnbaum, A. (1968). *Statistical theories of mental test scores*. Addison-Wesley.

Mannermaa, M. (1991). *Evolutionaarinen tulevaisuustutkimus. Tulevaisuustutkimuksen paradigmojen ja niiden metodologisten ominaisuuksien tarkastelua*. Tulevaisuuden tutkimuksen seura, Acta FUTURA Fennica n:o 2. VAPK-kustannus, Helsinki.

Mazzeo, J. (2016). Large-scale computer-based tests in the United States: Expectations, experiences, successes and some challenges for the future. Presentation on conference *Future Tests and Test Environments*, June 16, 2016, Umeå, Sweden.

McDonald, R. P. (1970). Theoretical canonical foundations of principal factor analysis, canonical factor analysis, and alpha factor analysis. *British Journal of Mathematical and Statistical Psychology, 23*, 1–21. http://dx.doi.org/10.1111/j.2044-8317.1970.tb00432.x

McDonald, R. P. (1999). *Test Theory: A Unified Treatment*. Lawrence Erlbaum Associates.

Meristö, T. (1993). Skenaariotyöskentely strategisessa johtamisessa. Teoksessa M. Vapaavuori (toim.), *Miten tutkimme tulevaisuutta?* (ss. 215–221) Tulevaisuuden tutkimuksen seura. Acta *FUTURA* Fennica n:o 5. Tulevaisuustutkimuksen keskus.

Metcalfe J., Kornell, N. & Finn, B. (2009). Delayed versus immediate feedback in children's and adult's vocabulary learning. *Memory & Cognition*, *37***,** 1077–1087. http://www.ncbi.nlm.nih.gov/pubmed/19933453

Metsämuuronen, J. (2001a). *Sosiaali- ja terveysalan tulevaisuutta etsimässä*. International Methelp Oy.

Metsämuuronen, J. (2001b). Uuden vuoisituhannen haasteet sosiaali- ja terveysalalla. Teoksessa J. Metsämuuronen, *Sosiaali- ja terveysalan tulevaisuutta etsimässä* (ss. 162–175). International Methelp Oy.

Metsämuuronen, J. (2009a). *Metodit arvioinnin apuna. Perusopetuksen oppimistulosarviointien ja -seurantojen menetelmäratkaisut Opetushallituksessa*. Oppimistulosten arviointi 1/2009. Opetushallitus.

Metsämuuronen, J. (2009b). *Tutkimuksen tekemisen perusteet ihmistieteissä*. 4. laitos. International Methelp Oy.

Metsämuuronen, J. (2012). Challenges of the Fennema–Sherman test in the international comparisons. *International Journal of Psychological Studies, 4*(3), 1–22. http://www.ccsenet.org/journal/index.php/ijps/article/view/16904/12480

Metsämuuronen, J. (2013a). Effect of repeated testing to the development of Biblical Hebrew language proficiency. *Journal of Educational and Psychological Development, 3*(1), 10–24. http://dx.doi.org/10.5539/jedp.v3n1p10

Metsämuuronen, J. (2013b). A new method to setting standard for the wide range of language proficiency levels. *International Education Research*, 1(1), 1–21. https://doi.org/10.12735/IER.V1I1P01.

Metsämuuronen, J. (2017a). *Oppia ikä kaikki. Matemaattinen osaaminen toisen asteen koulutuksen lopussa 2015.* Julkaisut 1:2017. Kansallinen koulutuksen arviointikeskus. https://karvi.fi/wp-content/uploads/2017/03/KARVI_0117-1.pdf

Metsämuuronen, J. (2017b). *Essentials of Research Methods in Human Sciences*. SAGE Publications, Inc.

Metsämuuronen, J. (2018). Common framework for mathematics—Discussions of possibilities to develop a set of general standards for assessing proficiency in mathematics. *International Electronic Journal of Mathematics Education*, *13*(2), 13–39. https://doi.org/10.12973/iejme/2693

Metsämuuronen, J. (2019). Educational assessment and some related indicators of educational equality and equity. *Education Quarterly Reviews*, 2(4), 770–788. https://doi.org/10.31014/aior.1993.02.04.105.

Metsämuuronen J (2020a). Somers' *D* as an alternative for the item–test and item–rest correlation coefficients in the educational measurement settings. *International Journal of Educational Methodology, 6*(1), 207–221. https://doi.org/10.12973/ijem.6.1.207

Metsämuuronen, J. (2020b). Dimension-corrected Somers' D for the item analysis settings. *International Journal of Educational Methodology, 6*(2), 297–317. https://doi.org/10.12973/ijem.6.2.297

Metsämuuronen J. (2021a). Goodman–Kruskal gamma and dimension-corrected gamma in educational measurement settings. *International Journal of Educational Methodology,* 7(1), 95–118. https://doi.org/10.12973/ijem.7.1.95

Metsämuuronen, J. (2021b). Directional nature of Goodman-Kruskal gamma and some consequences. Identity of Goodman-Kruskal gamma and Somers delta, and their connection to Jonckheere-Terpstra test statistic. *Behaviormetrika, 48*(2), 283–307. http://dx.doi.org/10.1007/s41237-021-00138-8

Metsämuuronen, J. (2022a). Deflation-corrected estimators of reliability. *Frontiers in Psychology, 12*:748672, https://doi.org/10.3389/fpsyg.2021.748672

Metsämuuronen, J. (2022b). Attenuation-corrected reliability and some other MEC-corrected estimators of reliability. *Applied Psychological Measurement, 46*(8). https://doi.org/10.1177/01466216221108131

Metsämuuronen, J. (2022c). Effect of various simultaneous sources of mechanical error in the estimators of correlation causing deflation in reliability. Seeking the best options of correlation for deflation-corrected reliability. *Behaviormetrika, 49*(1), 91–130 https://doi.org/10.1007/s41237-022-00158-y

Metsämuuronen, J. (2022d). Reliability for a score compiled from multiple booklets with equated scores. Preprint. https://www.researchgate.net/publication/358849481_Reliability_for_a_score_compiled_from_multiple_booklets_with_equated_scores

Metsämuuronen, J. (2022e). How to obtain the most error-free estimate of reliability? Eight sources of underestimation of reliability. *Practical Assessment, Research, and Evaluation, PARE, 27*(1), Art. 10. https://doi.org/10.7275/7nkb-j673

Metsämuuronen, J. (2022f). Typology of deflation-corrected estimators of reliability. *Frontiers in Psychology, 13*:891959. http://dx.doi.org/10.3389/fpsyg.2022.891959

Metsämuuronen, J. (2022g). Artificial systematic attenuation in eta squared and some related consequences. Attenuation-corrected eta and eta squared, negative values of eta, and their relation to Pearson correlation. *Behaviormetrika,* https://doi.org/10.1007/s41237-022-00162-2

Metsämuuronen, J. (2022h). Directional nature of the product-moment correlation coefficient and some consequences. *Frontiers in Psychology*, *13*:988660. https://doi.org/10.3389/fpsyg.2022.988660

Metsämuuronen, J., Hermonen, A., Nousiainen, S. & Laakso, M.-J. (2023). Digitaaliseen testaamiseen liittyviä erityiskysymyksiä kansallisen oppimistulosarvioinnin näkökannalta. Teoksessa J. Metsämuuronen & S. Nousiainen (toim.), *Matematiikkaa COVID-19-pandemian varjossa II. Menetelmälliset ratkaisut matematiikan 9. luokan arvioinnissa keväällä 2021* (ss. 83–126). Julkaisut 5:2023. Kansallinen koulutuksen arviointikeskus.

Metsämuuronen, J. & Lehikko, A. (2022). Challenges and possibilities of educational equity and equality in the post-COVID-19 realm in the Nordic countries. *Scandinavian Journal of Educational Research*. https://doi.org/10.1080/00313831.2022.2115549

Metsämuuronen, J. & Mattsson, M. (2013). Effect of repeated testing to the development of vocabulary, nominal structures, and verbal morphology. *Journal of Educational and Psychological Development, 3*(2), 89–101. http://dx.doi.org/10.5539/jedp.v3n2p89

Metsämuuronen, J. & Nousiainen, S. (2021). *Matematiikkaa COVID-19-pandemian varjossa. Matematiikan osaaminen 9. luokan lopussa keväällä 2021*. Julkaisut 27:2021. Kansallinen koulutuksen arviointikeskus. https://karvi.fi/wp-content/uploads/2021/12/KARVI_2721.pdf

Metsämuuronen, J. & Nousiainen, S. (2023). Yleiset menetelmäratkaisut matematiikan oppimistulosten arvioinnissa vuonna 2021. Teoksessa J. Metsämuuronen & S. Nousiainen (toim.), *Matematiikkaa COVID-19-pandemian varjossa II. Menetelmälliset ratkaisut matematiikan 9. luokan arvioinnissa keväällä 2021* (ss. 21–82). Julkaisut 5:2023. Kansallinen koulutuksen arviointikeskus.

Metsämuuronen, J. & Räsänen, P. (2018). Cognitive–linguistic and constructivist mnemonic triggers in teaching based on Jerome Bruner's thinking. *Frontiers in Psycholology*, 9:2543. https://doi.org/10.3389/fpsyg.2018.02543

Metsämuuronen, J. & Salonen, R. V. (2017). *Matematiikan osaamisen piirteitä ammatillisessa koulutuksessa 2015 ja pitkän ajan muutoksia*. Julkaisut 2:2017. Kansallinen koulutuksen arviointikeskus. https://karvi.fi/wp-content/uploads/2017/03/KARVI_0217.pdf

Metsämuuronen, J. & Seppälä, H. (2021). *COVID-19-pandemia, osaamisvaje ja osaamisen eriytyminen*. Policy Brief 1:2021. Kansallinen koulutuksen arviointikeskus. https://karvi.fi/wp-content/uploads/2021/12/KARVI_Policy_brief_0121.pdf

Metsämuuronen, J., & Suomilammi, M. (2023). Kolmen kansallisen populaation keskeiset erottelevat piirteet sekä heikkojen ja parempien oppilaiden osaamisen rajapintatarkastelua. Teoksessa J. Metsämuuronen & S. Nousiainen (toim.), *Matematiikkaa COVID-19-pandemian varjossa II. Menetelmälliset ratkaisut matematiikan 9. luokan arvioinnissa keväällä 2021* (ss. 127–172). Julkaisut 5:2023. Kansallinen koulutuksen arviointikeskus.

Metsämuuronen, J. & Tuohilampi, L. (2017). *Matematiikan osaamisen piirteitä lukiokoulutuksen lopussa 2015*. Julkaisut 3:2017. Kansallinen koulutuksen arviointikeskus. https://karvi.fi/wp-content/uploads/2017/03/KARVI_0317.pdf

Metsämuuronen, J. & Ukkola, A. (2019). *Alkumittauksen menetelmällisiä ratkaisuja*. Julkaisut 18:2019. Kansallinen koulutuksen arviointikeskus. https://karvi.fi/app/uploads/2019/08/KARVI_1819.pdf

Metsämuuronen, J. & Ukkola, A. (2022). Rudimentary stages of the mathematical thinking and proficiency. Mathematical skills of low-performing pupils at the beginning of the first grade. *LUMAT: International Journal on Math, Science and Technology Education, 10*(2), 56–83. https://doi.org/10.31129/10.1.1632

Milanzi, E., Molenberghs, G., Alonso, A., Verbeke, G., & De Boeck, P. (2015). Reliability measures in item response theory: Manifest versus latent correlation functions. *British Journal of Mathematical & Statistical Psychology, 68*(1), 43–64. http://dx.doi.org/10.1111/bmsp.12033

Miller, P. (2021). Online Learning Trends to Support Students and Teachers in 2022. *EdTech Digest*. Dec 01, 2021. https://www.edtechdigest.com/2021/12/01/online-learning-trends-to-support-students-and-teachers-in-2022/

Mislevy, R. J. (1992). *Linking educational assessments: Concepts, issues, methods, and prospects*. ETS Policy Information Center.

Molina, C. A., Theodore, N., Ahmed, A. K., Westbroek, E. M., Mirovsky, Y., Harel, R., Orru', E., Khan, M., Witham, T., & Sciubba, D. M. (2019). Augmented reality–assisted pedicle screw insertion: a cadaveric proof-of-concept study, *Journal of Neurosurgery: Spine SPI, 31*(1), 139–146. https://doi.org/10.3171/2018.12.SPINE181142

Moses, T. (2017). A review of developments and applications in item analysis. Teoksessa R. Bennett & M. von Davier (toim.), *Advancing human assessment. The methodological, psychological and policy contributions of ETS* (ss. 19–46). Educational Testing Service. Springer Open. https://doi.org/10.1007/978-3-319-58689-2_2

Mullis, I. V. S., & Martin, M. O. (toim.). (2017). *TIMSS 2019 Assessment Frameworks*. Boston College, TIMSS & PIRLS International Study Center. http://timssandpirls.bc.edu/timss2019/frameworks/

Mullis, I. V. S., Martin, M. O., Foy, P., Kelly, D. L., & Fishbein, B. (2020). *TIMSS 2019 International Results in Mathematics and Science*. Boston College, TIMSS & PIRLS International Study Center. https://timssandpirls.bc.edu/timss2019/international-results/

Nardi, A. & Ranieri, M. (2019). Comparing Paper-Based and Electronic Multiple-Choice Examinations with Personal Devices: Impact on Students' Performance, Self-Efficacy and Satisfaction. *British Journal of Educational Technology*, 50(3), 1495–1506. https://doi.org/10.1111/bjet.12644

Niemi, E. K., & Metsämuuronen, J. (toim.) (2008). *Miten matematiikan taidot kehittyvät? Matematiikan oppimistulokset peruskoulun viidennen vuosiluokan jälkeen vuonna 2008*. Koulutuksen seurantaraportit 2010:2. Opetushallitus.

Nobel (2014). *Press release. 06.10.2014.* https://www.nobelprize.org/nobel_prizes/medicine/laureates/2014/press.html

Nyyssölä, K. & Kumpulainen, T. (2020). *Perusopetuksen ja kouluverkon tulevaisuusnäkymiä*. Raportit ja selvitykset 2020:25. Opetushallitus. https://www.oph.fi/sites/default/files/documents/Perusopetuksen_ja_kouluverkon_tulevaisuudennakymia.pdf

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M. Huotilainen, M., Iivonen, A., Vaivio. M. Alku, P., Ilmoniemi, R.J., Luuk, A., Allik, J., Sinkkonen, J., & Alho, K. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature, 385*, 432–434. https://doi.org/10.1038/385432a0

OAK (2017). *Ohje opettajille. ViLLE Team.* Oppisanalytiikan keskus. Turun yliopisto. https://ville.cs.utu.fi/opintopolku/villeohje-opintopolku-%20opettajille.pdf

OECD (2013). *Technical Report of the Survey of Adult Skills (PIAAC)*. 2. laitos. OECD. http://www.oecd.org/skills/piaac/_Technical%20Report_17OCT13.pdf

OECD (2016). *Technical Report of the Survey of Adult Skills (PIAAC)*. 2. laitos. OECD. http://www.oecd.org/skills/piaac/PIAAC_Technical_Report_2nd_Edition_Full_Report.pdf

OECD (2019a), PISA 2018 Results (Volume III): What School Life Means for Students' Lives. PISA, OECD Publishing. https://doi.org/10.1787/acd78851-en

OECD (2019b). *Country note*. OECD Publications.

Suomi: https://www.oecd.org/pisa/publications/PISA2018_CN_FIN.pdf;

Ruotsi: https://www.oecd.org/pisa/publications/PISA2018_CN_SWE.pdf;

Norja: https://www.oecd.org/pisa/ publications/PISA2018_CN_NOR.pdf;

Tanska: https://www.oecd.org/pisa/publications/ PISA2018_CN_DNK.pdf;

Viro: https://www.oecd.org/pisa/publications/PISA2018_CN_EST.pdf.

OPH (1998). *Koulutuksellisuuden tuloksellisuuden arviointimalli*. Arviointi 7/98. Opetushallitus.

OPH (2014). *Perusopetuksen opetussuunnitelman perusteet 2014*. Määräykset ja ohjeet 2014:96. Opetushallitus. https://www.oph.fi/sites/default/files/documents/perusopetuksen_opetussuunnitelman_perusteet_2014.pdf

OPH (2018). *Työllisyyden ja osaamisen muutoksia. Osaamisen ennakointifoorumin skenaariotyön tuloksia*. https://www.oph.fi/sites/default/files/documents/tyollisyyden-ja-osaamisen-muutoksia-oef-vaihe-iii-er3-koulutus-kulttuuri-ja-viestinta_1.pdf

OPH (2020). *Perusopetuksen päättöarvioinnin kriteerit*. Opetushallituksen määräys OPH-5042-2020. Opetushallitus. https://www.oph.fi/sites/default/files/documents/Perusopetuksen%20p%C3%A4%C3%A4tt%C3%B6arvioinnin%20kriteerit%2031.12.2020.pdf

Osguthorpe, R. T., & Graham, C. R. (2003). Blended learning environments: Definitions and directions. *Quarterly Review of Distance Education, 4*(3), 227. https://www.learntechlib.org/p/97576/

Papas, C. (2021). *The 10 Best Blended Learning LMS Solutions (2022)*. https://elearningindustry.com/best-blended-learning-lms-solutions

Pearson, K. (1896). VII. Mathematical contributions to the theory of evolution. III. Regression, heredity and panmixia. Philosophical *Transactions of the Royal Society of London, 187*, 253–318. https://doi.org/10.1098/rsta.1896.0007

Pearson, K. (1900). I. Mathematical contributions to the theory of evolution. VII. On the correlation of characters not quantitatively measurable. *Philosophical Transactions of the Royal Society A. Mathematical, Physical and Engineering Sciences, 195*(262–273), 1–47. https://doi.org/10.1098/rsta.1900.0022

Pearson, K. (1913). On the measurement of the influence of "broad categories" on correlation. *Biometrika*, 9(1–2), 116–139. https://doi.org/10.1093/biomet/9.1-2.116

Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational psychology review*, *18*(4), 315–341. https://doi.org/10.1007/s10648-006-9029-9

Pekrun, R. (2017), Emotion and achievement during adolescence. *Child Development Perspectives, 11*(3), 215–221. https://doi.org/10.1111/cdep.12237

Pekrun, R., Elliot, A. J., & Maier, M. A. (2009). Achievement goals and achievement emotions: Testing a model of their joint relations with academic performance. *Journal of Educational Psychology, 101*, 115–135. https://doi.org/10.1037/a0013383

Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational psychologist*, *37*(2), 91–105. https://doi.org/10.1207/S15326985EP3702_4

Pekrun, R., & Linnenbrink-Garcia, L. (2012). Academic emotions and student engagement. Teoksessa S. L. Christenson, A. L. Reschly, & C. Wylie (toim.), *Handbook of research on student engagement* (ss. 259–282). Springer Science + Business Media. https://doi.org/10.1007/978-1-4614-2018-7_12

Pekrun, R., & Perry, R. P. (2014). Control-value theory of achievement emotions. Teoksessa R. Pekrun & L. Linnenbrink-Garcia, *International Handbook of Emotions in Education* (ss. 130–151). Routledge.

PISA (2010). *PISA Computer-Based Assessment of Student Skills in Science. OECD*. Publication: 24/8/2010. http://www.oecd.org/edu/school/programmeforinternationalstudentassessmentpisa/pisacomputer-basedassessmentofstudentskillsinscience.htm

PISA (2015a). Main results from the PISA 2012 Computer-based Assessments. Teoksessa OECD, *Students, Computers and Learning. Making the Connection* (ss. 81–103). http://dx.doi.org/10.1787/9789264239555-6-en

PISA (2015b). *Students, Computers and Learning. Making the Connection*. OECD. http://dx.doi.org/10.1787/9789264239555-en

PISA (2018). Annex A5. How comparable are the PISA 2018 computer- and paper-based tests? Teoksessa *PISA 2018 Results (Volume I): What Students Know and Can Do*. OECD. https://www.oecd-ilibrary.org/sites/8f293551-en/index.html?itemId=/content/component/8f293551-en

Price, W. N. II, & Cohen, I. G. (2019). Privacy in the age of medical big data. *Nature Medicine, 25*(1), 37–43. http://dx.doi.org/10.1038/s41591-018-0272-7

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Studies in Mathematic Psychology I. Danmarks Pædagogishe Institut. Nielsen & Lydiche.

Regeringen (2021). *Bred aftale om fremtidigt evaluerings- og bedømmelsessystem*. Børne- og Undervisningsministeriet, 29.10.2021. https://www.regeringen.dk/nyheder/2021/bred-aftale-om-fremtidigt-evaluerings-og-bedoemmelsessystem

Roediger, H.L. 3rd & Karpicke, J.D. (2006a). The power of testing memory. Basic research and implications for educational practice. *Perspectives on Psychological Science, 1*(3), 181–210. http://dx.doi.org/10.1111/j.1745-6916.2006.00012.x

Roediger, H. L., & Karpicke, J. D. (2006b). Test-enhanced learning: taking memory tests improves long-term retention. *Psychological Science*, *17*(3), 249–255. http://dx.doi.org/10.1111/j.1467-9280.2006.01693.x

Rowland, D. C., Roudi, Y., Moser, M. B., & Moser, E. I. (2016). Ten years of grid cells. *Annual Review of Neuroscience, 8*(39), 19–40. https://doi.org/10.1146/annurev-neuro-070815-013824

Räkköläinen, M, Metsämuuronen J., Holopainen J., & Hievanen R. (2017). *Kestävän kehityksen osaaminen, opetus ja koulutuksen järjestäjän toiminta ammatillisissa perustutkinnoissa.* Julkaisut 12:2017. Kansallinen koulutuksen arviointikeskus. https://karvi.fi/wp-content/uploads/2017/05/KARVI_1217.pdf

Räsänen, P., Aunio, P., Laine, A., Hakkarainen, A., Väisänen, E., Finell, J., Rajala, T., Laakso, M.-J., & Korhonen, J. (2021). Effects of gender on basic numerical and arithmetic skills: Pilot data from 3rd to 9th grade for a large-scale online dyscalculia screener. *Frontiers in Education*, 6:683672. https://doi.org/10.3389/feduc.2021.683672

Raykov, T., & Marcoulides, G. A. (2017). Thanks coefficient alpha, we still need you! *Educational and Psychological Measurement, 79*(1), 200–210. http://dx.doi.org/10.1177/0013164417725127

Raykov, T., West, B. T., & Traynor, A. (2015). Evaluation of coefficient alpha for multiple component measuring instruments in complex sample designs. *Structural Equation Modeling, 22*(3), 429–438. http://dx.doi.org/10.1080/10705511.2014.936081

Salonen, A. O. (2010). *Kestävä kehitys globaalin ajan hyvinvointiyhteiskunnan haasteena*. Tutkimuksia 318. Helsingin yliopisto, Käyttäytymistieteellinen tiedekunta. https://helda.helsinki.fi/bitstream/handle/10138/20067/kestavak.pdf

Salonen, R. V. (2023). Tunteiden mittaaminen matematiikan arvioinnissa—Tunnemittari, uskomukset ja kontrolli–arvo-teoria. Teoksessa J. Metsämuuronen & S. Nousiainen (toim.), *Matematiikkaa COVID-19-pandemian varjossa II. Menetelmälliset ratkaisut matematiikan 9. luokan arvioinnissa keväällä 2021* (ss. 173–189). Julkaisut x:2023. Kansallinen koulutuksen arviointikeskus.

Sandal, G. (2021). The metaverse is coming: How should organisations prepare? Nov. 30. *Futures Platform*. https://www.futuresplatform.com/blog/metaverse-how-should-organisations-prepare

Sattar, M., Palaniappan, S., Lokman, A., Shah, N., Khalid, U. & Hasan, R. (2020). Motivating medical students using virtual reality based education. *International Journal of Emerging Technologies In Learning (IJET), 15*(02), 160–174. http://dx.doi.org/10.3991/ijet.v15i02.11394

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research*, *99*(6), 323–338. http://dx.doi.org/10.3200/JOER.99.6.323-3

Schumacker, R. E., & Lomax, R. G. (2004). *Beginner's guide to structural equation modeling*. 2. laitos. Lawrence Erlbaum Associates.

Shaheen, N.L. (2021). Accessibility4Equity: Cripping technology-mediated compulsory education through sociotechnical praxis. *British Journal of Educational Technology, 53*(1), 77–92. https://doi.org/10.1111/bjet.13153

Shayer, M., & Ginsburg, D. (2009). Thirty years on—A large anti- Flynn effect? (II): 13- and 14-year-olds. Piagetian tests of formal operations norms 1976–2006/7. *British Journal of Educational Psychology, 79*(3), 409–418. https://doi.org/10.1348/978185408x383123

Shayer, M., Ginsburg, D., & Coe, R. (2007). Thirty years on—a large anti-Flynn effect? The Piagetian test Volume & Heaviness norms 1975–2003. *British Journal of Educational Psychology,* 77(1), 25–41. https://doi.org/10.1348/000709906x96987

Shdaifat, A. M., Obeidallah, R. A., Ghazal, G., Abu Sarhan, A., & Abu Spetan, N. R. (2020). A proposed iris recognition model for authentication in mobile exams. *International Journal of Emerging Technologies in Learning (iJET)*, *15*(12), 205–216. https://doi.org/10.3991/ijet.v15i12.13741

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika, 74*(1), 107–120. https://doi.org/10.1007/s11336-008-9101-0

Sijtsma, K., & Pfadt, J.M. (2021). Part II: On the Use, the Misuse, and the Very Limited Usefulness of Cronbach's Alpha: Discussing Lower Bounds and Correlated Errors. *Psychometrika 86*, 843–860. https://doi.org/10.1007/s11336-021-09789-8

Silverström, C. & Rautopuro J. (2015). *Språk och skrivande i årskurs 9. En utvärdering av lärresultat i modersmål och litteratur våren 2014*. Publikationer 18:2015. Nationella Centret för Utbildningsutvärdering. https://karvi.fi/wp-content/uploads/2015/09/KARVI_1815.pdf

Sitra (2020). *Megatrendit 2020*. Sitra. https://www.sitra.fi/aiheet/megatrendit/

Sjögren, A., Engdahlc, M., Hall, C., Holmlund, H., Lundin, M., Mühlrad, H., & Öckert, B. (toim.) (2021). *Swedish children and youth during the COVID-19 pandemic*. Working Papers 2021:3. The Institute for Evaluation of Labour Market and Education Policy. https://www.ifau.se/globalassets/pdf/se/2021/wp-2021-03-swedish-children-and-youth-during-the-covid-19-pandemic.pdf

Skinner, E., Furrer, C., Marchand, G., & Kindermann, T. (2008). Engagement and disaffection in the classroom: Part of a larger motivational dynamic? *Journal of educational psychology*, *100*(4), 765–781. https://doi.org/10.1037/a0012840

Soini, K. (2013). Kestävä kehitys ja kulttuuri. Teoksessa M. Laine (toim.), *Kestävä kasvatus—Kulttuuria etsimässä* (ss. 12–25). Suomen kulttuuriperintökasvatuksen seuran julkaisuja 6.

Somers, R. H. (1962). A new asymmetric measure of correlation for ordinal variables. *American Sociological Review, 27*(6), 799–811. http://dx.doi.org/10.2307/2090408

Steiger, J. H. (2007). Understanding the limitations of global fit assessment in structural equation modeling. *Personality and Individual differences, 42*(5), 893–898. https://doi.org/10.1016/j.paid.2006.09.017

Sternberg, R. J. & Grigorenko, E. L. (2002). *Dynamic testing*. Cambridge University Press.

Taivassalo, M. (2019). *Uudistuvat oppimisympäristöt ja digitaaliset ratkaisut oppimisen tukena—esimerkkejä erilaisista oppimisympäristöistä ja -ratkaisuista*. Zoomin alueellinen koulutus Turku 10.9.2019. Opetushallitus. https://www.oph.fi/sites/default/files/documents/uudistuvat_oppimisymparistot_minna_taivassalo.pdf

TOEFL (2007). *Test and Score Data Summary for TOEFL®. Computer-Based and Paper-Based Tests*. Educational Testing Service. https://www.ets.org/Media/Research/pdf/TOEFL-SUM-0506-CBT.pdf

TOEFL (2022). *TOEFL® Test Security: Detection*. https://www.ets.org/toefl/score-users/about/security/detection/

Tulving, E. (1967). The effects of presentation and recall of material in free-recall learning. *Journal of Verbal Learning and Verbal Behavior, 6*, 175–184. http://dx.doi.org/10.1016/S0022-5371(67)80092-6

Tuohilampi, L., & Giaconi, V. (2013). Minäkäsitys, motivaatio sekä tunteet matematiikkaan liittyen: kolmasluokkalaisten vertailua Suomessa ja Chilessä. Teoksessa M. Hähkiöniemi, H. Leppäaho, P. Nieminen, & V. Viiri, *Proceedings of the 2012 Annual Conference of Finnish Mathematics and Science Education Research Association: Matematiikan ja luonnontieteiden opetuksen tutkimusseuran konferenssijulkaisu 2012* (ss. 117–128). Jyväskylän yliopisto.

Ukkola, A. & Kivistö, A. (toim.) (2023). *Matematiikan viitekehys kansallisen koulutuksen arviointikeskuksen oppimistulosarviointiin vuosiluokilla 1–9*. Julkaisut 6:2023. Kansallinen koulutuksen arviointikeskus.

Ukkola, A. & Metsämuuronen, J. (2019). *Alkumittaus—Matematiikan ja äidinkielen ja kirjallisuuden osaaminen ensimmäisen luokan alussa*. Julkaisut 17:2019. Kansallinen koulutuksen arviointikeskus. https://karvi.fi/wp-content/uploads/2019/07/KARVI_1719.pdf.

Ukkola, A., & Metsämuuronen, J. (2021). *Matematiikan ja äidinkielen ja kirjallisuuden osaaminen kolmannen luokan alussa*. Julkaisut 20:2021. Kansallinen koulutuksen arviointikeskus. https://karvi.fi/wp-content/uploads/2021/08/KARVI_2021.pdf

Ukkola, A., & Metsämuuronen J. (2023). *Matematiikan ja äidinkielen taidot alkuopetuksen aikana. Perusopetuksen oppimistulosten pitkittäisarviointi 2018–2020*. Julkaisut 1:2023. Kansallinen koulutuksen arviointikeskus. https://karvi.fi/wp-content/uploads/2023/01/KARVI_0123.pdf

Ukkola, A., Metsämuuronen, J. & Paananen, M. (2020). *Alkumittauksen syventäviä kysymyksiä*. Julkaisut 10:2020. Kansallinen koulutuksen arviointikeskus. https://karvi.fi/wp-content/uploads/2020/08/KARVI_Alkumittaus.pdf

Vainikainen, M.-P. & Hautamäki, J. (2022). Three Studies on Learning to Learn in Finland: Anti-Flynn Effects 2001–2017. *Scandinavian Journal of Educational Research, 66*(1), 43–58 . https://doi.org/10.1080/00313831.2020.1833240

Van Boxel, M. (2016). *Implementing large-scale computer-based high-stakes testing in the Netherlands*. Presentation on conference Future Tests and Test Environments, June 16, 2016, Umeå, Sweden.

Van der Linden, W. J., & Glas, C.A.W. (toim.). (2000). *Computerized adaptive testing: Theory and practice*. Kluwer.

Van der Schoot, F. (2009). Cito variation of the Bookmark method. Reference supplement, Section I. Teoksessa S. Takala (toim.), *Relating language examinations to the Common European Framework of Reference for Languages: Learning, teaching, assessment (CEFR). A Manual*. Language Policy Division, Strasbourg. https://rm.coe.int/1680667a24

VATT (2018). *Economic Policy Council Report 2017*. Economic Policy Council. VATT. https://www.talouspolitiikanarviointineuvosto.fi/wordpress/wp-content/uploads/2018/01/Report2017.pdf

Veltheim, H. (2021). Future of education after COVID-19: AI becomes the teacher while humans mentor and coach. Jun 1. *Futures Platform*. https://www.futuresplatform.com/blog/future-education-after-covid-19-ai-becomes-teacher-while-humans-mentor-and-coach

Venäläinen, S., Laimi, T., Seppälä, S., Vuojus, T., Viitala, M., Ahlholm, M., Latomaa, S., Mård-Miettinen, K., Nirkkonen, M., Huhtanen, M., & Metsämuuronen, J. (2022). *Kielellisiä taitoja ja koulunkäyntivalmiuksia—Valmistavan opetuksen ja oman äidinkielen opetuksen tila ja vaikuttavuus -arviointi.* Julkaisut 19:2022. Kansallinen koulutuksen arviointikeskus. https://karvi.fi/wp-content/uploads/2022/09/KARVI_1922.pdf

Vie, J.-J., Popineau, F., Bruillard, É., & Bourda. Y. (2017) A review of recent advances in adaptive assessment. Teoksessa Learning Analytics: Fundaments, Applications, and Trends, 94, Springer International Publishing, (ss. 113–142). *Studies in Systems, Decision and Control 978-3-319-52976-9. hal-01488284*. https://hal.archives-ouvertes.fr/hal-01488284/document

Villavicencio, F. T., Bernardo, A. B. I. (2016). Beyond Math Anxiety: Positive Emotions Predict Mathematics Achievement, Self-Regulation, and Self-Efficacy. The *Asia-Pacific Education Researcher, 25*, 415–422. https://doi.org/10.1007/s40299-015-0251-4

Vojdanoska M., Cranney, J. & Newell, B.R. (2009). The testing effect: The role of feedback and collaboration in a tertiary classroom setting. *Applied Cognitive Psychology*, *24*(8), 1183–1195. http://dx.doi.org/10.1002/acp.1630

Wainer, H. (toim.). (2000). *Computerized adaptive testing: A Primer*. 2. laitos. Lawrence Erlbaum Associates

Watkins, M. W. (2018). Exploratory Factor Analysis: A Guide to Best Practice. *Journal of Black Psychology, 44*(3), 219–246. https://doi.org/10.1177/0095798418771807

Weiss, D.J., & Kingsbury, G.G. (1984). Application of computerized adaptive testing to educational problems. *Journal of Educational Measurement, 21*(4), 361–375. https://doi.org/10.1111/j.1745-984.1984.tb01040.x

Wise S.L. (2019). Controlling construct-irrelevant factors through computer-based testing: disengagement, anxiety, & cheating. *Education Inquiry, 10*(1), 21–33. https://doi.org/10.1080/20004508.2018.1490127

Woodley, M. A., & Meisenberg, G. (2013). In the Netherlands the anti-Flynn effect is a Jensen effect. *Personality and Individual Differences, 54*(8), 871–876. http://dx.doi.org/10.1016/j.paid.2012.12.02

YK (1987). *Report of the World Commission on Environment and Development, General Assembly Resolution 42/187*, 11 December 1987. Yhdistyneet Kansakunnat. http://www.un-documents.net/a42-427.htm

YK (2010). Culture and development. Assembly Resolution 65/166, 20 December 2010. Yhdistyneet Kansakunnat. http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CLT/pdf/8_UNGA_Resolution_A_RES_65_166_EN.pdf

YK (2020). *Policy Brief. Education during COVID-19 and beyond*. August 2020. Yhdistyneet Kansakunnat. https://www.un.org/development/desa/dspd/wp-content/uploads/sites/22/2020/08/sg_policy_brief_covid-19_and_education_august_2020.pdf

YTL (2022). *Digitaalisten kokeiden kuvaukset*. Ylioppilastutkintolautakunta. https://www.ylioppilastutkinto.fi/ylioppilastutkinto/digitaalinen-ylioppilastutkinto/digitaalisten-kokeiden-kuvaukset

Zumbo, B. D., Gadermann, A. M., & Zeisser, C. (2007). Ordinal versions of coefficients alpha and theta for Likert rating scales. *Journal of Modern Applied Statistical Methods, 6*(1), 21–29. http://dx.doi.org/10.22237/jmasm/1177992180

Åkerlund, C., Marjanen, J., & Lepola L. (2019). *Lärresultat i finska i åk 6. Resultat av en utvärdering våren 2018*. Publikationer 6:2019. Nationella Centret för Utbildningsutvärdering. https://karvi.fi/wp-content/uploads/2019/04/KARVI_0619.pdf

Åkerlund, C., Marjanen, J. & Peltola, E. (2022). *A-finska i åk 9. Utvärdering av lärresultat våren 2021*. Publikationer 4:2022. Nationella Centret för Utbildningsutvärdering. https://karvi.fi/wp-content/uploads/2022/03/NCU_0422.pdf

Liitteet

LIITE 1. Julkaistava testiversio

1.1 Suomenkielinen versio

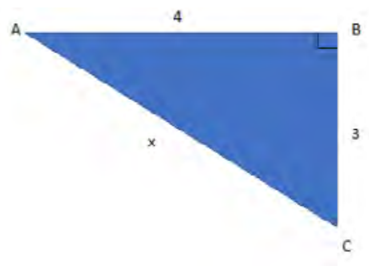
PERUSOPETUKSEN 9. VUOSILUOKAN MATEMATIIKAN SUMMATIIVINEN TESTI (Karvi 2021)

A-osa: Päässälaskutehtävät

	KUUNNELTAVAT PÄÄSSÄLASKUTEHTÄVÄT 1–2 Ohje oppilaalle: Kuulet jokaisen kysymyksen 2 kertaa, minkä jälkeen sinulla on 1 minuutti aikaa vastata.	
1. M20S1T10_003	Opettaja lukee/Äänitiedostosta kuuluu (kaksi kertaa): *Neliön pinta-ala on 25 cm². Kuinka pitkä on yksi sivu?*	Vastaus: _______________
2. M20S1T10_002	Opettaja lukee/Äänitiedostosta kuuluu (kaksi kertaa): *Virtasen perheessä kuluu maitoa 1,5 litraa päivässä. Kuinka monta litraa Virtasilla kuluu maitoa kahdessa viikossa?*	Vastaus: _______________

	PÄÄSSÄLASKUTEHTÄVÄT 3–6 Vastausaikaa yhteensä 4 minuuttia. Apuvälineitä ei saa käyttää. Digitaalisessa kokeessa ei saa käyttää kynää, paperia tai laskukonetta.	
3. M20S1T10_011	Laske 4,37−1,28	Vastaus: _______________
4. M20S2T10_028A	Kirjoita lausekkeen yksinkertaisin muoto $2 \cdot 4^2$	Vastaus: _______________
5. M12S1T10_163	Minulta menee 24 minuuttia, kun kävelen koulusta kotiin nopeudella 5 km/h. Kauanko kotimatkani kestää, kun ajan sen pyörällä nopeudella 15 km/h?	Vastaus: _______________
6. M20S1T10_012DEF 2	Luokassa on tuoleja ja jakkaroita. Kolmijalkaisia jakkaroita on kahdeksan. Niissä on yhteensä yhtä monta jalkaa kuin nelijalkaisissa tuoleissa. Kuinka paljon tuoleja ja jakkaroita on yhteensä?	Vastaus: _______________

B-osa: Monivalintatehtävät ja lyhyet vastaukset

<table>
<tr><td>7.
M15S4T15_174</td><td>Omenan hinta saadaan funktion $f(x)=2x+1$ arvosta, jossa x on omenoiden määrä (kg). Kuinka paljon maksaa 3,5 kg omenoita?</td><td>a) 24,5 €
b) 9 €
c) 8 €
d) 7 €</td></tr>
<tr><td>8.
M20S5T16_074 12356</td><td colspan="2">


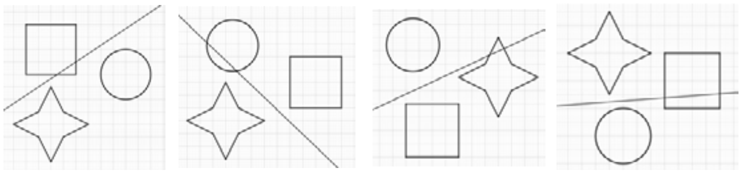
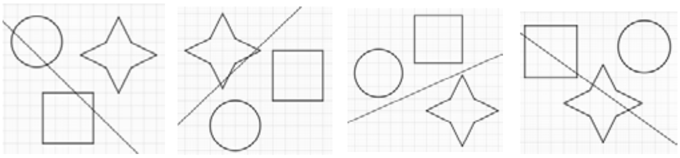
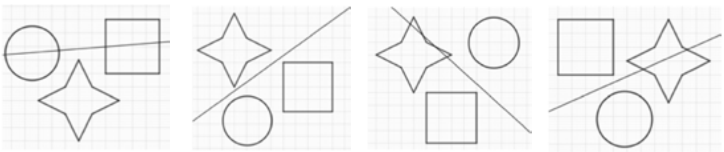
Ovatko väittämät oikein? Merkitse viivalle K= kyllä tai E= ei

1. Kolmio ABC on tylppäkulmainen. _______
2. Kolmio on symmetrinen pisteen B suhteen. _______
3. Kolmio on suorakulmainen. _______
4. Kulman A suuruus on 60 astetta. _______
5. Kolmion pinta-ala on 12 pinta-alayksikköä _______
</td></tr>
<tr><td>9.
M20S2T12_027 A-G 5</td><td>Sievennä 8:24</td><td>a) 1/20
b) 1/5
c) 1/3
d) 0,45
e) 4,5</td></tr>
<tr><td>10.
M20S2T13_027 A-G 6</td><td>Joukkue on pelannut tasapelin 20 % peleistä ja hävinnyt 25 %. Kuinka monta prosenttia peleistä se on voittanut?</td><td>a) 25 %
b) 10 %
c) 30 %
d) 50 %
e) 55 %</td></tr>
<tr><td rowspan="3">11.
M20S6T19_105</td><td colspan="2">Frekvenssi (f) ilmoittaa, kuinka monta kertaa havainto esiintyy tilastossa. Suhteellinen frekvenssi (f %) ilmoittaa, kuinka monta prosenttia havainnon osuus on kaikista havainnoista.

Erään luokan matematiikan kokeen arvosanat olivat

8, 7, 7, 7, 10, 6, 6, 8, 8, 8, 10, 9, 8, 7, 9, 9, 6, 5, 5, 7.</td></tr>
<tr><td>1. Mikä on arvosanan 7 frekvenssi?</td><td>a) 2
b) 3
c) 5
d) 6</td></tr>
<tr><td>2. Mikä on arvosanan 7 suhteellinen frekvenssi?</td><td>a) 25 %
b) 20 %
c) 50 %
d) 10 %</td></tr>
<tr><td>12.
M20S5T18_080 257</td><td colspan="2">Voiko seuraavien tietojen avulla laskea kolmion pinta-alan? Merkitse viivalle K=kyllä tai E=ei.

1. Kaksi kulmaa _____
2. Kolme kulmaa _____
3. Kaksi sivua ja niiden välinen kulma ____</td></tr>
</table>

13. M20S1T10_014	Seuraavat kuviot täyttävät tietyt tuntemattomat ehdot.  Seuraavat kuviot <u>eivät</u> täytä näitä tuntemattomia ehtoja.  Mikä seuraavista kuvioista täyttää tuntemattomat ehdot? Ympyröi.  **A** **B** **C** **D**	
14. M11S6T19_184	Katrilla on alla olevat pelimerkit pussissa, ja pussia on ravistettu.  Katri nostaa pussista yhden pelimerkin katsomatta. Millä todennäköisyydellä pelimerkissä oleva luku on jaollinen kolmella?	a) 1/11 b) 1/3 c) 4/11 d) 1/4 e) 4/7
15. M15S5T16_179	Oikokulmassa on kolme kulmaa, *a*, *a* ja 50 astetta. Kuinka suuri on kulma *a*?	a) 20 b) 55 c) 65 d) 75
16. M20S4T15_064 145	Nina on kylvyssä. Kylvyn jälkeen hän nostaa tulpan, ja vesi alkaa valua viemäriin. Funktio $f(x)$=–40x+180 kuvaa, kuinka paljon vettä kylpyammeessa on, kun tulpan poistamisesta on kulunut aikaa x minuuttia. Ovatko väittämät oikein (O) vai väärin (V)? Merkitse viivalle O tai V. 1. Ammeessa oli 140 litraa vettä, kun Nina oli kylvyssä. _______ 2. Amme on tyhjä 4 min 30 s jälkeen. _______ 3. Minuutin päästä tulpan poistamisesta ammeessa on 140 litraa vettä. _______	

17.

M20S6T19_093

Samuel on hakemassa toisen asteen koulutukseen. Taulukkoon hän on kerännyt arvosanansa. Hän haluaa koulutukseen, jossa lukuaineiden keskiarvoraja on ollut useana vuonna 8,2. Mikä arvosana Samuelin pitää vähintään saada matematiikasta, jotta hän saavuttaisi keskiarvorajan?

Lukuaineet	Arvosana
äidinkieli	8
englanti	8
ruotsi	7
biologia	9
maantieto	7
uskonto	8
historia	10
matematiikka	
kemia	9
fysiikka	7
yhteiskuntaoppi	9

Vastaus: ____

18.

M12S2T11_170A

Mitä tulee vastaukseksi laskulausekkeesta 2–5·2–5?

Vastaus: ______________________

19.

M20S2T13_025

Kalle ostaa takin. Paljonko takki maksaa alennuksen jälkeen?

a) 89,70 €
b) 199,33 €
c) 209,30 €
d) 269 €
e) 388,70 €

20.

M20S3T14_048B

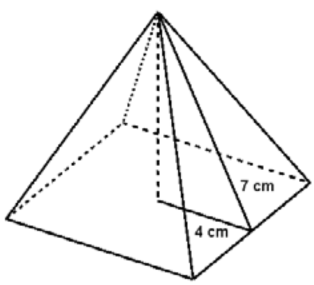
Laske polynomin x^2+6x-8 arvo, kun $x=3$. Ympyröi oikea vastaus.

a) 16
b) 19
c) 35
d) 64

21.

M12S5T16_177

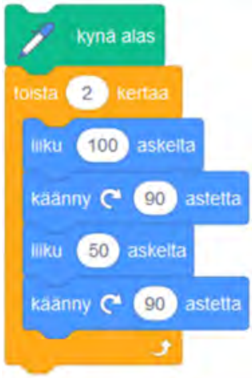
Mikä väitteistä on tosi? Ympyröi.

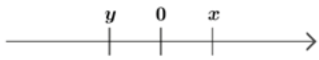


a) Pyramidin pohja on suorakulmainen kolmio
b) Pyramidin korkeus on 7 *cm*
c) Pyramidin pohjaneliön sivun pituus on 4 *cm*
d) Pyramidin pohjaneliön pinta-ala on 64 cm^2
e) Pyramidin sivutahkon pinta-ala on 56 cm^2

<table>
<tr>
<td>22.

M20S1T20_045D 1</td>
<td>Minkälaisen kuvion seuraava koodi tuottaa?


</td>
<td>a) kolmio
b) neliö
c) suorakulmio, joka ei ole neliö
d) kuusikulmio
e) ei mikään näistä</td>
</tr>
<tr>
<td>23.

M20S3T14_052 13</td>
<td colspan="2">

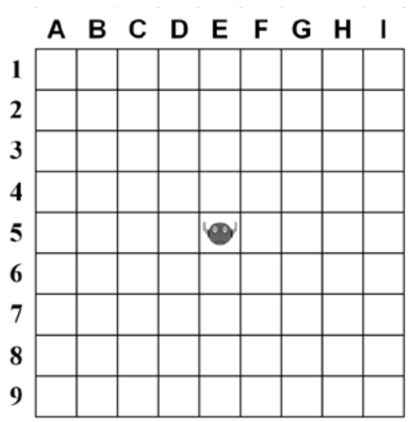
Tarkastele lukusuoraa. Pitävätkö seuraavat väittämät paikkansa?
Merkitse viivalle väitteen perään K=kyllä tai E=ei.

a) $x - y = y - x$ _____

b) $x - y > y - x$ _______</td>
</tr>
<tr>
<td>24.

M20S1T20_008 12</td>
<td colspan="2">Ohjelmoitavalle robotille annetaan sarja käskyjä, jotka se suorittaa järjestyksessä. Mahdolliset käskyt ovat E=”eteenpäin yksi ruutu”, T=”taaksepäin yksi ruutu”, O=”käänny 90 astetta oikealle” ja V=”käänny 90 astetta vasemmalle”.

Käytetään näille lyhenteitä E/T/O/V. Robotti liikkuu 9×9-ruudukossa, joka on kuvattu alla.



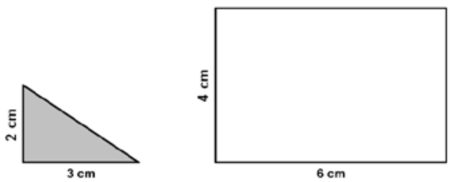
a) Robotti aloittaa ruudussa E5, katse ylöspäin. Sille annetaan käskysarja ”OEETV”.
Missä ruudussa robotti pysähtyy?

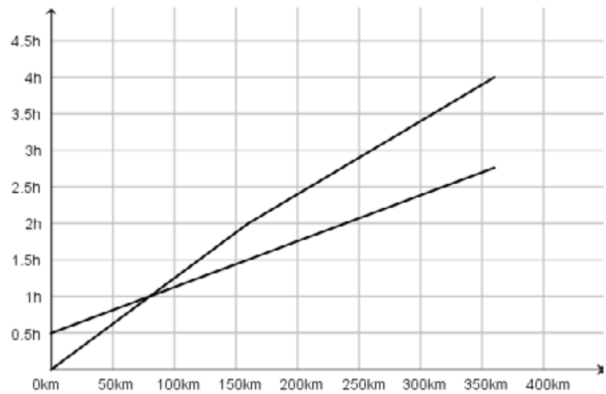
Vastaus: ____________

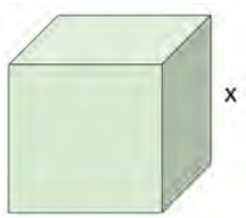
b) Robotti aloittaa ruudussa E5, katse ylöspäin. Sille annetaan käskysarja ”OEEVEVEVVO”.
Missä ruudussa robotti pysähtyy?

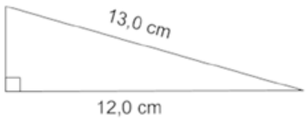
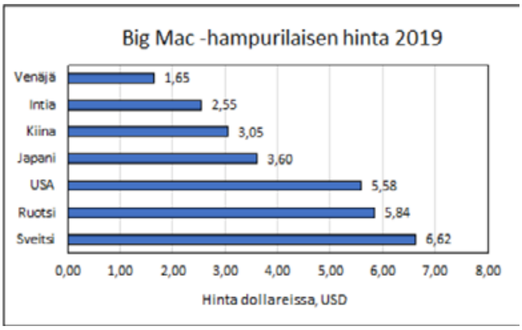
Vastaus: _____________</td>
</tr>
</table>

<table>
<tr><td>25.

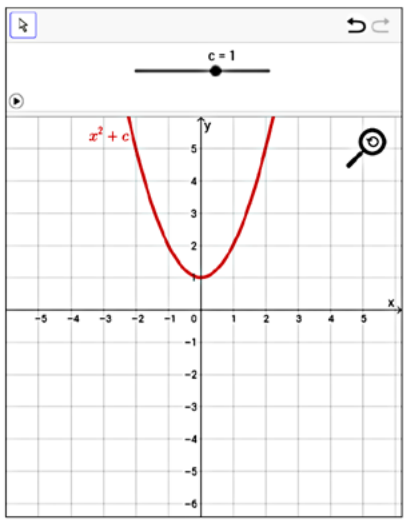
M12S5T17_178</td><td>Kuinka monta alla olevan kaltaista harmaata suorakulmaista kolmiota tarvitaan peittämään tarkalleen suorakulmion pinta?
</td><td>a) neljä
b) kuusi
c) kahdeksan
d) kymmenen
e) kaksitoista</td></tr>
<tr><td>26.

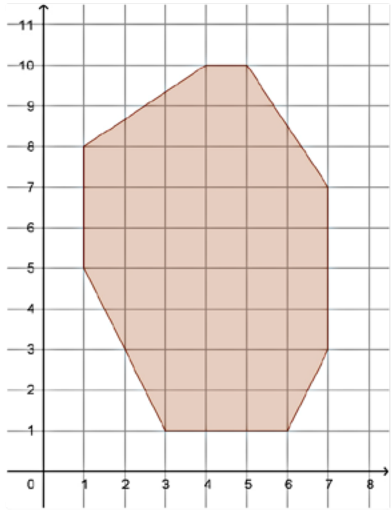
M15S4T15_183A 1</td><td colspan="2">Autoilija lähtee ajamaan Seinäjoelta Helsinkiin. Puoli tuntia myöhemmin Pendolino-juna lähtee Seinäjoelta Helsinkiin. Oheisessa kuvassa on auton ja junan kulkema matka. Arvioi kuvan perusteella: Millä nopeudella auto ajaa ensimmäiset kaksi tuntia?

Vastaus: ____</td></tr>
<tr><td rowspan="4">27.

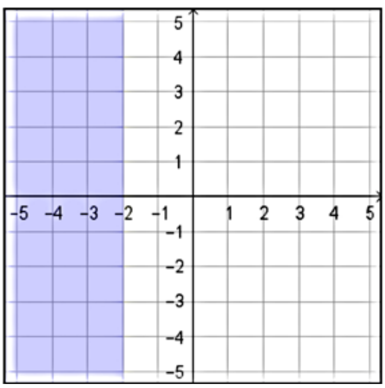
M20S3T14_046
A-F_BDE</td><td colspan="2">Kuution särmä on x.
</td></tr>
<tr><td>1. Mikä on tahkon pinta-ala?</td><td>a) $2x^2$
b) $4x$
c) x^2
d) $12x$
e) $6x^2$</td></tr>
<tr><td>2. Mikä on yhden tahkon piiri?</td><td>a) $2x^2$
b) $4x$
c) x^2
d) $12x$
e) $6x^2$</td></tr>
<tr><td>3. Mikä on särmien yhteispituus?</td><td>a) $2x^2$
b) $4x$
c) x^2
d) $12x$
e) $6x^2$</td></tr>
</table>

28. M20S1T10_019	Mikä on lukusarjan seuraava luku? 0, 1, 1, 2, 4, 7, 13, 24, 44, _____	
29. M20S5T17_084	Laske kateetin pituus.  	Vastaus: _______________
30. M20S2T13_103	Kuinka monta prosenttia kalliimpi Big Mac -hampurilainen on Sveitsissä kuin Intiassa? Anna vastaus prosentteina prosentin tarkkuudella.  	Vastaus: _______________

C-osa: Vaativammat ongelmanratkaisutehtävät

31. MYk19S2T11_193 A-F 13	Yhdistä lukujono sen määritelmään. 1. $a_n=2n-1$ 2. $a_n=n^3$	a) (1, 8, 27, 64, …) b) (1, 2, 3, 4, …) c) (1, 2, 3, 5, …) d) (1, 4, 9, 16, …) e) (1, 3, 5, 7, …)
32. M00S3T14_172B	Ratkaise yhtälö $5x–7=3x$. Perustele.	
33. M20S4T15_061	Kerro, miten vakion c arvon muuttuminen vaikuttaa funktion $f(x)=x^2+c$ kuvaajan kulkuun. Vastaus:  	

34. MYs18S5T18_201	Muotokilpailun palkintolautakunta myöntää muistolaatan kilpailun parhaille teoksille. Lautakunnan taiteellinen avustaja tekee ensimmäisen version laatan pienoismallista käyttämällä GeoGebra-ohjelman koordinaatistopiirrosta. Hän aloittaa suorakulmiosta, jonka leveys on 6 ja korkeus 9 pituusyksikköä. Leikkaamalla pois tämän suorakulmion kaikki neljä kulmaa eri tavoilla hän päätyy viereisen kuvion monikulmioon. Määritä tämän monikulmion pinta-ala. Perustele vastauksesi. Vastaus: 
35. M20S4T14_070 12	Taksi A maksaa 3 euroa kilometriltä ja siinä on aloitusmaksu 4,50 euroa. Taksi B puolestaan maksaa 5 euroa kilometriltä ja siinä on aloitusmaksu 1,50 euroa. Muodosta funktiot taksien hinnoittelusta (merkitse kilometrejä muuttujalla *x*). a) Taksi A b) Taksi B
36. MYs17S3T14_196 B	Ratkaise yhtälöpari. $\begin{cases} y - x + 1 = 0 \\ 4y = 12 - x. \end{cases}$ Kirjoita välivaiheet. Vastaus:

37. MYs17S3T14_196 C	Ratkaise yhtälö $2^{(3x+1)}=8$ Kirjoita välivaiheet. Vastaus:
38. M20S6T19_095AB 12	Hotellissa yöpyy eräänä yönä ihmisiä eri maista. Yöpyjistä kolme on suomalaisia, viisi saksalaisia, yksi italialainen, seitsemän ruotsalaisia, neljä venäläisiä, kaksi virolaisia, seitsemän ranskalaisia, kaksi kanadalaisia, yhdeksän puolalaisia ja viisi norjalaisia. Jokainen vieras majoittuu omassa huoneessaan. Ranskalaisten lisäksi hotellin vieraista ranskaa puhuvia ovat yksi kanadalainen, kaksi puolalaista ja yksi saksalainen. Millä todennäköisyydellä satunnaisesti valittu hotellivieras puhuu ranskaa? Muodosta lauseke ja anna vastaus prosentin tarkkuudella. a) Lauseke: b) Vastaus:
39. MYk18S3T14_191A-F 6	Yhdistä xy-tason sinisellä varjostettuun alueeseen sitä vastaava epäyhtälö.  Ympyröi **A)** $x > -y + 2$ **B)** $x < -2x$ **C)** $y > x - 4$ **D)** $x < -2$ **E)** $y > 2$ **F)** $x + y > 0$ **G)** $2x + y > 0$ **H)** $y < 3x + 4$ **I)** $x + y < 2$

Pisteitysohjeet

KUUNNELTAVIEN PÄÄSSÄLASKUJEN 1–2 VASTAUKSET:

M20S1T10_003

1. 5 cm (1p)

M2S1T10_002

2. 21 litraa (1p)

PÄÄSSÄLASKUJEN 3–6 VASTAUKSET:

M20S1T10_011

3. 3,09 (1p)

M20S2T10_028A

4. 32 (1p)

M12S1T10_163

5. 8 min (1p)

M20S1T10_012DEF 2

6. 14 kpl (1p)

B-OSA TEHTÄVIEN 7–30 VASTAUKSET:

M15S4T15_174

7. c) (1p)

M20S5T16_074 12356

8. 1. ei (1p) 2. ei (1p) 3. kyllä (1p) 4. ei (1p) 5. ei (1p)

M20S2T12_027 A-G 5

9. c) (1p)

M20S2T13_027 A-G 6

10. e) (1p)

M20S6T19_105

11. 1. c) (1p), 2. b) (1p)

M20S5T18_080 257

12. 1. ei (1p), 2. ei (1p), 3. kyllä (1p)

M20S1T10_014

13. D (1p)

M11S6T19_184

14. c) (1p)

M15S5T16_179

15. c) (1p)

M20S4T15_064 145

16. 1. väärin (1p), 2. oikein (1p), 3. oikein (1p)

M20S6T19_093

17. 9 (1p)

M12S2T11_170A

18. -13 (1p)

M20S2T13_025

19. c) (1p)

M20S3T14_048B

20. b) (1p)

M12S5T16_177

21. d) (1p)

M20S1T20_045D 1

22. c) (1p)

M20S3T14_052 13

23. 1. ei (1p), 2. kyllä (1p)

M20S1T20_008 12

24. 1. F5 (1p), 1. F4 (1p)

M12S5T17_178

25. c) (1p)

M15S4T15_183A 1

26. 80 km/h (hyväksytään 75km/h – 85km/h), (1p)

M20S3T14_046 A-F_BDE

27. 1. c) (1p), 2. b) (1p), 3. d) (1p)

M20S1T10_019

28. 81 (1p)

M20S5T17_084

29. 5,0 cm (1p)

M20S2T13_103

30. 4,07/2,55 = 1,596… = 160 %, (1p)

C-OSA TEHTÄVIEN VASTAUKSET:

MYk19S2T11_193 A-F 13

31. 1. a_n=(1,3,5,7…) (1p), 2. a_n= (1, 8, 27, 64…) (1p)

M00S3T14_172B

32. min 0 p., maks 2 p.

    Yhtälön ratkaisun periaate oikein (esimerkiksi lisätty ja vähennetty sekä saatu 5x − 3x = 7. (1 p.)
    Ratkaisu oikein ($x = 3\frac{1}{2}$), hyväksytään myös x = 3,5 (+1 p.)
    Vastauksen pyöristäminen (-1 p.)
    ”Merkkivirhe” ratkaisussa (-2 p.)

M20S4T15_061

33. Hyväksyttäviä perusteluja:

    Vakio c kertoo
    funktion pienimmän arvon
    tai
    kuvaajan matalimman kohdan y-koordinaatin
    tai
    kuvaajan ja y-akselin leikkauspisteen y-koordinaatin. (1p)

MYs18S5T18_201

34. Monikulmion pinta-ala ja perustelut:

    Huom. Useita ratkaisuvaihtoehtoja. Alla yksi.
    Laatta sisältyy suorakulmioon, jonka pinta-ala on $6 \cdot 9 = 54$.
    ⇒ Laatan pinta-ala saadaan vähentämällä tästä kolmioiden alat.
    Kolmioiden alat ovat 3, 3, 4, ja 1, joiden summa on 11.
    Tulos on $54 - 11 = 43$.
    Pisteitys: maks 3p

    - perustelu, miten on laskenut (esim. jakanut kolmioihin) (1p)
    - välivaiheiden laskut oikein (1p)
    - oikea vastaus 43 pinta-alayksikköä (1p)

M20S4T14_070 12

35. 1. A taksi: $f(x)= 3x+4{,}5$ (1p) 2. B taksi: $h(x)=5x+1{,}5$ (1p)

MYs17S3T14_196 B

36. Onnistuttu saamaan yhden muuttujan yhtälö, esim. sijoittamalla tai laskemalla yhteen

    $\Rightarrow y= \frac{11}{5}$ ja $x= \frac{16}{5}$

    maks 2p

    - muokattu yhtälö ratkaistavaan muotoon (1p)
    - oikea vastaus (1p)

MYs17S3T14_196 C

37. $8=2^3$

    $3x + 1 = 3$ eli $x= \frac{2}{3}$

    maks 2p

    - perustelu/laskutoimitus (1p)
    - oikea vastaus (1p)

M20S6T19_095AB 12

38. a) lauseke: P(ranskaa puhuvat) = 11/45 = 0,24444… (1p)

    b) vastaus: 24 % (1p)

MYk18S3T14_191A-F 6

39. D (1p)

1.2 Svensk version

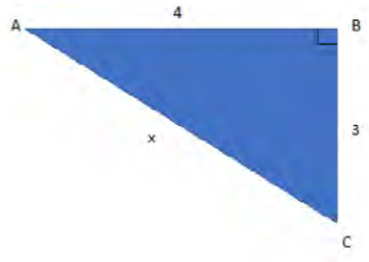
UTVÄRDERINGEN AV LÄRRESULTATEN I MATEMATIK I ÅK 9 (NCU 2021)

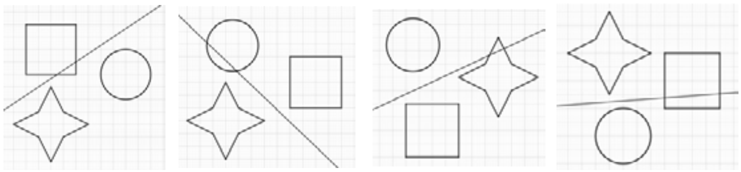
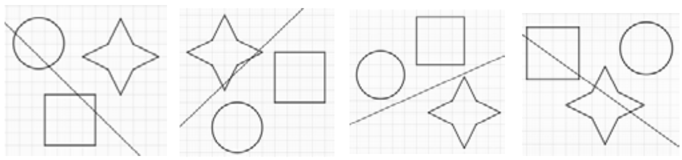
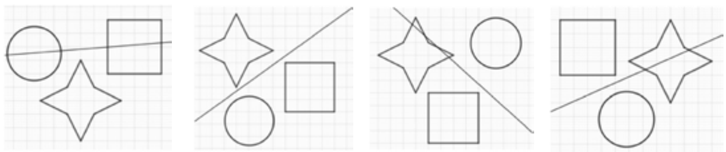
A-del: Huvudräkningar

	HUVUDRÄKNINGAR 1–2 Du hör varje fråga 2 gånger och du har en minut att besvara.	
1. M20S1T10_003	Läraren läser / Hörs från ljudfilen (två gånger): *En kvadrat har arean 25 cm². Hur lång är kvadratens sida?*	Svar: ________________
2. M20S1T10_002	Läraren läser / Hörs från ljudfilen (två gånger): *Familjen Lindholm förbrukar en och en halv liter mjölk om dagen. Hur många liter mjölk går det åt hos Lindholms under två veckor?*	Svar: ________________

	HUVUDRÄKNINGAR 3–6 Du har 4 minuter att besvara. Du kan inte använda hjälpmedel (till exempel räknare eller anteckningsmedel).	
3. M20S1T10_011	Beräkna 4,37−1,28	Svar: ________________
4. M20S2T10_028A	Beräkna värdet $2 \cdot 4^2$	Svar: ________________
5. M12S1T10_163	Då jag promenerar hem från skolan med hastigheten 5 km/h tar hemfärden 24 minuter. Hur lång tar hemfärden om jag cyklar med medelhastigheten 15 km/h?	Svar: ________________
6. M20S1T10_012DEF 2	I ett klassrum finns stolar och pallar. Alla åtta pallar har tre ben. Pallarna har sammanlagt lika många ben som stolarna, som alla har fyra ben. Hur många pallar och stolar finns det totalt i klassrummet?	Svar: ________________

B-del: Flervalsuppgifter och korta svar

7. M15S4T15_174	Priset på äpplen kan beräknas med funktionen $f(x)=2x+1$ där x är mängden äpplen (kg). Hur mycket kostar 3,5 kg äpplen?	a) 24,5 € b) 9 € c) 8 € d) 7 €
8. M20S5T16_074 12356	 Vilka påståenden stämmer? Märka J=ja eller N=nej 1. Triangeln ABC är trubbvinklig. ________ 2. Triangeln är symmetrisk i förhållande till punkten B. ________ 3. Triangeln är rätvinklig. _____ 4. Vinkeln A är 60 grader. ______ 5. Triangelns area är 12 arealenheter. _______	
9. M20S2T12_027 A-G 5	Förenkla 8:24	a) 1/20 b) 1/5 c) 1/3 d) 0,45 e) 4,5
10. M20S2T13_027 A-G 6	Ett lag har spelat 20 % av sina matcher oavgjort och förlorat 25 % av matcherna. Hur många procent av matcherna har laget vunnit?	a) 25 % b) 10 % c) 30 % d) 50 % e) 55 %
11. M20S6T19_105	Frekvensen (f) anger hur många gånger en observation förekommer i ett statiskt material. Den relativa frekvensen (f %) anger hur många procentandelar av en viss observation är jämfört av alla observationer. En klass har efter ett prov i matematik följande vitsord som statistisk material: 8, 7, 7, 7, 10, 6, 6, 8, 8, 8, 10, 9, 8, 7, 9, 9, 6, 5, 5, 7.	
	1. Vilken är frekvensen för vitsordet 7?	a) 2 b) 3 c) 5 d) 6
	2. Vilken är relativa frekvensen för vitsordet 7?	a) 25 % b) 20 % c) 50 % d) 10 %
12. M20S5T18_080 257	Kan man med hjälp av följande information beräkna triangelns areal? Märka J=ja eller N=nej 1. Två vinklar ______ 2. Tre vinklar ______ 3. Två sidor och den mellanliggande vinkeln ____	

13. M20S1T10_014	Följande figurer uppfyller vissa obekanta villkor.  Följande figurer uppfyller <u>inte</u> de okända villkoren.  Vilken av följande figurer uppfyller de okända villkoren? Utmärk.  A B C D	
14. M11S6T19_184	Katrin har följande spelmärke i en påse, och påsen har skakats om.  Katrin lyfter ett spelmärke ur påsen utan att titta. Vilken är sannolikheten att talet på spelmärket är delbar med tre?	a) 1/11 b) 1/3 c) 4/11 d) 1/4 e) 4/7
15. M15S5T16_179	En rak vinkel består av tre vinklar, *a, a* och 50 grader. Hur stor är vinkeln *a*?	a) 20 b) 55 c) 65 d) 75
16. M20S4T15_064 145	Nina badar. Efter badet drar hon ur proppen och vattnet börjar rinna ner i avloppet. Funktionen $f(x)=-40x+180$ beskriver hur mycket vatten det finns kvar i badkaret när det har gått x minuter från det att hon dragit ur proppen. Vilka påståenden stämmer? Märk J=ja eller F=fel 1. När Nina badade fanns det 140 l vatten i badkaret. _______ 2. Badkaret töms på 4 min och 30 s. _______ 3. En minut efter att Nina drog ur proppen finns det 140 l vatten i badkaret. _______	

17.

M20S6T19_093

Samuel söker till utbildning på andra stadiet. Han söker till en utbildning där antagningsgränsen för medeltalet på avgångsbetyget varit 8,2 de senaste åren. Han har sammanställt sina vitsord i en tabell. Vilket är det lägsta vitsordet Samuel borde ha i matematik för att nå upp till antagningsgränsen?

Läsämnen	Vitsord
svenska	8
engelska	8
finska	7
biologi	9
geografi	7
religion	8
historia	10
matematik	
kemi	9
fysik	7
samhällslära	9

Svar: ____

18.

M12S2T11_170A

Vilket svar ger uttrycket $2-5\cdot2-5$?

Svar: ______________________

19.

M20S2T13_025

Kevin köper en jacka. Vad kostar jackan med rabatt?

a) 89,70 €
b) 199,33 €
c) 209,30 €
d) 269 €
e) 388,70 €

20.

M20S3T14_048B

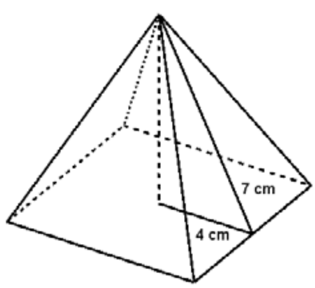
Beräkna värdet för polynomet x^2+6x-8 när $x=3$. Välj rätt svar.

a) 16
b) 19
c) 35
d) 64

21.

M12S5T16_177

Vilket av påståendena om pyramiden är sann?

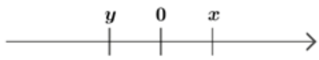


a) Bottnen är en rätvinklig triangel.
b) Höjden är 7 *cm*
c) Den kvadratiska basytans sida är 4 *cm* lång
d) Den kvadratiska basytans areal är 64 cm^2
e) Sidoytans areal är 56 cm^2

<table>
<tr>
<td>22.

M20S1T20_045D 1</td>
<td>En hurdan figur skapar koden?

</td>
<td>a) en triangel
b) en kvadrat
c) en rektangel som inte är en kvadrat
d) en sexhörning
e) ingen av ovanstående</td>
</tr>
<tr>
<td>23.

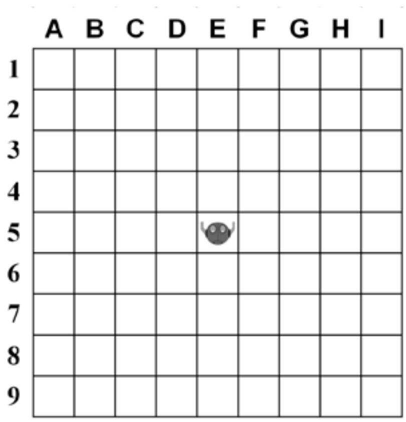
M20S3T14_052 13</td>
<td colspan="2">

Studera tallinjen. Stämmer påståenden?
Märk J=ja eller N=nej

a) $x - y = y - x$ _____

b) $x - y > y - x$ ______</td>
</tr>
<tr>
<td>24.

M20S1T20_008 12</td>
<td colspan="2">En programmerbar robot får en serie kommandon som den utför i en bestämd ordning. Möjliga kommandon är F=”gå en ruta framåt”, B=”gå en ruta bakut”, H=”vänd 90 grader åt höger” och V=”vänd 90 grader åt vänster”

För kommandona använder vi förkortningarna F/B/H/V. Roboten rör sig i ett 9x9-rutor stort rutnät som finns avbildat här nedanför.


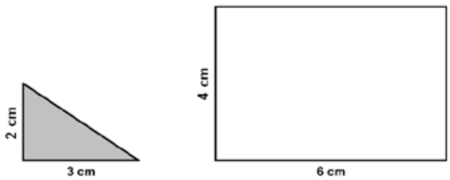
a) Roboten startar i ruta E5 med blicken vänd uppåt. Den får kommandoserien ”HFFBV”.
I vilken ruta stannar roboten?

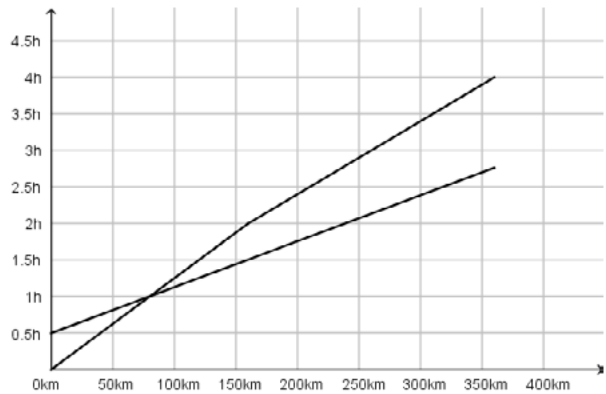
Svar: ____________

b) Roboten startar i ruta E5 med blicken vänd uppåt. Den får kommandoserien ”HFFVFVFVVH”.
I vilken ruta stannar roboten?

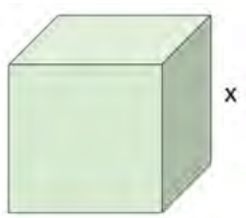
Svar: _____________</td>
</tr>
</table>

<table>
<tr><td rowspan="1">25.

M12S5T17_178</td><td>Nedan finns en grå, rätvinklig triangel. Hur många dylika trianglar skulle du behöva för att kunna täcka rektangelns areal exakt?

</td><td>a) fyra
b) sex
c) åtta
d) tio
e) tolva</td></tr>
<tr><td>26.

M15S4T15_183A 1</td><td colspan="2">En bilist börjar köra från Seinäjoki till Helsingfors. En halvtimme senare avgår ett Pendolino-tåg från Seinäjoki till Helsingfors. Följande figur visar sträckorna som bilen och tåget förflyttar sig. Med vilken hastighet kör bilen de första två timmarna? Bestäm med hjälp av figuren.


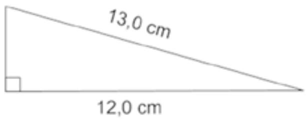
Svar: ____</td></tr>
<tr><td rowspan="4">27.

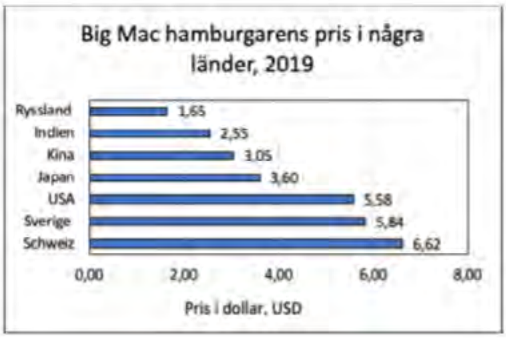
M20S3T14_046
A-F_BDE</td><td colspan="2">En kub har kantlängden x.

</td></tr>
<tr><td>1. Vilken areal har en sidoyta?</td><td>a) $2x^2$
b) $4x$
c) x^2
d) $12x$
e) $6x^2$</td></tr>
<tr><td>2. Vilken är omkretsen för en sidoyta?</td><td>a) $2x^2$
b) $4x$
c) x^2
d) $12x$
e) $6x^2$</td></tr>
<tr><td>3. Vilken är kanternas sammanlagda längd?</td><td>a) $2x^2$
b) $4x$
c) x^2
d) $12x$
e) $6x^2$</td></tr>
</table>

<table>
<tr><td>28.

M20S1T10_019</td><td colspan="2">Vilket är det följande talet i talföljden?

0, 1, 1, 2, 4, 7, 13, 24, 44, _____</td></tr>
<tr><td>29.

M20S5T17_084</td><td>Beräkna katetens längd.

</td><td>Svar:
_______________</td></tr>
<tr><td>30.

M20S2T13_103</td><td>Hur många procent dyrare är en Big Mac hamburgare i Schweiz jämfört med i Indien? Ge svaret i procent med en procents precision.

</td><td>Svar:
_______________</td></tr>
</table>

C-del: Svårare problemlösningsuppgifter

<table>
<tr><td>31.

MYk19S2T11_193
A-F 13</td><td>Förena följande talföljd med rätt definition.

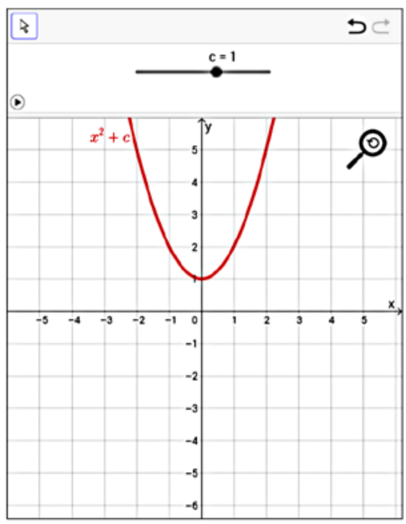
1. $a_n=2n-1$

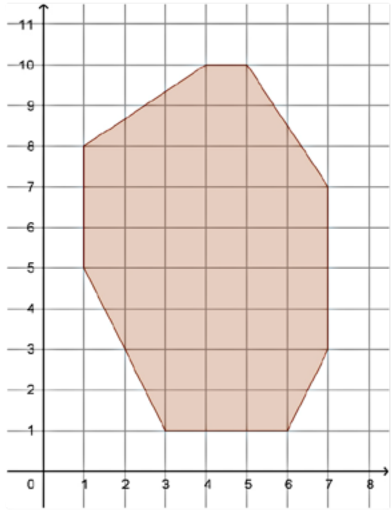
2. $a_n=n^3$</td><td>a) (1, 8, 27, 64, …)
b) (1, 2, 3, 4, …)
c) (1, 2, 3, 5, …)
d) (1, 4, 9, 16, …)
e) (1, 3, 5, 7, …)</td></tr>
<tr><td>32.

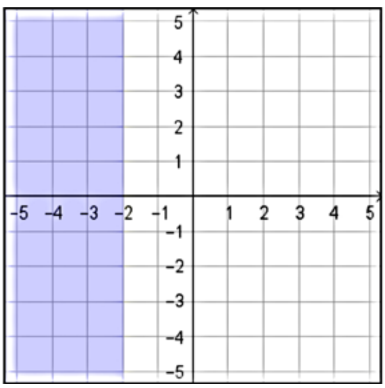
M00S3T14_172B</td><td colspan="2">Lös ekvationen $5x-7=3x$.

Motivera.</td></tr>
<tr><td>33.

M20S4T15_061</td><td colspan="2">Beskriv hur grafen för funktionen $f(x)=x^2+c$ förändras när värdet på konstanten c varieras.

Svar:

</td></tr>
</table>

34. MYs18S5T18_201	Tävlingsjuryn för en formgivningstävling belönar de bästa bidragen med en minnestavla. Juryns konstnärliga assistent gör den första versionen av en miniatyrmodell av tavlan genom att använda ett koordinatsystem i programmet GeoGebra. Hon utfår från en rektangel, vars bredd är 6 och höjd är 9 längdenheter. Genom att på olika sätt skära bort rektangelns alla fyra hörn får hon månghörningen i figuren invid som resultat. Beräkna arealen av denna månghörning. Motivera ditt svar. Svar: 
35. M20S4T14_070 12	Det kostar 3 euro per kilometer att anlita Taxibolag A och dessutom en grundavgift på 4,50 €. Det kostar 5 euro per kilometer att anlita Taxibolag B och dessutom en grundavgift på 1,50 €. Bilda funktionerna som beskriver prissättningen för taxibolagen (variabeln *x* motsvarar antalet kilometer). a) Taxi A b) Taxi B
36. MYs17S3T14_196 B	Lös ekvationsparet. $\begin{cases} y - x + 1 = 0 \\ 4y = 12 - x. \end{cases}$ Skriv ut mellansteg. Svar:

37. MYs17S3T14_196 C	Lös ekvationen $2^{(3x+1)}=8$ Skriv ut mellansteg. Svar:
38. M20S6T19_095AB 12	En natt övernattar personer från olika länder på ett hotell. Hotellgästerna består av tre finländare, fem tyskar, en italienare, sju svenskar, fyra ryssar, två estländare, sju fransmän, två kanadensare, nio polacker och fem norrmän. Alla hotellgäster övernattar i ett eget rum. Bland hotellgästerna finns det förutom fransmännen en kanadensare, två polacker och en tysk som talar franska. Med vilken sannolikhet kan man välja en slumpmässig gäst som talar franska? Bilda uttryck och svara med en procents precision. a) Uttryck: b) Svar:
39. MYk18S3T14_191A-F 6	Det blå området i xy-planet motsvaras av en olikhet. Välj den olikhet som motsvarar området.  Märk rätt alternativ. **A)** $x > -y + 2$ **B)** $x < -2x$ **C)** $y > x - 4$ **D)** $x < -2$ **E)** $y > 2$ **F)** $x + y > 0$ **G)** $2x + y > 0$ **H)** $y < 3x + 4$ **I)** $x + y < 2$

Poängsättningsanvisningar

HUVUDRÄKNINGARNA 1–2:

M20S1T10_003

1. 5 cm (1p)

M2S1T10_002

2. 21 liter (1p)

HUVUDRÄKNINGARNA 3–6:

M20S1T10_011

3. 3,09 (1p)

M20S2T10_028A

4. 32 (1p)

M12S1T10_163

5. 8 min (1p)

M20S1T10_012DEF 2

6. 14 st (1p)

B-DEL: UPPGIFTER 7–30

M15S4T15_174

7. c) (1p)

M20S5T16_074 12356

8. 1. nej (1p) 2. nej (1p) 3. ja (1p) 4. nej (1p) 5. nej(1p)

M20S2T12_027 A-G 5

9. c) (1p)

M20S2T13_027 A-G 6

10. e) (1p)

M20S6T19_105

11. 1. c) (1p), 2. b) (1p)

M20S5T18_080 257

12. 1. nej (1p), 2. nej (1p), 3. ja (1p)

M20S1T10_014

13. D (1p)

M11S6T19_184

14. c) (1p)

M15S5T16_179

15. c) (1p)

M20S4T15_064 145

16. 1. fel (1p), 2. rätt (1p), 3. rätt (1p)

M20S6T19_093

17. 9 (1p)

M12S2T11_170A

18. -13 (1p)

M20S2T13_025

19. c) (1p)

M20S3T14_048B

20. b) (1p)

M12S5T16_177

21. d) (1p)

M20S1T20_045D 1

22. c) (1p)

M20S3T14_052 13

23. 1. nej (1p), 2. ja (1p)

M20S1T20_008 12

24. 1. F5 (1p), 1. F4 (1p)

M12S5T17_178

25. c) (1p)

M15S4T15_183A 1

26. 80 km/h (accepteras 75km/h – 85km/h), (1p)

M20S3T14_046 A-F_BDE

27. 1. c) (1p), 2. b) (1p), 3. d) (1p)

M20S1T10_019

28. 81 (1p)

M20S5T17_084

29. 5,0 cm (1p)

M20S2T13_103

30. 4,07/2,55 = 1,596… = 160 %, (1p)

C-DEL: UPPGIFTER 31–39

MYk19S2T11_193 A-F 13

31. 1. a_n=(1,3,5,7…) (1p), 2. a_n= (1, 8, 27, 64…) (1p)

M00S3T14_172B

32. min 0 p., maks 2 p.

    Ekvationen löst rätt (t.ex. addition och subtraktion med resultat 5x − 3x = 7 el. dyl.. (1 p.)
    Rätt löst (x = $3\frac{1}{2}$), även x = 3,5 godkänns (+1 p.)
    Svaret avrundat (-1 p.)
    Teckenfel i lösningen (-2 p.)

M20S4T15_061

33. godkända motivationer

    Konstanten c beskriver
    funktionens minimivärde
    eller
    den lägsta punkten på y-axeln
    eller
    y-koordinaten för skärningspunkten mellan grafen och y-axeln (1 p.)

MYs18S5T18_201

34. Det finns många olika sätt att lösa uppgiften. Nedan är en av dem.

    Minnestavlan ingor i en rektangel med en yta på 6 · 9 = 54.
    ⇒ Plattans area erhålls genom att subtrahera trianglarnas ytor från detta.
    Ytorna av trianglar är 3, 3, 4, ja 1, som summas 11.
    Resultatet är 54 − 11 = 43.

    maks 3 p

    - 1 p för motivering hur hen räknat (t.ex. delat in i trianglar)
    - 1 p för korrekta mellansteg
    - 1 p för korrekt svar, 43 (arealenheter)

M20S4T14_070 12

35. 1. A taxi: $f(x)= 3x+4,5$ (1p) 2. B taxi: $h(x)=5x+1,5$ (1p)

MYs17S3T14_196 B

36. Studerande har lyckats med att ta fram en ekvation av en variabel, t.ex. genom substitutions- eller additionsmetoden.

    $\Rightarrow y= \frac{11}{5}$ och $x= \frac{16}{5}$

    maks 2p

    - 1 p. ekvationen har bearbetats till lösbar form
    - 1 p. korrekt svar

MYs17S3T14_196 C

37. $8=2^3$

    $3x + 1 = 3$ eller $x= \frac{2}{3}$

    maks 2p

    - 1p motiv/räkneoperation
    - 1p korrekt svar

M20S6T19_095AB 12

38. a) uttryck: P(fransktalande) = 11/45 = 0,24444… (1 p)

    b) svar: 24 % (1 p)

MYk18S3T14_191A-F 6

39. D (1p)

1.3 English version

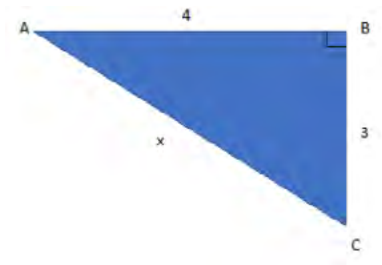
COMPREHENSIVE EDUCATION GRADE 9 ASSESSMENT OF THE LEARNING OUTCOMES IN MATHEMATICS 2021 (FINEEC 2021)

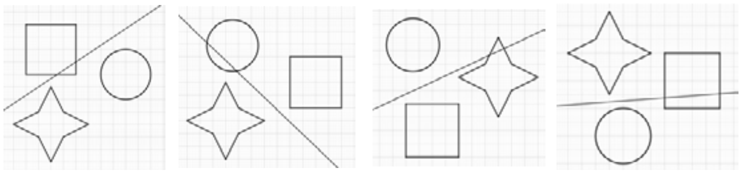
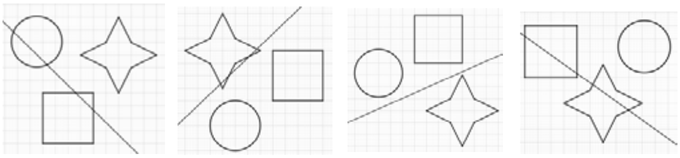
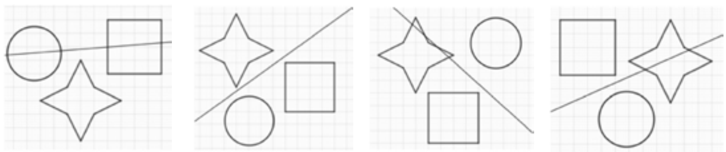
Part A: Mental calculations

	MENTAL CALCULATIONS 1–2 You will hear each question 2 times, after which you have 1 minute to answer it.	
1. M20S1T10_003	Teacher reads/Heard from an audiofile (2 times) *The area of a square is 25 cm². How long is one edge?*	Answer: ________
2. M20S1T10_002	Teacher reads/Heard from an audiofile (2 times) *The family Virtanen consumes 1.5 liters of milk every day. How many liters of milk does family Virtanen consume in two weeks?*	Answer: ________

	MENTAL CALCULATIONS 3–6 Total time for calculations 4 minutes. No aids can be used. In this digital test, you cannot use pencil, paper, nor calculator.	
3. M20S1T10_011	Calculate 4,37−1,28	Answer: ________
4. M20S2T10_028A	Write the simplest form of the expression $2 \cdot 4^2$	Answer: ________
5. M12S1T10_163	It takes me 24 minutes, when I walk home from school with a speed of 5 km/h. How long does my journey home take when I cycle it with a speed of 15 km/h?	Answer: ________
6. M20S1T10_012DEF 2	In the classroom there are chairs and stools. There are eight three-legged stools. Those have altogether as many legs as the four-legged chairs have. How many stools and chairs are there altogether?	Answer: ________

Part B: Multiple choice and short answers

7. M15S4T15_174	The price of apples can be obtained by calculating the value of the function $f(x)=2x+1$, where x is the amount of apples (kg). How much does 3.5kg of apples cost?	a) 24,5 € b) 9 € c) 8 € d) 7 €
8. M20S5T16_074 12356	 Are the statements true? Mark on the line Y= yes or N= no 1. Triangle ABC is an acute-angled triangle. _______ 2. Triangle is symmetrical about the point B._______ 3. Triangle is a right triangle. ____ 4. The size of the angle A is 60 degrees. _____ 5. The area of the triangle is 12 units of area. ______	
9. M20S2T12_027 A-G 5	Simplify 8:24	a) 1/20 b) 1/5 c) 1/3 d) 0,45 e) 4,5
10. M20S2T13_027 A-G 6	20% of the games the team has played have been drawn matches, and the team has lost 25% of the games. What percentage of the games has the team won? Circle.	a) 25 % b) 10 % c) 30 % d) 50 % e) 55 %
11. M20S6T19_105	Frequency (f) determines how many times does the finding occur in the statistic. Relative frequency (f%) informs how much is the finding out of all the data as percentage. The Math grades of a class were 8, 7, 7, 7, 10, 6, 6, 8, 8, 8, 10, 9, 8, 7, 9, 9, 6, 5, 5, 7.	
	1. What is the frequency of the grade 7? Circle	a) 2 b) 3 c) 5 d) 6
	2. What is the relative frequency of the grade 7? Circle.	a) 25 % b) 20 % c) 50 % d) 10 %
12. M20S5T18_080 257	Can we calculate the area of the triangle when we have the following information? Mark on the line Y= yes or N= no 1. Two angles _____ 2. Three angles _____ 3. Two sides and the angle between _____	

<table>
<tr>
<td>13.
M20S1T10_014</td>
<td colspan="2">The following figures fulfill certain unknown conditions

The following figures do not fulfill these unknown conditions

Which of the following figures fulfill the unknown conditions? Circle.

A B C D</td>
</tr>
<tr>
<td>14.
M11S6T19_184</td>
<td>Katri has the chips shown below in her pouch and the pouch has been shaken


Katri picks up one chip without looking. What is the probability that the chip has a number which is divisible by three?</td>
<td>a) 1/11
b) 1/3
c) 4/11
d) 1/4
e) 4/7</td>
</tr>
<tr>
<td>15.
M15S5T16_179</td>
<td>A straight angle is made of three angles a, a, and 50 degrees. How big is the angle a? Circle.</td>
<td>a) 20
b) 55
c) 65
d) 75</td>
</tr>
<tr>
<td>16.
M20S4T15_064 145</td>
<td colspan="2">Nina is taking a bath. After the bath she removes the plug, and the water starts running to the drains.

The function $f(x)=-40x+180$ describes how much water there is in the tub, when x minutes have passed from the moment of removing the plug.

Is the statement true (T) or false (F)? Mark on the T or F.

1. There is 140 liters of water in the tub, when Nina was taking the bath. ______
2. The tub is empty after 4min 30s. _______
3. After one minute of the removal of the plug there is 140 liters of water in the tub. _______</td>
</tr>
</table>

17.

M20S6T19_093

Samuel is applying for secondary education. He has collected all his grades in the table. He wants to get into a school, where the mean of academic subjects has been 8.2 for several years. Which grade should Samuel get from Math in order for him to reach this requirement of the mean?

Subject	Grade
mother tongue	8
English	8
Swedish	7
biology	9
geography	7
religion	8
history	10
mathematics	
chemistry	9
physics	7
social studies	9

Answer: ____

18.

M12S2T11_170A

What is the solution of the expression 2−5·2−5?

Answer: _______________________

19.

M20S2T13_025

Kalle buys a jacket. What does the jacket cost after the reduction?

a) 89,70 €
b) 199,33 €
c) 209,30 €
d) 269 €
e) 388,70 €

20.

M20S3T14_048B

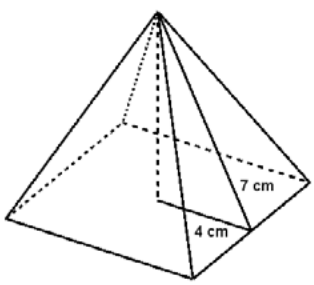
Calculate the value of the polynomial x^2+6x-8 when $x=3$. Circle the correct solution.

a) 16
b) 19
c) 35
d) 64

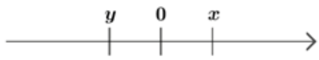
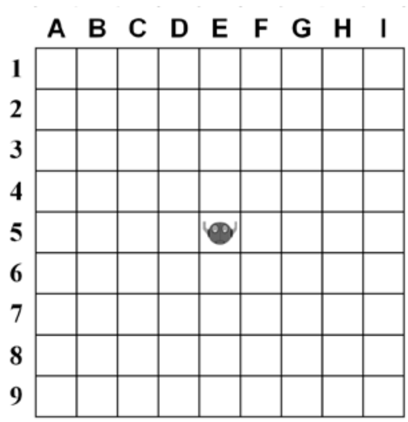
21.

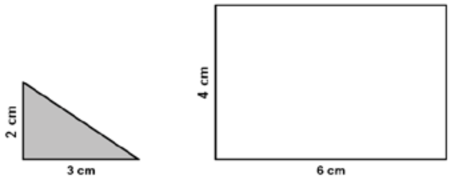
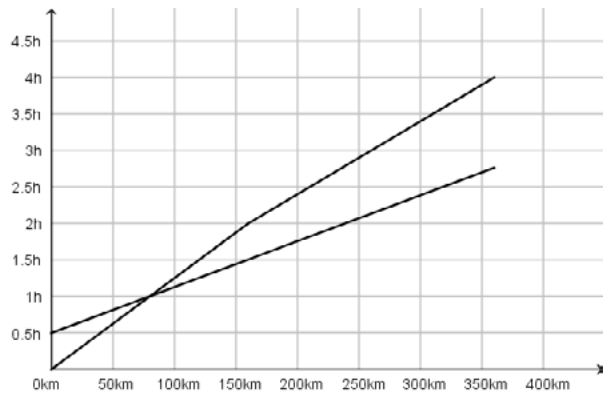
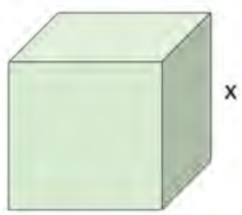
M12S5T16_177

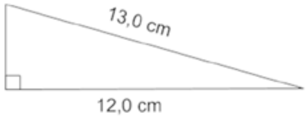
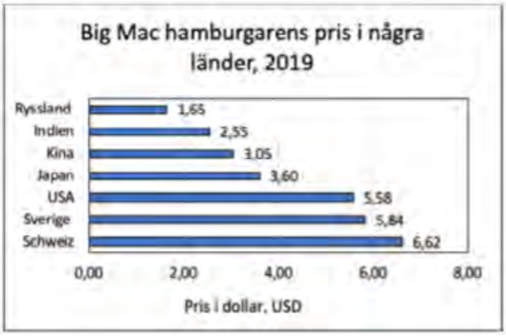
Which of the statements are true? Circle.



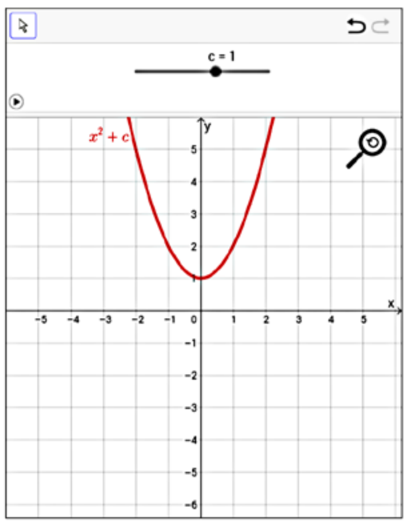
a) The base of the pyramid is a right triangle.
b) The height of the pyramid is 7 *cm*
c) The length of the edge of the square base is 4 *cm*
d) The area of the square base is 64 cm^2
e) The area of the lateral face of the pyramid is 56 cm^2

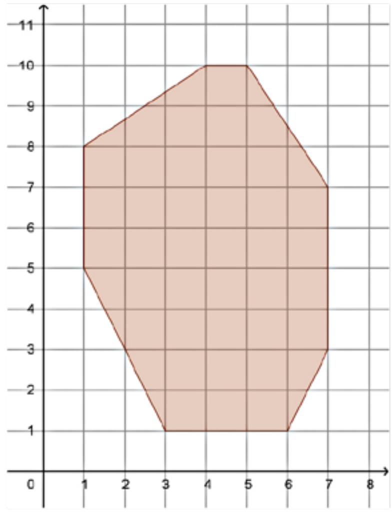
22. M20S1T20_045D 1	What kind of figures is formed with the following piece of code?  	a) triangle b) square c) rectangle, which is not a square d) hexagon e) none of the above ones
23. M20S3T14_052 13	 Study the number line. Are the following statements true? Mark on the line after the statement Y= yes or N= no. a) $x - y = y - x$ _____ b) $x - y > y - x$ _______	
24. M20S1T20_008 12	The following commands are given to the robot which is being programmed and it carries out the commands in the given order. The possible commands are E="forward one block", T="backwards one block", O="right turn 90 degrees" and V="left turn 90 degrees". Let us use the abbreviations E/T/O/V. The robot is moving inside a 9 x 9 -grid, which is shown below  a) The robot starts from the block E5, facing upwards. It is given the commands "OEETV". In which block does the robot stop? Solution: _____________ b) The robot starts from the block E5, facing upwards. It is given the commands "OEEVEVEVVO". In which block does the robot stop? Solution: ______________	

25. M12S5T17_178	How many of the grey-colored right-angled triangles shown below do we need to fill exactly the area of the given rectangle? 	a) four b) six c) eight d) ten e) twelve
26. M15S4T15_183A 1	A driver starts their journey from Seinäjoki to Helsinki. After half an hour a pendolino train leaves from Seinäjoki towards Helsinki. In the graph below the journeys of the car and train are shown. Estimate by studying the graph: What is the average speed of the car for the first two hours?  Answer: ___________________________________	
27. M20S3T14_046 A-F_BDE	The edge of a cube is x. 	
	1. What is the area of the face?	a) $2x^2$ b) $4x$ c) x^2 d) $12x$ e) $6x^2$
	2. What is the perimeter of one face?	a) $2x^2$ b) $4x$ c) x^2 d) $12x$ e) $6x^2$
	3. What is the total length of all the edges?	a) $2x^2$ b) $4x$ c) x^2 d) $12x$ e) $6x^2$

28. M20S1T10_019	What is the following term of the given number pattern? 0, 1, 1, 2, 4, 7, 13, 24, 44, _____	
29. M20S5T17_084	Calculate the length of the cathetus (leg). 	Answer: _______________
30. M20S2T13_103	What percentage more expensive is the Big Mac hamburger in Switzerland (Sveitsi) than in India (Intia)? Give your answer with the accuracy of one percent. 	Answer: _______________

Part C: More demanding problemsolving tasks

31. MYk19S2T11_193 A-F 13	Connect the number pattern to its general term/definition of the patters. 1. $a_n=2n-1$ 2. $a_n=n^3$	a) (1, 8, 27, 64, …) b) (1, 2, 3, 4, …) c) (1, 2, 3, 5, …) d) (1, 4, 9, 16, …) e) (1, 3, 5, 7, …)
32. M00S3T14_172B	Solve the equation $5x-7=3x$. Justify.	
33. M20S4T15_061	Explain, how does the change of the constant c chance the path of the function $f(x)=x^2+c$. Answer: 	

34. MYs18S5T18_201	The award committee of a design competition grants a memorial plaque to the best pieces of the competition. The artistic assistant of the committee makes the first version of the model of this plaque using Geogebra-program's coordinate system. They start with a rectangle that has a width of 6 units and height of 9 units. By cutting off differently all the four corners of this rectangle they end up having the next polygonal figure shown. Determine the area of this polygon and explain/justify your solution. Answer: 
35. M20S4T14_070 12	Taxi A costs 3 euros for each kilometer and it has an initial fee of 4.50 euros. Taxi B on the other hand costs 5 euros per kilometer and it has an initial fee of 1.50 euros. Form functions for the prizing of the taxi rides (mark kilometers with the variable *x*). a) Taxi A b) Taxi B
36. MYs17S3T14_196 B	Solve the pair of simultaneous equations. $\begin{cases} y - x + 1 = 0 \\ 4y = 12 - x. \end{cases}$ Solution

<table>
<tr><td>37.

MYs17S3T14_196 C</td><td>Solve the equation $2^{(3x+1)}=8$

Solution.</td></tr>
<tr><td>38.

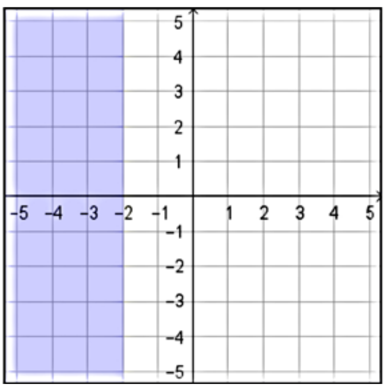
M20S6T19_095AB 12</td><td>People from different countries are guests in a hotel for an overnight stay. Three of the guests are Finns, five Germans, one Italian, seven Swedes, four Russians, two Estonians, seven French, two Canadians, nine Poles and five Norwegians. Each guest stays in their own room.

In addition to the French guests, there are also other French-speaking guests, of which one is Canadian, two are Polish and one is German. What is the probability that a randomly picked hotel guest speaks French? Form an expression and give the answer to the nearest percent.

a) Expression:

b) Solution:</td></tr>
<tr><td>39.

MYk18S3T14_191A-F 6</td><td>Connect the correct inequality which defines the blue-colored area marked on the xy-plane.



Circle.

A) $x > -y + 2$ B) $x < -2x$ C) $y > x - 4$
D) $x < -2$ E) $y > 2$ F) $x + y > 0$
G) $2x + y > 0$ H) $y < 3x + 4$ I) $x + y < 2$</td></tr>
</table>

Scoring keys

ANSWERS TO THE TASKS 1 JA 2

M20S1T10_003

1. 5 cm (1p)

M2S1T10_002

2. 21 liter (1p)

ANSWERS TO THE TASKS 3–6

M20S1T10_011

3. 3,09 (1p)

M20S2T10_028A

4. 32 (1p)

M12S1T10_163

5. 8 min (1p)

M20S1T10_012DEF 2

6. 14 pcs (1p)

B-PART ANSWERS FOR TASKS 7–30

M15S4T15_174

7. c) (1p)

M20S5T16_074 12356

8. 1. no (1p) 2. no (1p) 3. yes (1p) 4. no (1p) 5. no (1p)

M20S2T12_027 A-G 5

9. c) (1p)

M20S2T13_027 A-G 6

10. e) (1p)

M20S6T19_105

11. 1. c) (1p), 2. b) (1p)

M20S5T18_080 257

12. 1. no (1p), 2. no (1p), 3. yes (1p)

M20S1T10_014

13. D (1p)

M11S6T19_184

14. c) (1p)

M15S5T16_179

15. c) (1p)

M20S4T15_064 145

16. 1. wrong (1p), 2. right (1p), 3. right (1p)

M20S6T19_093

17. 9 (1p)

M12S2T11_170A

18. -13 (1p)

M20S2T13_025

19. c) (1p)

M20S3T14_048B

20. b) (1p)

M12S5T16_177

21. d) (1p)

M20S1T20_045D 1

22. c) (1p)

M20S3T14_052 13

23. 1. no (1p), 2. yes (1p)

M20S1T20_008 12

24. 1. F5 (1p), 1. F4 (1p)

M12S5T17_178

25. c) (1p)

M15S4T15_183A 1

26. 80 km/h (acceptable 75km/h – 85km/h), (1p)

M20S3T14_046 A-F_BDE

27. 1. c) (1p), 2. b) (1p), 3. d) (1p)

M20S1T10_019

28. 81 (1p)

M20S5T17_084

29. 5,0 cm (1p)

M20S2T13_103

30. 4,07/2,55 = 1,596… = 160 %, (1p)

C-PART CORRECT ANSWERS FOR ITEMS 31–39:

MYk19S2T11_193 A-F 13

31. 1. a_n=(1,3,5,7…) (1p), 2. a_n= (1, 8, 27, 64…) (1p)

M00S3T14_172B

32. min 0 p., max. 2 p.

    The principle of the solution of the equation is corrected (e.g., added and substracted and got $5x - 3x = 7$. (1 p.)
    Solution correct ($3\frac{1}{2}$), also $x = 3{,}5$ is acceptable (+1 p.)
    Rounding of the answer (-1 p.)
    ”sign error” in the solution (-2 p.)

M20S4T15_061

33. Acceptable answers:

    Constant c indicates
    the minimum value of the function
    or
    the y-coordinate of the lowest point of the graph
    or
    the y-coordinate of the graph of the function and y-axis. (1p)

MYs18S5T18_201

34. Area of polygon and justifications:

    Several options to solve. One example is given below:
    The plate is embedded to a rectangle of which area is. The area of rectangle is $6 \cdot 9 = 54$.
    Subtract the areas of the triangles from the area of the plate.
    Areas of the triangles 3, 3, 4, ja 1 summing up to 11.
    Answer is 54 – 11 = 43.
    Marks: max. 3p

    - justify the calculations, (for example. divided to triangles) (1p)
    - intermediate calculations correct (1p)
    - correct answer 43 units of area (1p)

M20S4T14_070 12

35. 1. A taxi: $f(x)= 3x+4,5$ (1p) 2. B taxi: $h(x)=5x+1,5$ (1p)

MYs17S3T14_196 B

36. Formed an equation with one variable by, for example, using the addition method or subtraction method.

    $\Rightarrow y= \frac{11}{5}$ and $x= \frac{16}{5}$

    Max. 2p

    - equation modified to solvable form (1p)
    - correct answer (1p)

MYs17S3T14_196 C

37. $8=2^3$

    $3x + 1 = 3$ or $x = \frac{2}{3}$

    Max. 2p

    - justification/calculation (1p)
    - correct answer (1p)

M20S6T19_095AB 12

38. Max. 2p.

    a) expression: P(speak French) = 11/45 = 0,24444.. (1p)

    b) correct answer: 24 % (1p)

MYk18S3T14_191A-F 6

39. D (1p)

LIITE 2. Osioparametrit

	tehtävän mekaaniset ominaisuudet			vaikeustaso- ja erottelukykyparametrejä					validiteettiparametreja			IRT parametreja (vaikeustaso)					
Nimi	Järjestys	Testin osa (A, B, C)	Maksimi-pisteet	ratkaisu-osuus (p)	Ri	D	G	keski-hajonta	Sisältö-alue (S1-S6)	Tavoite (T10-T20)	Ajattelun taso (H1- H4)	B:1	B:2	B:3	SE(B):1	SE(B):2	SE(B):3
M20S1T10_003 1 (1)	1	A1	1	0,561	0,528	0,623	0,635	0,496	S1, S5	T10, T18	H2	-0,486			0,021		
M20S1T10_002 1 (1)	2	A1	1	0,661	0,419	0,515	0,526	0,474	S1, S2	T10	H2	-0,982			0,036		
M20S1T10_011 1 (1)	3	A2	1	0,765	0,376	0,516	0,527	0,424	S1, S2	T10, T11	H2	-1,598			0,040		
M20S2T10_028ABCD 1 (1)	4	A2	1	0,318	0,268	0,318	0,325	0,466	S1, S2	T10	H3	0,810			0,033		
M12S1T10_163 1 (1)	5	A2	1	0,506	0,498	0,579	0,591	0,500	S1, S2	T10, T11	H3	0,035			0,018		
M20S1T10_012DEF 1 (2)	6	A2	1	0,591	0,518	0,619	0,631	0,492	S1, S2	T10, T11	H2	-0,630			0,031		
M15S4T15_174 1	7	B	1	0,654	0,472	0,581	0,593	0,476	S4	T15	H2	-0,982			0,023		
M20S5T16_074A-F 1	8.1	B	1	0,905	0,262	0,503	0,514	0,293	S5	T16, T17, T18	H4	-2,871			0,033		
M20S5T16_074A-F 2	8.2	B	1	0,624	0,267	0,312	0,319	0,484	S5	T16, T17, T18	H4	-0,860			0,021		
M20S5T16_074A-F 3	8.3	B	1	0,899	0,359	0,681	0,695	0,302	S5	T16, T17, T18	H4	-2,785			0,032		
M20S5T16_074A-F 5	8.4	B	1	0,664	0,359	0,437	0,447	0,472	S5	T16, T17, T18	H4	-1,116			0,022		
M20S5T16_074A-F 6	8.5	B	1	0,676	0,361	0,439	0,450	0,468	S5	T16, T17, T18	H4	-1,068			0,022		
M20S2T13_027A-G 5	9	B	1	0,878	0,374	0,658	0,671	0,327	S2	T13, T12	H3	-2,066			0,038		
M20S2T13_027A-G 6	10	B	1	0,916	0,349	0,716	0,730	0,277	S2	T13, T12	H3	-2,975			0,058		
M20S6T19_105 1	11.1	B	1	0,877	0,418	0,731	0,744	0,329	S6	T19, T13	H2	-2,647			0,030		
M20S6T19_105 2	11.2	B	1	0,709	0,423	0,539	0,551	0,454	S6	T19, T13	H2	-1,428			0,023		
M20S5T18_080A-H 2	12.1	B	1	0,731	0,324	0,417	0,427	0,444	S5	T18	H2	-1,401			0,023		
M20S5T18_080A-H 5	12.2	B	1	0,483	0,460	0,531	0,542	0,500	S5	T18	H2	-0,138			0,021		
M20S5T18_080A-H 7	12.3	B	1	0,758	0,390	0,532	0,544	0,428	S5	T18	H2	-1,640			0,024		
M20S1T10_014 1	13	B	1	0,483	0,447	0,516	0,527	0,500	S1	T10	H3	-0,177			0,021		
M11S6T19_184 1	14	B	1	0,730	0,493	0,649	0,662	0,444	S6, S2	T19, T11	H3	-1,145			0,036		
M15S5T16_179 1	15	B	1	0,632	0,514	0,626	0,638	0,482	S5	T16	H2	0,884			0,035		
M20S4T15_064A-E 1	16.1	B	1	0,568	0,342	0,388	0,397	0,495	S4	T15	H3	-0,538			0,021		
M20S4T15_064A-E 4	16.2	B	1	0,692	0,186	0,211	0,216	0,462	S4	T15	H3	-1,189			0,022		
M20S4T15_064A-E 5	16.3	B	1	0,479	0,352	0,397	0,406	0,500	S4	T15	H3	-0,038			0,021		
M20S6T19_093 1 (1)	17	B	1	0,517	0,386	0,446	0,456	0,500	S6	T19	H3	-0,295			0,031		
M12S2T11_170A 1 (1)	18	B	1	0,685	0,416	0,519	0,531	0,465	S2	T11	H2	-0,823			0,030		
M20S2T13_025 1	19	B	1	0,675	0,408	0,507	0,518	0,469	S2	T13	H3	-1,043			0,035		

	tehtävän mekaaniset ominaisuudet			vaikeustaso- ja erottelukykyparametrejä					validiteettiparametreja			IRT parametreja (vaikeustaso)					
Nimi	**Järjestys**	**Testin osa (A, B, C)**	**Maksimi-pisteet**	**ratkaisu-osuus (p)**	**Ri**	**D**	**G**	**keski-hajonta**	**Sisältö-alue (S1-S6)**	**Tavoite (T10-T20)**	**Ajattelun taso (H1- H4)**	B:1	B:2	B:3	SE(B):1	SE(B):2	SE(B):3
M20S3T14_048B 1	20	B	1	0,806	0,482	0,714	0,727	0,396	S3	T14	H2	-1,861			0,037		
M12S5T16_177 1	21	B	1	0,528	0,517	0,602	0,614	0,499	S5	T16	H4	-0,268			0,029		
M20S1T20_045A-D (D)	22	B	1	0,571	0,454	0,531	0,543	0,495	S1, S5	T20, T16	H2	-0,509			0,035		
M20S3T14_052 1	23.1	B	1	0,606	0,453	0,536	0,548	0,489	S3, S1	T14, T10	H3	-0,693			0,036		
M20S3T14_052 3	23.2	B	1	0,599	0,426	0,499	0,510	0,490	S3, S1	T14, T10	H3	-0,657			0,036		
M20S1T20_008ABC 1 (1)	24.1	B	1	0,607	0,527	0,632	0,644	0,488	S1	T20	H2	-0,867			0,021		
M20S1T20_008ABC 1 (2)	24.2	B	1	0,476	0,540	0,630	0,642	0,500	S1	T20	H2	-0,212			0,021		
M12S5T17_178 1	25	B	1	0,515	0,476	0,552	0,563	0,500	S5	T17	H3	-0,928			0,023		
M15S4T15_183A 1 (1)	26	B	1	0,392	0,487	0,579	0,590	0,488	S4	T15	H2	0,319			0,021		
M20S3T14_046A-F_B 1	27.1	B	1	0,512	0,554	0,646	0,658	0,500	S3, S5	T14, T18	H1	-0,277			0,021		
M20S3T14_046A-F_D 1	27.2	B	1	0,591	0,569	0,678	0,691	0,492	S3, S5	T14, T18	H1	-0,667			0,031		
M20S3T14_046A-F_E 1	27.3	B	1	0,571	0,455	0,529	0,541	0,495	S3, S5	T14, T18	H1	-0,509			0,035		
M20S1T10_019 1 (1)	28	B	1	0,214	0,429	0,585	0,597	0,410	S1	T10	H3	1,362			0,025		
M20S5T17_084 1 (1)	29	B	1	0,360	0,564	0,675	0,687	0,480	S5	T17	H2	0,490			0,022		
M20S2T13_103 1 (1)	30	B	1	0,046	0,312	0,767	0,778	0,210	S6, S2	T13	H2	3,330			0,045		
MYk19S2T11_193A-F 1	31.1	C	1	0,269	0,275	0,339	0,347	0,443	S2	T11	H2	1,193			0,024		
MYk19S2T11_193A-F 3	31.2	C	1	0,368	0,390	0,456	0,466	0,482	S2	T11	H2	0,583			0,022		
M00S3T14_172B	32	C	2	0,400	0,634	0,640	0,653	0,858	S3	T14	H3	0,385	0,157		0,042	0,047	
M20S4T15_061	33	C	1	0,197	0,499	0,702	0,714	0,398	S4	T15	H4	1,686			0,038		
MYs18S5T18_201	34	C	3	0,155	0,616	0,767	0,779	0,916	S5	T18, T17	H3	2,174	0,717	1,285	0,051	0,069	0,073
M20S4T14_070 1 (1)	35.1	C	1	0,327	0,571	0,701	0,713	0,469	S4	T15, T14	H3	0,752			0,037		
M20S4T14_070 1 (2)	35.2	C	1	0,313	0,561	0,696	0,708	0,464	S4	T15, T14	H3	0,829			0,037		
MYs17S3T14_196B	36	C	2	0,073	0,410	0,701	0,712	0,461	S2	T14	H3	2,759	1,168		0,039	0,056	
MYs17S3T14_196C	37	C	2	0,032	0,300	0,736	0,746	0,322	S2	T14	H3	4,021	0,996		0,063	0,086	
M20S6T19_095AB 1 (1)	38.1	C	1	0,063	0,261	0,589	0,601	0,243	S6	T19	H3	3,280			0,044		
M20S6T19_095AB 1 (2)	38.2	C	1	0,086	0,319	0,623	0,635	0,280	S6	T19	H3	2,944			0,039		
MYk18S3T14_191A-F 6	39	C	1	0,240	0,250	0,316	0,323	0,427	S3	T14	H2	1,252			0,024		

Selitteitä:
Testin osa: A = päässälaskut, B = monivalinta-, yhdistämis- ja lyhytvastaukset, C = ongelmaratkaisutehtävät
Erottelukyvyn indikaattorit: osio–summa korrelaatiot R = Pearson korrelaatio, D = Somerin delta, G = Goodmanin–Kruskalin gamma
Sisältöalueet: S1 = Ajattelun taidot ja menetelmät, S2 = Luvut ja laskutoimitukset, S3 = Algebra, S4 = Funktiot, S5 = Geometria, S6 = Tietojen käsittely ja tilastot sekä todennäköisyys
Tavoitteet: [Oppilas…] T10 = osaa tehdä matemaattisia päätelmiä, laskea päässä ja käyttää taitojaan eri konteksteissa, T11 = osaa peruslaskutoimituksia rationaaliluvuilla, T12 = ymmärtää

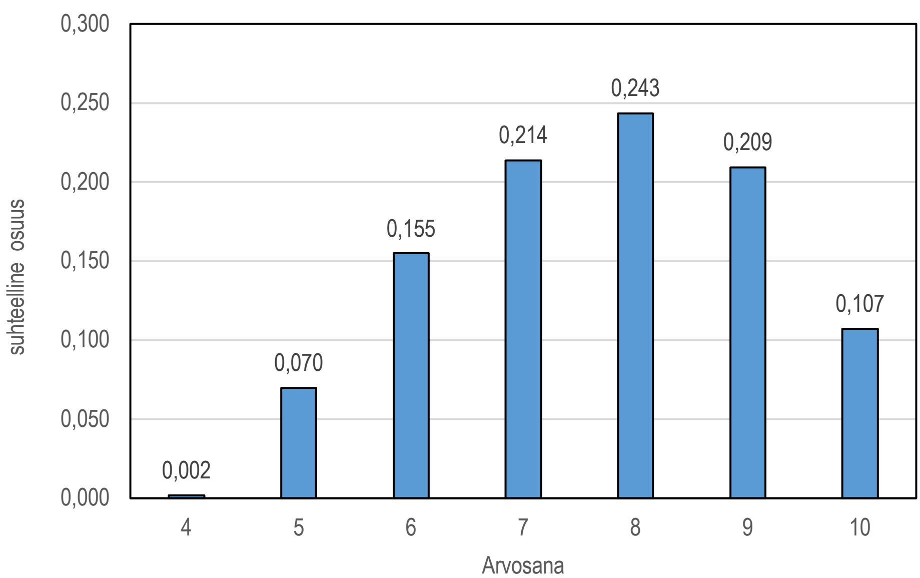
reaaliluvun käsitteen, T13 = ymmärtää prosentin käsitteen ja osaa laskea prosenttiosuuden, prosenttiluvun osoittaman määrän, muutos- ja vertailuprosentin, T14 = ymmärtää tuntemattoman käsitteen ja osaa ratkaista yhtälöitä, T15 = ymmärtää funktion käsitteen, osaa tulkita funktion kuvaajia ja osaa tuottaa funktion kuvaajia, T16 = ymmärtää geometrian käsitteitä ja geometristen käsitteiden välisiä yhteyksiä, T17 = ymmärtää suorakulmaiseen kolmioon ja ympyrään liittyviä ominaisuuksia sekä osaa hyödyntää suorakulmaiseen kolmioon ja ympyrään liittyviä ominaisuuksia, T18 = osaa laskea pinta-aloja ja tilavuuksia, T 19 = osaa määrittää tilastollisia tunnuslukuja ja laskea todennäköisyyksiä, T20 = osaa ajatella algoritmisesti, käyttää algoritmisia ja matemaattisia taitojaan ongelmien ratkaisemiseen sekä ratkaista ongelmia myös ohjelmoiden.
Ajattelun hierarkkiset tasot: H1 = muistaminen (recall), H2 = asian hallinta (comprehension), H3 = soveltaminen (application), H4 = korkeammat taidot (higher skills) ml. analysointi, syntetisointi, arviointi, luominen

LIITE 3. Julkaistuun testisarjaan liittyvä arvosanaehdotus ja sen perustelu

3.1 Arvosanaehdotuksen taustaa

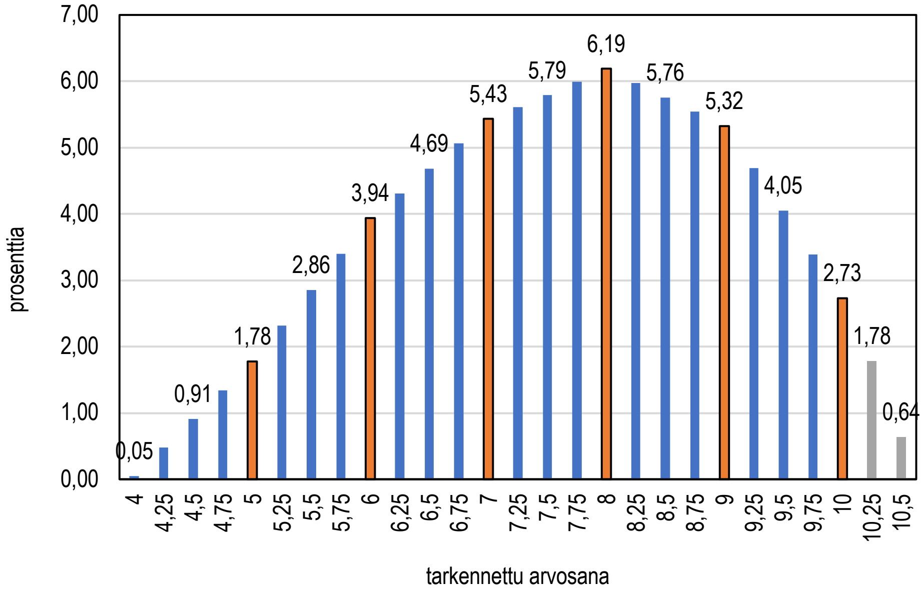
Julkaisun testisarjan arvosanaehdotus perustuu oppilaan saamaan kokonaispistemäärään testissä. Erittäin pienen pistemäärän (alle 10 pistettä) ja erittäin korkean pistemäärän (yli 50 pistettä) saaminen testissä on hyvin harvinaista. Arvosanaehdotus on muodostettu otosoppilaiden (n = 12 829) viimeisimmän Koski-rekisterissä olleet arvosanan jakauman perusteella seuraavasti:

1. Koski-rekisteristä saatiin perusarvosanoille 4–10 prosentuaaliset jakaumat niille oppilaille, jotka olivat kokonaisotoksessa (n = 12 829). Jakaumaa hieman korjattiin, sillä tosiallisessa jakaumassa arvosana 5 näytti pyöristyneet usein 6:ksi ja arvosanaa 10 näytettiin pelätyn käyttää ja näin arvosana 9 osuus oli hieman odotusta suurempi. Näitä suhteellisia osuuksia käytetään populaatiojakauman estimaatteina (Kuvio 1).



KUVIO 1. Otosoppilaiden matematiikan arvosanojen korjattu jakauma Koski-rekisterissä 2020 (n = 12 829)

2. Väliarvosanojen 4,5 ja 5,5, … ja 9,5 laskennalliset prosentuaaliset jakaumat intrapoloitiin perusarvosanojen väliin näiden keskiarvona. Väliarvosanoja tarkemmat arvosanat 4,25 ja 4,75, 5,25 ja 5,75, …. 9,25 ja 9,75 laskennalliset prosentuaaliset osuudet intrapoloitiin puolikkaiden ja perusarvosanojen väliin näiden keskiarvona. Perusarvosanojen ja näiden väliin jäävien väliarvosanojen jakauma vastaa muodon osalta kokonaisarvosanojen jakaumaa, mutta on tätä yksityiskohtaisempi (Kuvio 2).



KUVIO 2. Tarkennettu arvosanajakauma Koski-rekisterin perusteella

3. Laskettiin kunkin arvosanan portaan laskennallinen kumulatiivinen prosenttijakauma, jota käytetään suoraan arvosanaehdotuksen muodostamisessa.
4. Julkaistavan version summapistemäärän jakauman perusteella (n = 4 451) etsittiin ne kumulatiivisen jakauman pisteet, joissa arvosana vaihtuu. Näin eri arvosanaluokkaa kuuluvien pistemäärien vaihteluväli voi poiketa toisistaan. Esimerkiksi arvosanaehdotusta 5- (4,75) vastaa kansallisessa aineistossa pistemäärät 3–10, koska näitä pistemääriä sai hyvin harva oppilas; arvauskynnys saa aikaan sen, että tosiasiallisesti laskenta alkaa vasta 11 pisteestä ja tästä ylöspäin. Tätä pienemmät arvosanat ovat harvinaisia kansallisessa aineistossa.
5. Kaksi erityisluokkaa luotiin arvosanan 10 yläpuolelle, 10+ ja 10,5, jotka heijastavat poikkeuksellisen erinomaista suoritusta. Tämän kaltainen suoritus on syntynyt niin, että oppilas on joko saanut kaikki tai lähes kaikki tehtävät oikein. Jos oppilas on saanut arvosanaehdotuksekseen 10,5, hän kuuluu harvinaiseen 0.64 prosentin joukkoon, joka otoksessa selvisi testistä poikkeuksellisen hyvin.

3.2 Arvosanaehdotus

Taulukkoon 1 on koottu arvosanaehdotus tilanteessa, että oppilas on tehnyt julkaistun testin valvotuissa olosuhteissa.

TAULUKKO 1. Arvosanaehdotus julkaistuissa tehtäväsarjassa

pistemäärä tehtäväsarjassa	arvosana-ehdotus	huomioita
0	4	Pistemäärää 3 pienempiä pistemääriä on vaikea saada testissä; pistemäärä 0−2 osoittaa erittäin heikkoa näytettyä osaamista
1	4,25	
2	4,5	
3−10	4,75	Pistemäärän 10 jälkeen osaamisessa ilmenee suuri loikka; on huomattavasti vaikeampaa saada testissä 11 pistettä kuin 10 tai tätä vähemmän. Tämä johtuu arvauskynnyksestä.
11	5	Todellinen osaaminen alkaa 11 pisteestä
12−13	5,25	
14	5,5	
15−16	5,75	
17	6	
18−19	6,25	
20	6,5	
21−22	6,75	
23	7	
24−25	7,25	
26−27	7,5	
28−29	7,75	
30−32	8	
33−34	8,25	
35−36	8,5	
37−38	8,75	
39−40	9	
41−42	9,25	
43−45	9,5	
46−47	9,75	
48−50	10	
51−55	10,25	Testissä on mukana joukko erittäin vaikeita tehtäviä. Yli 50 pisteen saaminen on erittäin harvinaista ja osoittaa poikkeuksellista matemaattista kyvykkyyttä 9. luokalla.
56−60	10,5	

Taulukkoon 2 on koottu arvio yksittäisten osioiden linkittymisestä arvosanaan. Taustalla käytetyn IRT-mallituksen keskeinen oletus on, että testattavan osaamisparametri (θ) ja osion vaikeusparametri (β) ovat identtisiä, kun testattavalla on 50/50 todennäköisyys saada tehtävä suoritettua oikein. Kun siis tiedetään, mikä on testipisteeseen (ja näin ollen arvosanaan) liittyvä θ-arvo, voidaan karkeasti arvioida mikä olisi se arvosana, jonka saanut oppilas saisi tehtävän oikein 50%:n todennäköisyydellä (p = 0.50). Tämänkaltainen tieto ei ole tietenkään tarkka, mutta saattaa helpottaa opettajaa ja testien kehittäjiä arvioimaan karkeasti, minkä tyyppinen saattaisi olla esimerkiksi arvosanaa 8 vastaava tehtävä kansallisen aineiston perusteella 9. luokan matematiikan osaamisen arvioinnissa. Opettaja tai testin kehittäjä voisi siis muokata tehtävän vaikeustasoa vastaavan *uuden* tehtävä ja arvioida, että ko. uuden tehtävän vaikeustaso on likipitäen sama kuin nyt julkaistun tehtävän. Tämä kaltainen tieto on tarkentuvaa ja on mahdollista, että tulevissa arvioinneissa tarjotaan asiaan lisää tietoa. Myös käytetyt IRT-mallit saattavat tuottaa toistaan hieman poikkeavia tuloksia. Taulukot 1 ja 2 ovat ensimmäisiä laatuaan Karvin arvioinneissa ja ne on tehty yksiparametrisella mallilla, ns. Raschin mallilla.

TAULUKKO 2. Yksittäisiin osioihin liittyvä arvosanaehdotus

Numero	Osion nimi	Järjestys	B-parametri[1]	Arvosana[2]
1	M20S1T10_003 1 (1)	1	-0,486	7,25
2	M20S1T10_002 1 (1)	2	-0,982	6,5
3	M20S1T10_011 1 (1)	3	-1,598	5,75
4	M20S2T10_028ABCD 1 (1)	4	0,81	9
5	M12S1T10_163 1 (1)	5	0,035	8
6	M20S1T10_012DEF 1 (2)	6	-0,63	7,25
7	M15S4T15_174 1	7	-0,982	6,5
8	M20S5T16_074A-F 1	8.1	-2,871	4,75
9	M20S5T16_074A-F 2	8.2	-0,86	6,75
10	M20S5T16_074A-F 3	8.3	-2,785	4,75
11	M20S5T16_074A-F 5	8.4	-1,116	6,25
12	M20S5T16_074A-F 6	8.5	-1,068	6,25
13	M20S2T13_027A-G 5	9	-2,066	5
14	M20S2T13_027A-G 6	10	-2,975	4,75
15	M20S6T19_105 1	11.1	-2,647	4,75
16	M20S6T19_105 2	11.2	-1,428	6
17	M20S5T18_080A-H 2	12.1	-1,401	6
18	M20S5T18_080A-H 5	12.2	-0,138	7,75
19	M20S5T18_080A-H 7	12.3	-1,64	5,5
20	M20S1T10_014 1	13	-0,177	7,75
21	M11S6T19_184 1	14	-1,145	6,25
22	M15S5T16_179 1	15	0,884	9
23	M20S4T15_064A-E 1	16.1	-0,538	7,25
24	M20S4T15_064A-E 4	16.2	-1,189	6,25
25	M20S4T15_064A-E 5	16.3	-0,038	8
26	M20S6T19_093 1 (1)	17	-0,295	7,5
27	M12S2T11_170A 1 (1)	18	-0,823	7
28	M20S2T13_025 1	19	-1,043	6,5

Numero	Osion nimi	Järjestys	B-parametri[1]	Arvosana[2]
29	M20S3T14_048B 1	20	-1,861	5,25
30	M12S5T16_177 1	21	-0,268	7,75
31	M20S1T20_045A-D (D)	22	-0,509	7,25
32	M20S3T14_052 1	23.1	-0,693	7,25
33	M20S3T14_052 3	23.2	-0,657	7,25
34	M20S1T20_008ABC 1 (1)	24.1	-0,867	6,75
35	M20S1T20_008ABC 1 (2)	24.2	-0,212	7,75
36	M12S5T17_178 1	25	-0,928	6,75
37	M15S4T15_183A 1 (1)	26	0,319	8,25
38	M20S3T14_046A-F_B 1	27.1	-0,277	7,5
39	M20S3T14_046A-F_D 1	27.2	-0,667	7,25
40	M20S3T14_046A-F_E 1	27.3	-0,509	7,25
41	M20S1T10_019 1 (1)	28	1,362	9,5
42	M20S5T17_084 1 (1)	29	0,49	8,5
43	M20S2T13_103 1 (1)	30	3,33	10
44	MYk19S2T11_193A-F 1	31.1	1,193	9,25
45	MYk19S2T11_193A-F 3	31.2	0,583	8,75
46	M00S3T14_172B	32	0,385	8,25
47	M20S4T15_061	33	1,686	9,75
48	MYs18S5T18_201	34	2,174	10
49	M20S4T14_070 1 (1)	35.1	0,752	9
50	M20S4T14_070 1 (2)	35.2	0,829	9
51	MYs17S3T14_196B	36	2,759	10
52	MYs17S3T14_196C	37	4,021	10
53	M20S6T19_095AB 1 (1)	38.1	3,28	10
54	M20S6T19_095AB 1 (2)	38.2	2,944	10
55	MYk18S3T14_191A-F 6	39	1,252	9,5

1) polytomisilla osioilla 1:n pisteen saamisen B-arvo
2) karkea ehdotus perustuen 1-parametriseen IRT-malliin

Tähän raporttiin on koottu vuoden 2021 matematiikan oppimistulosarvioinnin menetelmälliset ratkaisut. Erillisissä artikkeleissa kuvataan yleiset menetelmälliset ratkaisut, digitaalisen testialustan ominaisuudet ja siihen liittyviä haasteita, ja mittauksissa käytetyn uuden tunnemittarit ominaisuudet. Lisäksi tutkitaan mekanismeja, joilla kansallinen osaamisen taso on laskenut viimeisten vuosien aikana ja erityisesti, kun on siirrytty paperi-kynä-testauksesta digitaaliseen testaukseen.

Mittaus oli monella tavalla aiempia arviointeja laajempi ja monipuolisempi. Tavanomaiseen tapaan osaamista mittaavat mittarit ovat osuvia (valideja) ja tarkkoja (reliaabeleja) antamaan uskottava kuva matematiikan osaamisesta perusopetuksen päättövaiheessa.

Kansallinen koulutuksen arviointikeskus (Karvi) on itsenäinen koulutuksen arviointiviranomainen. Se toteuttaa koulutukseen sekä opetuksen ja koulutuksen järjestäjien toimintaan liittyviä arviointeja varhaiskasvatuksesta korkeakoulutukseen. Lisäksi arviointikeskus toteuttaa perusopetuksen ja toisen asteen koulutuksen oppimistulosten arviointeja. Keskuksen tehtävänä on myös tukea opetuksen ja koulutuksen järjestäjiä ja korkeakouluja arviointia ja laadunhallintaa koskevissa asioissa sekä kehittää koulutuksen arviointia.

ISBN 978-952-206-787-6 pdf
ISSN 2342-4184 (verkkojulkaisu)

Kansallinen koulutuksen arviointikeskus
PL 380
(Hakaniemenranta 6)
00531 Helsinki
Puhelinvaihde: 029 533 5500
karvi.fi