KANSALLINEN
KOULUTUKSEN
ARVIOINTIKESKUS

ALKUMITTAUKSEN MENETELMÄLLISIÄ RATKAISUJA

Metsämuuronen Jari
Ukkola Annette

Julkaisut 18:2019

ALKUMITTAUKSEN MENETELMÄLLISIÄ RATKAISUJA

Metsämuuronen Jari
Ukkola Annette

KARVI

Kansallinen koulutuksen arviointikeskus
Julkaisut 18:2019

JULKAISIJA Kansallinen koulutuksen arviointikeskus

KANSI JA ULKOASU Juha Juvonen (org.) & Sirpa Ropponen (edit)
TAITTO PunaMusta

ISBN 978-952-206-546-9 nid.
ISBN 978-952-206-547-6 pdf

ISSN 2342-4176 (painettu)
ISSN 2342-4184 (verkkojulkaisu)
ISSN 2342-4176

PAINATUS PunaMusta Oy, Tampere

© Kansallinen koulutuksen arviointikeskus

Tiivistelmä

Julkaisija
Kansallinen koulutuksen arviointikeskus (KARVI)

Julkaisun nimi
Alkumittauksen menetelmällisiä ratkaisuja

Tekijät
Metsämuuronen Jari ja Ukkola Annette

Kansallinen koulutuksen arviointikeskus (Karvi) järjesti syksyllä 2018 oppimistulosten arvioinnin perusopetuksen 1. vuosiluokan oppilaille. Tästä arvioinnista käytettiin nimitystä alkumittaus. Sen tarkoituksena oli tuottaa kansallista tietoa siitä, mitä oppilaat osaavat, kun he aloittavat perusopetuksen ensimmäisen luokan. Alkumittaus toimi samalla lähtötasomittauksena perusopetuksen oppimistulosten pitkittäisarvioinnissa. Pitkittäisarvioinnin tarkoituksena on arvioida oppilaiden osaamisen kehittymistä peruskoulun aikana kahdessa keskeisessä oppiaineessa, *matematiikassa* ja *äidinkielessä ja kirjallisuudessa* perusopetuksen opetussuunnitelman perusteiden tavoitteiden mukaan. Tässä raportissa kuvataan hankkeen menetelmällisiä ratkaisuja.

Kaikkiaan koulun alkuvaiheen osaamista kartoittava tiedonkeruu on validiteetiltaan ja reliabiliteetiltaan riittävä luotettavien johtopäätösten tekemiseen kansallisella tasolla. Oppilasaineistosta saatava tieto oppilaiden osaamisesta koulun alkuvaiheessa on hyvin yleistettävissä koko populaatioon erityisesti, kun käsitellään painotettuja tuloksia ja tunnuslukuja. Sen sijaan on syytä olla enemmän varovainen, kun selitetään osaamisen eroja huoltaja-aineiston perusteella. Huoltaja-aineiston vastaajien lapset edustavat nimittäin selvästi paremmin menestyneitä oppilaita kuin vastaamatta jättäneiden huoltajien lapset. Kokonaisosaamisen keskiarvo oli 516 niillä oppilailla, joiden huoltajilta saatiin taustatietoja kyselyllä. Niillä oppilailla, joiden huoltajat jättivät vastaamatta kyselyyn, kokonaisosaamisen keskiarvo oli 480.

Avainsanat: alkuopetus, arviointi, lukutaito, matematiikka, oppimistulokset, perusopetus, äidinkieli ja kirjallisuus, arviointimenetelmät

Sammanfattning

Utgiven av
Nationella centret för utbildningsutvärdering (NCU)

Publikationens namn
Alkumittauksen menetelmällisiä ratkaisuja

Författare
Jari Metsämuuronen och Annette Ukkola

Nationella centret för utbildningsutvärdering (NCU) anordnade hösten 2018 för första gången en utvärdering av lärresultat för elever i årskurs 1 i den grundläggande utbildningen. För denna utvärdering används benämningen inledande mätning. Dess syfte var att producera nationell information om vad eleverna kan när de börjar i årskurs 1. Den inledande mätningen startade den longitudinella utvärderingen av lärresultat i den grundläggande utbildningen. Syftet med den longitudinella utvärderingen är att utvärdera utvecklingen av elevernas kunskaper och färdigheter under den grundläggande utbildningen i två centrala läroämnen, *matematik* och *modersmål och litteratur* enligt målen i Grunderna för läroplanen för den grundläggande utbildningen 2014. Elevernas kunskaper utvärderas i årskurs tre, sex och nio. I denna rapport presenteras de metodlösningar för den utvärderingen.

Allt som allt är validiteten och reliabiliteten för den datainsamling som kartlägger kunskaper i det skede då man börjar skolan tillräckliga för att tillförlitliga slutsatser ska kunna dras på nationell nivå. Den information som elevmaterialet ger om elevernas kunskaper då de börjar skolan kan väl generaliseras för att gälla för hela populationen, särskilt då man behandlar viktade resultat och nyckeltal. Däremot finns anledning att vara mer försiktig då man förklarar skillnader i kunskaper utifrån vårdnadshavarmaterialet. Barnen till respondenterna i vårdnadshavarmaterialet representerar nämligen elever med klart bättre framgång jämfört med elever till vårdnadshavare som inte svarat på enkäten. Medeltalet för de övergripande kunskaperna var 516 hos de elever vars vårdnadshavare gav bakgrundsinformation genom enkäten. Medeltalet för de övergripande kunskaperna var 480 hos de elever vars vårdnadshavare inte svarade på enkäten.

Nyckelord: grundläggande utbildning, lärresultat, läsfärdighet, matematik, modersmål och litteratur, nybörjarundervisning, utvärdering, metoder

Abstract

Publisher
Finnish Education Evaluation Centre (FINEEC)

Title of publication
Alkumittauksen menetelmällisiä ratkaisuja

Authors
Jari Metsämuuronen och Annette Ukkola

In autumn 2018, the Finnish Education Evaluation Centre (FINEEC) arranged an assessment of learning outcomes among first-grade pupils in basic education for the first time. The assessment was called the starting level measurement. The purpose was to produce national data on the pupils' starting level at the beginning of the first grade of basic education. The starting level measurement marked the beginning of a longitudinal assessment of learning outcomes in basic education. The goal of the longitudinal assessment is to assess the development of pupils' skills during comprehensive school in two key subjects, *mathematics* and *mother tongue and literature*, according to the goals set in the national core curriculum for basic education. The learning outcomes will be assessed in grades three, six and nine. This report describes the methodological solutions in the assessment.

The validity and reliability of data collection carried out to map pupils' competence at the beginning of primary school is sufficient to make conclusions at the national level. The information pupil data provides on pupils' competence at the beginning of primary school can easily be generalised to the entire population, especially when weighted results and key figures are discussed. More caution is needed, on the other hand, when information from data provided by guardians is used to explain competence differences. This is because the children of guardians whose responses are included in the data represent pupils who performed clearly better than the children of guardians who did not respond. The average of the points scored for overall competence by pupils whose guardians provided background information in the survey organised was 516. The average scored by pupils whose guardians did not respond to the survey was 480.

Key words: basic education, early primary education, evaluation, learning outcomes, literacy, mathematics, mother tongue and literature, assessment, methods

Sisällys

Tiivistelmä ... 3
Sammandrag ... 5
Summary ... 7

1 Johdannoksi ... 11

2 Mittariversiot, tehtävien osa-alueet ja niiden yhteys muihin mittauksiin ... 15
2.1 Matematiikan ja äidinkielen viitekehys ... 16
2.2 Luku- ja kirjoitustaidottomien oppilaiden osaamisen mittaaminen ja sähköinen tiedonkeruu ... 18
2.3 Osaamismittarit ja niiden luotettavuus ... 20
2.3.1 Mittauksen validiteettiin liittyviä kysymyksiä ... 20
2.3.2 Mittareiden metrisiä ominaisuuksia ... 21
2.4 Esimerkkejä eritasoisista tehtävistä ... 28
2.5 Asennemittarit ... 31
2.6 Huoltajakysely ja muu tiedonkeruu ... 32

3 Osioanalyysiin ja vertaistamiseen liittyviä erityiskysymyksiä ... 37
3.1 Osioanalyysin vaikutus lopullisiin mittareihin ... 38
3.2 Vertaistaminen ja muuttujamuunnokset ... 39

4 Termit ja menetelmät ... 43
4.1 Käytettävät termit ... 44
4.2 Analyysimenetelmät ... 45
4.3 Painotus ja ryvästyminen ... 47

5 Otos ja aineisto ... 51
5.1 Aineistojen koonti ja yhdistäminen ... 52
5.2 Otos ja kato ... 53
5.3 Lopullinen aineisto ja sen ominaispiirteitä ... 56

6 Yhteenveto ... 61

7 Viitteet ... 65

1

Johdannoksi

1

Vuoden 2018 syksyllä kerättiin ensimmäisen kerran osaamiseen liittyvää tietoa koulun aloittavalta ikäluokalta. Aiemmassa matematiikan pitkittäisarvioinnissa (Metsämuuronen, 2010a; 2010b; 2013a; 2013b; 2017a) kouluun tulon vaiheen osaaminen mallinnettiin 3. luokalla osoitetun osaamisen ja asiantuntijamielipiteen perusteella (ks. Metsämuuronen, 2010a). Tiedonkeruu oli ainutlaatuinen myös siksi, että matematiikkaan ja äidinkieleen liittyvää osaamista mitattiin samalla tehtäväsarjalla limittäin ja siksi että arviointitehtävät toteutettiin ääniohjeistetusti ekaluokkalaisten ehdoilla suunnitellulla digitaalisella järjestelmällä.

Aineiston keruun ajankohta oppilaan koulu-uran alkuvaiheessa on monella tavalla haasteellinen, eikä pelkästään sen vuoksi, että oppilaiden ei edellytetä osaavan lukea eikä laskea kouluun tullessaan. Toisaalta oppilaiden osaamisen tiedetään olevan varsin vaihtelevaa kouluun tullessa. Aiemman arvioinnin perusteella oli odotettavissa, että kouluun tullessa oppilaista löytyisi ainakin neljä erilaista populaatiota (Metsämuuronen & Salonen, 2010, s. 55). Osaavimpien oppilaiden ryhmästä voidaan käyttää nimitystä *erittäin edistyneet*. Tämän ryhmän osaamisen taso saattaa vastata jopa 3. luokan tasoa. Erittäin edistyneet oppilaat osaavat lukea ja laskea sujuvasti. Suurin ryhmä ovat *normaalisti edistyneet*, joiden osaamisen taso vastaa ikätasoa. He tunnistavat matematiikan peruskäsitteitä (kuten ympyrä, kolmio ja neliö) ja numeroita ja tunnistavat kirjaimia ja pystyvät ehkä lukemaan yksinkertaisia sanoja. Näitä heikommin menestyvät koulutulokkaat jaetaan vielä kahteen ryhmään: *aloittelevat oppilaat*, jotka ovat akateemisilta valmiuksiltaan erittäin puutteellisia, ja *kehittyvät oppilaat*, joilla on vain vähäinen käsitys kirjaimista ja numeroista. Miten siis rakentaa testistö, jolla voidaan mitata kaikkien näiden ryhmien osaamista ilman, että oppilaan edellytetään pystyvän lukemaan, kirjoittamaan tai suorittamaan matemaattisia operaatioita?

Tässä artikkelissa kuvataan aineistoon ja aineiston analyysiin liittyviä menetelmällisiä ratkaisuja. Ensin esitellään sähköinen järjestelmä ja mittaristot (luku 2), vertaistamiseen ja muuttujamuunnoksiin liittyviä näkökulmia (luku 3), käytetyt menetelmät ja termit (luku 4). Luvussa 5 esitellään aineisto koonti, katoanalyysi ja lopullisen aineiston keskeisiä piirteitä.

2

Mittariversiot, tehtävien osa-alueet ja niiden yhteys muihin mittauksiin

2

2.1 Matematiikan ja äidinkielen viitekehys

Ennen osaamismittareiden (tuonnempana tehtäväsarjojen) laatimista tutkittiin tarkasti sekä Esiopetuksen opetussuunnitelman perusteet (OPH, 2016) että Perusopetuksen opetussuunnitelman perusteissa (OPH, 2014) ilmaistut alkuopetusta koskevat tavoitteet. Näiden perusteella laadittiin ns. viitekehys, joka toimi osaamista koskevien mittareiden perustana. Viitekehyksessä matematiikan osaaminen jaetaan kolmeen keskeiseen sisältöalueeseen: *Geometria ja mittaaminen*, *Luvut ja laskutoimitukset* ja *Matemaattinen ajattelu*. Myös äidinkieli ja kirjallisuuden osaaminen jaetaan kolmeen keskeiseen sisältöalueeseen: *Kielen, kirjallisuuden ja kulttuurin ymmärtäminen*, *Tekstien tulkitseminen* ja *Tekstin tuottaminen*. Nämä jaetaan ala-alueisiin Taulukossa 1 kuvatulla tavalla. Opetussuunnitelman perusteissa on myös neljännet sisältöalueet: matematiikassa *Tietojenkäsittely ja tilastot* ja äidinkielessä *Vuorovaikutustilanteissa toimiminen*. Nämä alueet päätettiin jättää arvioinnin ulkopuolelle, joskin yksittäisillä mittaamisen osa-alueen osioilla saatiin selville lähtötasoa myös tietojenkäsittelyn ja tilastojen osa-alueelta (ks alaviite 1 luvussa 2.3.1). Lisäksi opetussuunnitelmissa kuvataan seitsemän laaja-alaisen osaamisen aluetta, joista arviointiin otettiin mukaan viisi: *Ajattelu ja oppimaan oppiminen*, *Kulttuurinen osaaminen, vuorovaikutus ja ilmaisu*, *Itsestä huolehtiminen ja arjen taidot*, *Monilukutaito* sekä *Tieto- ja viestintäteknologinen osaaminen*.

Näille sisältöalueille ei ole asetettu omia tavoitteita tai kriteereitä, mutta kaikki eri sisältöalueiden tavoitteet liittyvät vähintään yhteen laaja-alaisen osaamisen alueeseen. Siksi voidaan myös tulkita, että kaikki tehtävät mittaavat laaja-alaista osaamista omalla tavallaan. Näiden lisäksi myös erittäin helpot tehtävät erotettiin tehtäväsarjassa omaksi kokonaisuudekseen osaksi laaja-alaista osaamista.

Koska mittaus oli sähköinen (ks. luku 2.2), oleellista oli myös saada tietoa lapsen tietoteknisistä valmiuksista ja yleisestä ohjeiden ymmärtämisen tasosta. Niinpä tehtäväsarjan alkuun suunniteltiin lyhyt harjoitustehtävien sarja, jolla sekä koulutettiin oppilaat käyttämään varsinaista tiedonkeruujärjestelmää että myös kerättiin tietoa tehtäväsarjan suorittamiseen liittyvistä valmiuksista.

TAULUKKO 1. Arvioinnissa käytetyt OPS:n sisältöalueet

Oppiaine/laaja-alaisuus	Sisältöalue	Ala-alue	Tarkin kuvaustaso[1]
Matematiikka	1 Geometria ja mittaaminen	1 Geometria	
		2 Mittaaminen	
		3 Aika ja kello	
	2 Luvut ja laskutoimitukset	1 Laskut	
		2 Luvut	
	3 Matemaattinen ajattelu	1 Sanalliset kuuntelutehtävät	
		2 Päättely	
		3 Suhteet	
Äidinkieli	1 Kieli, kirjallisuus ja kulttuuri	1 Sanasto	
		2 Kirjallisuus	
	2 Tekstien tulkinta	1 Lukeminen	1 Kirjaintaso
			2 Tavutaso
			3 Kirjaintaso
			4 Sanataso
			5 Virketaso
			6 Tarkka lukeminen
		2 Kuvien lukeminen	
		3 Puhutun ymmärtäminen	1 Äännetaso
			2 Tavutaso
			3 Sanataso
			4 Virketaso
	3 Tekstin tuottaminen	1 Kirjallinen tuottaminen[2]	1 Sanataso
			2 Virketaso
			3 Käsin kirjoitus
Laaja-alainen osaaminen	0 Helpot tehtävät		
	1 Ajattelu ja oppimaan oppiminen		
	2 Kulttuurinen osaaminen, vuorovaikutus ja ilmaisu		
	3 Itsestä huolehtiminen ja arjen taidot		
	4 Monilukutaito		
	5 Tieto- ja viestintäteknologinen osaaminen		
Tietotekniikka/Yleiset valmiudet			Harjoitustehtävät

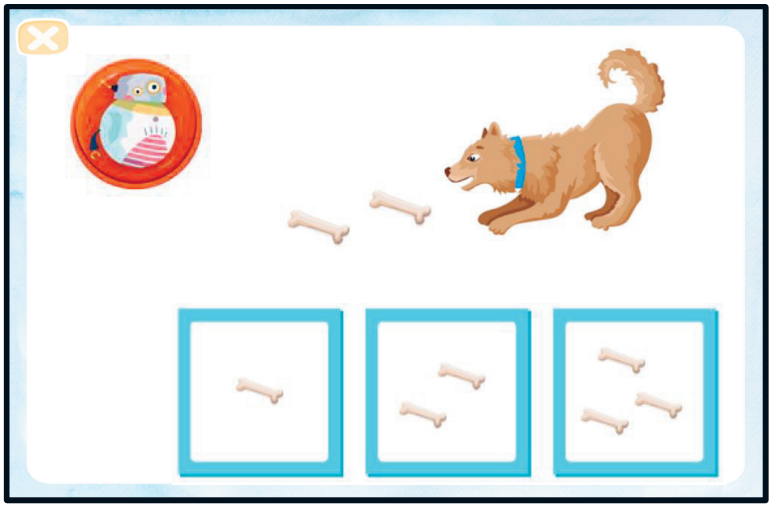
1) Näin tarkkaa sisältöaluejakoa ei valtakunnallisissa suunnitelmissa ole. Se lisättiin aineiston kuvaamisen rikastamiseksi.
2) Tätä mitattiin pienemmällä otoksella. Kirjallisen tuottamisen lisäksi tähän sisältöalueeseen olisi voitu sijoittaa Kuvien tuottaminen ja Puheen tuottaminen. Lisäksi äidinkielen sisältöalueista mittauksen ulkopuolelle jätettiin Vuorovaikutustilanteissa toimiminen.

2.2 Luku- ja kirjoitustaidottomien oppilaiden osaamisen mittaaminen ja sähköinen tiedonkeruu

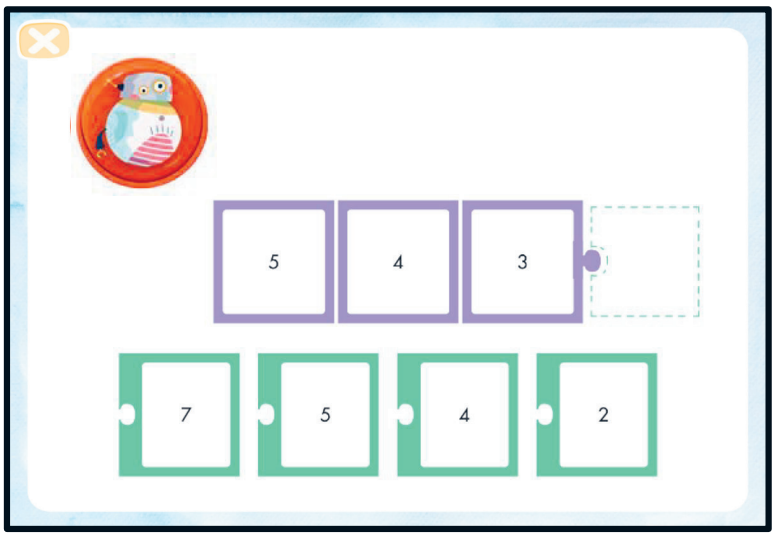
Varhaisten ikävuosien aikana tapahtuva osaamisen ja asenteiden mittaaminen on luonnollisesti haasteellisempaa kuin myöhempinä kouluvuosina, jolloin luku- ja kirjoitustaidot ovat kehittyneet. Yhtäältä haasteena on oppilaiden luku- ja kirjoitustaidon laaja vaihtelu – osa ei tunne vielä kirjaimia, kun osa lukee jo sujuvasti. Tämän suhteen tiedonkeruu olisi voitu järjestää esimerkiksi niin, että opettaja lukee oppilailleen tehtävät ääneen ja oppilaat olisivat edenneet tehtäväsarjassa opettajan määrittelemään tahtiin. Tällöin ei tietenkään voida yhdenmukaistaa opettajien käyttäytymistä, äänen painoja tai lukemisen nopeutta. Osa oppilaista saattaisi myös kokea testin valtaosin liian helppona, mikäli joutuu odottelemaan pidemmän aikaa vähäisemmillä taidoilla testiin osallistuvien oppilaiden vastaamista, ja toiset taas eivät osaisi pitää kynää kunnolla kädessään. Toiseksi haasteellista on se, että kansallisessa oppimistulosmittauksessa on tarkoituksen mukaista pyrkiä kattamaan testitehtävillä osaamisen eri tasoja niin laajasti kuin se on mahdollista. Tehtävien ja lopullisen tehtäväsarjan laadinnassa oleellista oli huolehtia siitä, että tehtävillä voidaan kartoittaa koulun alkamisen lähtötaso mahdollisimman laajasti niin, että kaikkien aloittavien oppilaiden osaamisen tasoa voitiin uskottavasti mitata. Tehtäväsarjaan suunniteltiin tietoisesti tehtäviä, joissa ei yhtäältä tarvittu luku-, lasku- tai kirjoitustaitoa, ja toisaalta tehtäviä, joilla voitiin mitata sitä, kuinka pitkälle parhaat oppilaat olivat edistyneet.

Tiedonkeruu päätettiin tehdä sähköisenä, mikä mahdollisti sen, että oppilaat voivat edetä tehtävissä kukin omaan tahtiinsa ja tehtävissä tarvittavat ohjeistukset tulevat kaikille oppilaille samanlaisina. Useita olemassa olevia testijärjestelmiä vertailtiin, mutta haasteeksi tuli aineistojen oikeuksien siirtymisessä Karvin hallintaan. Niinpä Karvi kehitti yhdessä pelillisiin testijärjestelmiin erikoistuneen NordicEdu-yhtiön kanssa pienemmillekin oppijoille soveltuvan testijärjestelmän, jonka ensimmäinen versio oli käytössä 6. luokan ruotsin B-oppimäärän (finska) oppimistulosten arvioinnissa 2018 keväällä (Åkerlund, Marjanen, & Lepola, 2019). Karvin puolesta kehitystyön vastuuhenkilönä oli arvioinnin digitalisoinnin projektipäällikkö Laura Pihala. Finskan osaamisen arviointiin kehitetty tiedonkeruujärjestelmä ei sellaisenaan soveltunut lukutaidottomien ekaluokkalaisten käyttöön. Niinpä sitä kehitettiin edelleen. Useita tehtäväpohjia ja järjestelmän käyttötapoja testattiin 3–10-vuotiailla lapsilla. Näissä havainnoinneissa todettiin, että kaksi tehtäväpohjaa, *paina* ja *siirrä* (Kuvat 1 ja 2), toimivat parhaiten ja niillä voitiin kattaa mitattavat sisältöalueet riittävällä laajuudella. Niinpä kaikki alkumittaukseen valitut sähköiset tehtävät olivat näitä kahta tyyppiä. Näiden lisäksi pienemmältä oppilasjoukolta koottiin myös käsin kirjoitettu kirjallinen tuotos.

Tehtäväsarja itsessään sijaitsi NordicEdun serverillä. Otoskoulut pääsivät heille lähetettyjen linkkien kautta järjestelmän kirjautumissivulle. Oppilaat kirjautuivat järjestelmään yksilöllisillä kuvatunnuksilla, joiden käyttö onnistui pääosin erittäin hyvin myös lukutaidottomilta oppilailta. Kouluja kehotettiin käyttämään testin suorittamiseen mieluiten tablettitietokoneita, jotka olivat esitestauksessa osoittautuneet pienten lasten kanssa helpommiksi käyttää kuin perinteinen tai kannettava tietokone hiirineen tai kosketuslevyineen. Tehtävien tekeminen oli kuitenkin mahdollista myös tietokoneella. Järjestelmä optimoitiin käytettäväksi kolmella selaimella, jotka olivat Safari, Chrome ja Firefox. Järjestelmää ei voinut käyttää Internet Explorerilla. Lopulliseen tiedonkeruuseen osallistuneista oppilaista 56 % käytti Safaria, 39 % Chromea ja 5 % Firefoxia. Käyttöjärjestelmä, selain tai ruutukoko ei ollut merkitsevässä yhteydessä osaamisen tasoon kokonaisaineistossa.



KUVA 1. Esimerkki *Paina*-tyyppisestä tehtävästä ("*Paina sitä kuvaa, jossa on yksi luu vähemmän kuin koiralla.*")



KUVA 2. Esimerkki *Siirrä*-tyyppisestä tehtävästä ("*Siirrä alhaalta oikea luku jonoon.*")

2.3 Osaamismittarit ja niiden luotettavuus

Tehtäväsarjaan tehtiin tietoisesti tehtäviä, joissa yhtäältä ei tarvittu luku-, lasku-, tai kirjoitustaitoa, ja joilla toisaalta voitiin mitata sitä, kuinka pitkälle parhaat oppilaat olivat edistyneet. Tehtävien tuli kuvata osaamista sekä laaja-alaisesti että riittävällä tarkkuudella. Mittareiden laadintaa kuvaavassa perusohjeistuksessa (Metsämuuronen, 2009a) erityisesti mainitaan, että tavoitteena on maksimoida reliabiliteetin eli tarkkuuden ja johdonmukaisuuden sijaan validiteetti eli se mitataanko oikeaa asiaa. Tosin ymmärrämme, että jos mittari ei ole reliaabeli eli jos se tuottaa epäjohdonmukaisia tuloksia, se ei voi olla validikaan (Metsämuuronen & Räsänen, 2018, s. 337).

Taulukkoon 2 on koottu mittareita koskevia keskeisiä tunnuslukuja. Näistä kahta vähemmän tunnettua, *alfaU*:ta ja *JTR*-keskiarvoa (*MJTR*) käsitellään perusteellisemmin toisaalla (Metsämuuronen, 2018a; 2018b; 2019a), mutta oleellisia laskennallisia seikkoja ja erityisesti niiden tulkintaa käsitellään myös tässä osuudessa. Taulukon 2 perusteella pohditaan ensin mittauksen validiteettiin liittyviä seikkoja ja sen jälkeen reliabiliteettiin ja metrisiin ominaisuuksiin liittyviä seikkoja.

2.3.1 Mittauksen validiteettiin liittyviä kysymyksiä

Ensimmäinen huomio Taulukosta 2 liittyy mittauksen validiteettiin. Mittauksen validiteettia voidaan tarkastella useasta näkökulmasta. Ensiksi, kansallisen arvioinnin näkökulmasta *rakennevaliditeettia* voidaan arvioida sen perusteella, kuinka hyvin mittarin rakenne vastaa kansallisissa opetussuunnitelmien perusteissa kuvattua tai taustalla olevaa, tarkoitettua (teoreettista) rakennetta (ks. Metsämuuronen & Räsänen, 2018, s. 338). Vertaamalla Taulukkoja 1 ja 2 huomataan, että mittariston rakenne vastaa pitkälti teoreettista rakennetta. Taulukossa 1 nähtävistä 30 pää- ja alasisältöalueista kaikki ovat mittauksen kohteena.[1] Tekstin tuottamisen alasisältöinä olevia kuvallista ja puheen tuottamista ei tehtäväsarjoilla mitattu, mutta muutoin kaikkia osa-alueita tarkastellaan tehtäväsarjassa. Kattavuutta voi pitää erittäin osuvana.

Toiseksi, kansallisen arvioinnin näkökulmasta *sisällön validiteettia* voidaan arvioida sen perusteella, kuinka hyvin mittarin sisältö vastaa opetussuunnitelman perusteissa kuvattuja sisältöalueen (esimerkiksi geometrian ja mittaamisen tai tekstien tulkinnan) tavoitteita. Tätä pyrittiin saavuttamaan yhtäältä sillä, että tehtävien laatijoina oli kokeneita varhaiskasvatuksen ja alkuopetuksen opettajia ja asiantuntijoita, ja toisaalta sillä, että tehtävien määrä oli riittävä laaja-alaisuuden saavuttamiseksi. Taulukosta 2 havaitaan, että osassa sisältöalueista – erityisesti tarkemmilla ala-alueilla – osioiden määrä jäi pieneksi, joskin kaikilla sisältöalueilla osioita on ainakin neljä.

1 Näiden lisäksi sisältöalueista kaksi (Tietojenkäsittely ja tilastot ja Vuorovaikutustilanteissa toimiminen) jätettiin pois arvioinnista. Tilastojen lukutaitoa mitataan alkeellisesti esimerkiksi lämpömittaritehtävässä. Näin olleen tämänkin sisältöalueen alkeellinen lähtötasotieto on olemassa, mutta varsinaisesti sen osaamista arvioidaan vasta pitkittäisarvioinnin seuraavassa vaiheessa.

Kolmanneksi, kansallisen arvioinnin näkökulmasta *ekologista validiteettia* voidaan arvioida sen perusteella, katetaanko mittauksessa riittävällä laajuudella eri ajattelun tasoja esimerkiksi Bloomin (Bloom ym., 1956) tai siitä johdetun Andersonin–Krathwohlin (Anderson & Krathwohl, 2000) luokituksen näkökulmasta. Jos mittarissa olisi yksinomaan esimerkiksi muistamista tai tunnistamista edellyttäviä tehtäviä, eikä mittauksessa voisi arvioida lainkaan asian ymmärtämistä, soveltamista tai korkeampaa ajattelua (analysoiminen, arvioiminen, luominen), mittausta ei voisi pitää ekologisen validiteetin osalta hyvänä. Koska mittauksessa oli tärkeää saada tietää laajasti oppilaiden lähtötasoa, tämä luontevasti ohjasi tehtävän laatimisessa ja mittarin laatimisessa valitsemaan tehtäviä, jotka mittaavat niin tunnistamista, ymmärtämistä, soveltamista kuin jossain määrin myös korkeampia taitoa lapsen ikätaso huomioiden. Näin siis esimerkiksi korkeampiin taitoihin luettavat ”arvioiminen” ja ”luominen” ovat 6–7-vuotiaan oppilaan näkökulmasta erilaisia kuin ylemmillä luokilla. Voidaan arvioida, että mittauksen ekologinen validiteetti on riittävä – ellei hyvä – kuvaamaan erilaisia ajattelun tasoja.

Lopuksi, mittauksen *ilmivaliditeettia* tai *koettua validiteettia* (face validity), eli asiantuntijan henkilökohtaista tuntemusta siitä, kuinka osuva mittaus on, voi kansallisen arvioinnin yhteydessä arvioida esimerkiksi sillä, kuinka hyvin se vastaa kohdejoukon osaamisen tasoa. Liian vaikea tai liian helppo tehtäväsarja ei ”tunnu” hyvältä. Tulosten perusteella näyttää siltä, että koko tehtäväsarjan ja eri sisältöalueiden mittareiden vaikeustaso on kokonaisuutena ollut varsin osuva.

2.3.2 Mittareiden metrisiä ominaisuuksia

Kaikkiaan koko tehtäväsarjassa oli 115 osiota, joista saattoi saada yhteensä 138 pistettä, joskaan kukaan oppilaista ei saanut täysiä pisteitä. Kaikki tehtävät pisteitettiin automaattisesti koneellisesti eli opettajan pisteittämiä tuottamistehtäviä ei varsinaisessa tehtäväsarjassa ollut. Äidinkielen kirjallista tuottamista kartoitettiin lisäksi erillisellä kirjoitustehtävällä 508 oppilaan näytteen avulla.

Taulukosta 2 huomataan, että äidinkielen kokonaistehtäväsarjassa on 104 osiota ja matematiikan vastaavassa vain 58. Mittareiden pituuksien ero johtuu siitä, että lähes kaikissa matemaattisissa tehtävissä on myös kielellinen komponentti: tehtävänannot kuultiin puhuttuna (esimerkiksi: ”Paina *suurinta* lukua.”). Näin ollen valtaosa matematiikan tehtävistä tuotti tietoa myös kuullun ymmärtämisestä ja näin ollen äidinkielen tehtäväsarjan kokonaispistemäärästä muodostui noin kaksinkertainen matematiikan vastaavaan nähden.

TAULUKKO 2. Osaamisen sisältöalueita kuvaavien mittareiden ominaisuuksia

Mittarin sisältö	osioiden määrä (k)	maksimi piste-määrä	vaikeus-taso (p)	reliabi-liteetti alfaR	reliabi-liteetti alfaU	JTR keski-arvo
Koko tehtäväsarja (harjoitustehtävät, MAT, ja ÄI)	115	138	0,618	0,929	0,962	0,734
Harjoitustehtävät	8	11	0,961	0,246	0,859	0,915
Matematiikka koko tehtäväsarja	58	62	0,541	0,879	0,940	0,748
Äidinkieli koko tehtäväsarja	104	126	0,625	0,925	0,960	0,738
MAT1 Geometria ja mittaaminen	17	17	0,607	0,646	0,850	0,789
MAT1.1 Geometria	9	9	0,511	0,473	0,799	0,830
MAT1.2 Mittaaminen	10	10	0,728	0,618	0,836	0,827
MAT1.3 Aika ja kello	4	4	0,554	0,432	0,700	0,863
MAT2 Luvut ja laskutoimitukset	25	25	0,466	0,826	0,912	0,789
MAT2.1 Laskut	17	17	0,388	0,776	0,888	0,804
MAT2.2 Luvut	8	8	0,632	0,680	0,858	0,858
MAT3 Matemaattinen ajattelu	30	36	0,574	0,766	0,871	0,736
MAT3.1 Sanalliset kuuntelutehtävät	10	13	0,488	0,629	0,779	0,788
MAT3.2 Päättely	6	9	0,491	0,404	0,620	0,780
MAT3.3 Suhteet	18	18	0,648	0,649	0,851	0,783
Laaja-alainen osaaminen (=koko tehtäväsarja)	115	138	0,618	0,929	0,962	0,734
LA0 Helpot tehtävät	28	31	0,898	0,708	0,920	0,823
LA1 Ajattelu ja oppimaan oppiminen	64	68	0,577	0,876	0,940	0,742
LA2 Kulttuurinen osaaminen, vuorovaikutus ja ilmaisu	17	20	0,514	0,745	0,832	0,763
LA3 Itsestä huolehtiminen ja arjen taidot	26	27	0,670	0,732	0,873	0,764
LA4 Monilukutaito	70	85	0,624	0,905	0,948	0,753
LA5 Tieto- ja viestintäteknologinen osaaminen	25	31	0,637	0,595	0,845	0,771
ÄI1 Kieli, kirjallisuus ja kulttuuri	68	87	0,628	0,883	0,936	0,734
ÄI1.1 Sanasto	50	71	0,652	0,852	0,918	0,740
ÄI1.2 Kirjallisuus	16	16	0,517	0,698	0,837	0,783
ÄI2 Tekstien tulkinta	73	93	0,604	0,903	0,945	0,737
ÄI2.1 Lukeminen	38	55	0,612	0,892	0,934	0,785
ÄI2.1.1 Kirjaintaso	25	42	0,663	0,862	0,918	0,817
ÄI2.1.2 Tavutaso	16	33	0,595	0,856	0,903	0,807
ÄI2.1.3 Sanataso	14	31	0,577	0,854	0,899	0,807
ÄI2.1.4 Virketaso	12	12	0,439	0,763	0,838	0,800
ÄI2.1.5 Tarinataso	9	9	0,422	0,690	0,792	0,806
ÄI2.1.6 Tarkka lukeminen	4	4	0,419	0,458	0,661	0,851
ÄI2.2 Kuvien lukeminen	25	31	0,678	0,725	0,852	0,755
ÄI2.3 Puhutun ymmärtäminen	79	84	0,659	0,888	0,951	0,751
ÄI2.3.1 Äännetaso	12	12	0,877	0,752	0,935	0,905
ÄI2.3.2 Tavutaso	3	3	0,769	0,490	0,808	0,928
ÄI2.3.3 Sanataso	37	40	0,652	0,779	0,901	0,762
ÄI2.3.4 Virketaso	15	18	0,577	0,678	0,829	0,789
ÄI3 Tekstien tuottaminen						

Taulukosta 2 huomataan myös, että osioiden määrät eri mittareissa vaihtelevat runsaasti. Yleisesti tiedetään, että pidempi mittari on reliabiliteetiltaan korkeampi kuin lyhyempi mittari. Tämä ei tarkalleen ottaen pidä paikkaansa, mutta perinteiset tavat laskea reliabiliteetti (esimerkiksi perinteinen alfakerroin, Taulukossa 2 *alfaR*) riippuvat voimakkaasti summapistemäärän varianssista. Jos summamuuttujassa on vain vähäisesti arvoja – kuten tapahtuu, kun osioita on vähän – varianssi jää teknisistä syistä johtuen pieneksi. Jos lisäksi osioiden ja summan väliset korrelaatiokertoimet (osio–summa-korrelaatiot) jäävät mataliksi, myös reliabiliteetilla on taipumusta jäädä matalaksi, vaikka summamuuttuja kykenisikin erottelemaan vastaajat toisistaan riittävällä tarkkuudella. Tämän indikaattorina Taulukossa 2 on kaksi uudempaa indeksiä, *alfaU* (Metsämuuronen, 2018b) ja *JTR* keskiarvo (*MJTR*, Metsämuuronen, 2019a). Koska kaksi jälkimmäistä on lukijoille vieraampia, niiden logiikka ja laskenta esitellään lyhyesti.

AlfaU

Oletetaan, että testissä on k osiota, jotka kaavoissa indeksoidaan symbolilla g ja jotka muodostavat summamuuttujan X, ja summa muodostaa järjestysmuuttujan R. Yksi summamuuttujan reliabiliteetin arvioimiseksi käytettävistä indikaattoreista, perinteinen alfakerroin (Kuder & Richardson, 1937; Gulliksen, 1950; Cronbach, 1951), voidaan esittää muodossa

$$\alpha_R = \frac{k}{k-1}\left(1 - \frac{\sum_{g=1}^{k}\sigma_g^2}{\left(\sum_{g=1}^{k}\sigma_g\rho_{gX}\right)^2}\right) \tag{1}$$

(Lord & Novick, 1968, s. 344) jossa σ_g^2 viittaa osioiden (populaatio)varianssiin[2] ja ρ_{gX} osioiden ja summamuuttujan väliseen korrelaatiokertoimeen. ρ_{gX} on eräs keskeisistä osion erottelukykyä kuvaavista indikaattoreista, ja se aliarvioi aina erottelukykyä, mikäli osion ja summan vapausasteet[3] $df(g)$ ja $df(X)$ poikkeavat toisistaan (ks. Metsämuuronen, 2016). Siksi tiedetään, että ρ_{gX} on yksi syy tunnettuun ilmiöön, että alfakerroin aliarvioi mittarin reliabiliteettia (ks. Metsämuuronen, 2016; 2018a; 2018b). Metsämuuronen (2018b) esittää, että parempi arvio mittarin erottelukyvylle saadaan, kun ρ_{gX}:n tilalle vaihdetaan toinen erottelukyvyn indikaattori, U_{gX} (Metsämuuronen, 2018c; 2019c), joka on itse asiassa dimensionkorjattu järjestys-polyseriaalinen korrelaatiokerroin (dimension-corrected rank polyserial coefficient of correlation, Metsämuuronen, 2019b). Dimensio*korjaamaton* järjestys-polyseriaalinen korrelaatiokerroin JT_{gX} puolestaan perustuu Jonckheere–Terpstra testisuureeseen (*JT*, Jonckheere, 1954; Terpstra, 1952) joka edelleen perustuu Mannin ja Whitneyn *U*-testisuureeseen (*U*, Mann & Whitney, 1947). Niinpä kertoimien nimissä U_{gX} ja JT_{gX} nähdään suoraan niiden historiallinen tausta. JT_{gX} on yleensä korrelaatiokerrointa parempi arvioitaessa binäärisen (0/1-jakoisen) osion erottelukykyä, ja sen dimensiokorjattu versio U_{gX} antaa

2 Varianssi voidaan laskea kahdella tavalla: otoskaavan ja populaatiokaavan avulla. Näissä laskuissa käytetään aina populaatiokaavaa.

3 $df(g)$ = luokkien määrä miinus 1 osiossa ja $df(X)$ = luokkien määrä miinus 1 summassa.

korrelaatiokerrointa tarkemman arvion erottelukyvystä polytomisilla eli useita arvoja saavilla osioilla (Metsämuuronen, 2019b, 2019c). Dimensiokorjatussa alfakertoimessa, jota merkitään α_U (*alfaU*), aliarvion antava ρ_{gX} on korvattu paremmin erottelukykyä vastaavalla estimaattorilla U_{gX}

$$\alpha_U = \frac{k}{k-1}\left(1 - \frac{\sum_{g=1}^{k}\sigma_g^2}{\left(\sum_{g=1}^{k}\sigma_g U_{gX}\right)^2}\right) \tag{2}$$

(Metsämuuronen, 2018b, s. 12), missä

$$U_{gX} = 1 - \left(JT_{gX} - 1\right) \times \left(A - 1\right) \tag{3}$$

missä

$$A = \frac{df(g) - 1}{df(g)}\left(1 - \frac{1}{df(g)}\right)^2 \tag{4}$$

(Metsämuuronen, 2019c, ss. 11–12) ja missä JT_{gX} lasketaan havaitun JT-arvon (JT_{gX}^{Obs}) avulla:

$$JT_{gX} = 2 \times \frac{JT_{gX}^{Obs}}{JT_{gX}^{Max}} - 1 \tag{5}$$

(Metsämuuronen, 2019b, s. 13). JT_{gX} voidaan laskea havaittujen Mannin–Whitneyn U-arvojen (U_{gXij}^{Obs}) ja maksimaalisten U-arvojen ($U_{gXij}^{Max} = n_i n_j$) avulla seuraavasti:[4]

$$JT_{gX} = 2 \times \frac{\sum_{i<j}^{r} U_{gXij}^{Obs}}{\sum_{i<j}^{r} n_i n_j} - 1 \tag{6}$$

4 Jonckheere–Terpstra-testisuureen laskemista kuvaa tarkemmin esimerkiksi Metsämuuronen (2017c, ss. 216–225).

(Metsämuuronen, 2019b), missä n_i ja n_j ovat otoskokoja ryhmissä i ja j ja $i < j$. Binääriaineistossa n_i ja n_j ovat n_0 ja n_1 ja Likert-tyyppisessä mittauksessa yksi pari on esimerkiksi $i = 1$ ja $j = 4$. Kaava tarkoittaa siis sitä, että lasketaan, kuinka paljon on sellaisia havaintoja, jotka sijoittuvat osiossa summamuuttujan mukaiseen suuruusjärjestykseen. Voidaan osoittaa (Metsämuuronen, 2019b), että JT_{gX} on Somersin $D(g|X)$:n (Somers, 1962) erityistapaus tilanteesta, jossa osioon vastaamista selitetään summamuuttujalla. Tämä tarkoittaa yhtäältä sitä, että JT_{gX} on ns. suunnattu korrelaatiokerroin, ja toisaalta sitä, että U_{gX}:n kaavassa (5) voidaan suoraan käyttää Somersin deltan erityistapausta $D(g|X)$, jossa summamuuttuja X selittää käyttäytymistä osiossa g:[5]

$$U_{gX} = 1 - (Somersin\, D(g|X) - 1) \times (A - 1) \tag{7}$$

jossa A on kuten kaavassa 4. *AlfaU*:n arvot ovat yleensä hieman korkeampia kuin *alfaR*:n, mikä näkyykin Taulukosta 2. Ero näkyy selkeimmin testeissä, joiden vaikeustaso on äärimmäinen – siis joko erittäin helppo tai erittäin vaikea – joissa *alfaR*:n sisältämä korrelaatiokerroin antaa selvän aliarvion, mutta *alfaU*:n sisältämä U_{gX} on vakaa (ks. Metsämuuronen, 2018b). Taulukon 2 testeistä näitä ovat esimerkiksi äärimmäisen helpot harjoitustehtävät ($p = 0{,}96$, $\alpha_R = 0{,}25$, $\alpha_U = 0{,}86$), melko vaikea tarkan lukemisen tehtäväkokonaisuus ($p = 0{,}42$, $\alpha_R = 0{,}46$, $\alpha_U = 0{,}66$) ja matematiikan päättelytehtävät ($p = 0{,}49$, $\alpha_R = 0{,}43$, $\alpha_U = 0{,}62$). Toisaalta *alfaR* ei ole vahva lyhyissä testeissä, joissa summamuuttujan varianssi jää matalaksi. Esimerkkinä näistä Taulukossa 2 ovat tavutason lukeminen ($k = 3$, $\alpha_R = 0{,}49$, $\alpha_U = 0{,}81$) sekä aikaan ja kelloon liittyvät tehtävät ($k = 4$, $\alpha_R = 0{,}43$, $\alpha_U = 0{,}70$).

JTR ja MJTR

Sitä, antaako perinteinen alfakerroin matalan arvion teknisistä syistä, voidaan arvioida *MJTR*-kertoimen avulla (Metsämuuronen, 2019b). Edellä kaavojen (5) ja (6) ytimessä on havaitun *JT*-testisuureen (JT_{gX}^{Obs}) ja maksimaalisen *JT*-testisuureen (JT_{gX}^{Max}) välinen suhde (*Ratio*). Tätä suhdetta (*JT Ratio*) merkitään lyhenteellä *JTR*, joka kaavoista (5) ja (6) yhdistämällä on

$$JTR = \frac{JT_{gX}^{Obs}}{JT_{gX}^{Max}} = \frac{\sum_{i<j}^{r} U_{gXij}^{Obs}}{\sum_{i<j}^{r} n_i n_j} \tag{8}$$

(Metsämuuronen, 2019b, kaava 15). Binäärin muuttujan tapauksessa (8) yksinkertaistuu muotoon

5 Somersin *D*:n laskemista kuvaa tarkemmin esimerkiksi Metsämuuronen (2017c, ss. 266–270).

$$JTR = \frac{U_1}{n_0 n_1} = \frac{W_1 - n_1\left(n_1+1\right)/2}{n_0 n_1} \tag{9}$$

jossa U_1 on osion ja summamuuttujan välinen U-testisuureen arvo, joka liittyy ryhmään $j = 1$ ja W_1 on arvon 1 saaneiden vastaajien järjestyspisteiden summa. Taulukossa 3 havainnollistetaan JTR:n laskemista (hypoteettisen) osion g_1 sekä summamuuttujan X ja tästä muodostetun järjestyspistemuuttujan R avulla. Sijoittamalla Taulukossa 3 olevat arvot kaavaan (8) saadaan $JTR = JT_{g_2X}^{Obs} \big/ JT_{g_2X}^{\max} = 54/55 = 0{,}982$. Tämä tarkoittaa sitä, että vain yhden vastaajan yhdessä vastausvaihtoehdossa kaikkiaan 55:stä mahdollisesta järjestyksestä (1,8 %) vastaajien järjestys osiossa *ei* ollut looginen, ja näin 98,2 %:lla havainnoista se oli loogisesti järjestynyt.

TAULUKKO 3. *JTR*:n laskeminen käsin

vastaaja	g_1	X	R	U_{ij}	g_1	n_i x n_j	g_1
1	1	24	1,5	U_{12}	2	$n_1\ n_2$	2
2	1	24	1,5	U_{13}	4	$n_1\ n_3$	4
3	2	25	3	U_{14}	8	$n_1\ n_4$	8
4	3	26	4	U_{15}	6	$n_1\ n_5$	6
5	4	27	5	U_{23}	2	$n_2\ n_3$	2
6	3	28	6	U_{24}	4	$n_2\ n_4$	4
7	4	30	7	U_{25}	3	$n_2\ n_5$	3
8	4	31	8	U_{34}	7	$n_3\ n_4$	8
9	4	33	9	U_{35}	6	$n_3\ n_5$	6
10	5	36	10,5	U_{45}	12	$n_4\ n_5$	12
11	5	36	10,5	SUM	54	SUM	55
12	5	40	12				
JT_{gX}^{obs}	54	JTR	0,982				
$JT_{gX}^{\max}$	55	ρ_{gX}	0,881				

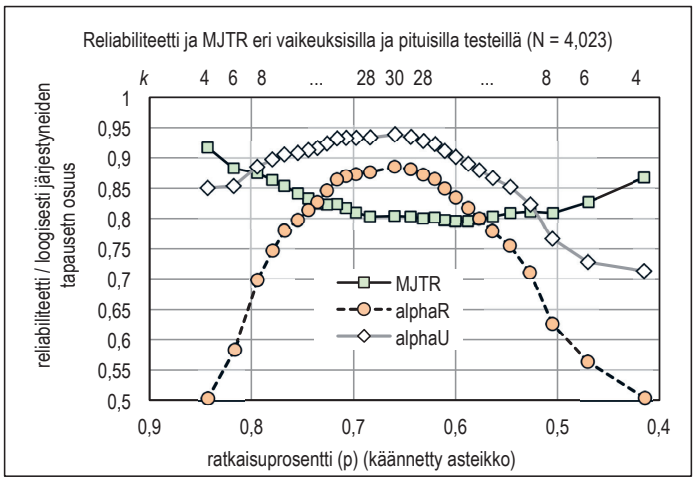
JTR kertoo loogisesti järjestyneiden havaintojen suhteellisen osuuden maksimaalisen loogisesti järjestyneestä tapauksesta. Jos $JTR = 1$, kaikki havainnot osiossa ovat ryhmittyneet niin, että pienemmän summapistemäärän saanut vastaaja sijoittuu pienemmäksi myös osiopistemäärillä tarkasteltuna. Jos taas $JTR = 0{,}5$, havainnot osiossa ovat satunnaisessa järjestyksessä ilman loogista yhteyttä summamuuttujan tuottamaan järjestykseen. Kun $JTR = 0$, kaikki parhaiten suoriutuneet vastaajat ovat osiossa ryhmittyneet (patologisesti) *heikoimmiksi*.

Kaikkien osioiden keskimääräinen *JTR* (*Mean* $JTR(g|X)$ eli *MJTR*) kuvaa loogisesti järjestyneiden havaintojen osuutta summapistemäärän muodostaneissa osioissa yhteensä:

$$MJTR = \frac{1}{k}\sum_{g=1}^{k} JTR(g|X) \tag{11}$$

(Metsämuuronen, 2019a, kaava 23). Jos kaikissa osioissa kaikki vastaajat ovat loogisesti samassa järjestyksessä kuin summamuuttujassa, *MJTR* = 1. Vastaavasti jos kaikissa osioissa kaikki vastaajat ovat täysin satunnaisessa järjestyksessä, *MJTR* = 0,5. *MJTR* näyttää kertoimena olevan huomattavasti vakaampi kuin reliabiliteetti (ks. Metsämuuronen, 2019a). Kuvassa 3 havainnollistetaan tätä Metsämuurosen (2019a) aineiston pohjalta. Mitä vähemmän summamuuttujassa on luokkia, sitä suuremmaksi *MJTR*:llä on taipumusta tulla, ja vastaavasti mitä enemmän luokkia on, sitä pienemmäksi *MJTR* menee. Erot eivät kuitenkaan ole lainkaan samaa luokkaa kuin reliabiliteetin yhteydessä. *MJTR*:ää voi siis käyttää karkeana indikaattorina kuvaamaan sitä, onko matala alfakertoimen arvo seurausta mekaanisista tekijöistä vai onko mittarin erottelukyky kuitenkin korkea matalasta alfasta huolimatta.

Selkeästi alfakertoimesta poikkeavan viesti saadaan *MJTR*:n avulla esimerkiksi erittäin helpoista harjoitustehtävistä, joissa α_R = 0,25 ja *MJTR* = 0,92 eli peräti 92 % kaikista havainnoista oli loogisessa järjestyksessä, vaikka sisäinen yhdenmukaisuus väittää muuta. Näin käy myös melko helpossa puhutun ymmärtämisen tavutason tehtäväkokonaisuudessa, joka lisäksi oli erittäin lyhyt (k = 3, p = 0,77, α_R = 0,49, *MJTR* = 0,93), ja erittäin vaikeissa tarkan lukemisen tehtävissä (k = 4, p = 0,42, α_R = 0,46, *MJTR* = 0,85). Erityisesti näiden summien suhteen perinteinen reliabiliteetti antaa harhaisen kuvan mittarin erottelukyvystä.



KUVA 3. *MJTR*:n ja reliabiliteetin yhteys eri pituisilla ja eri vaikeusasteisilla testeillä (Metsämuuronen, 2019b)

Kokonaisarvio mittareiden erottelukyvystä

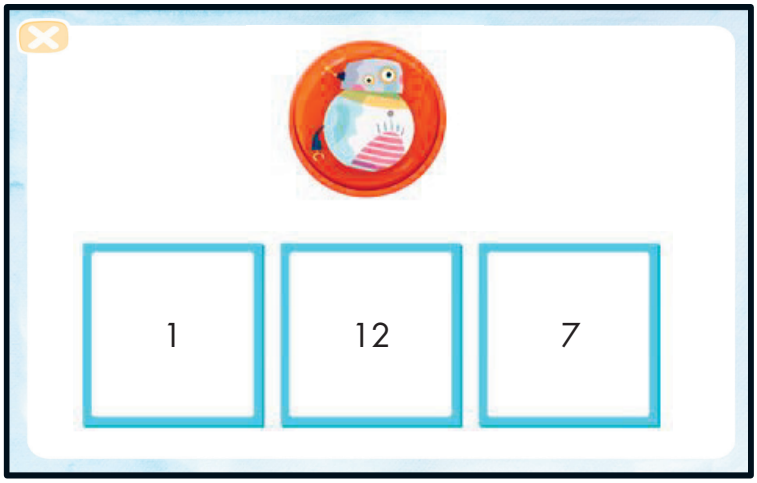
Odotetusti reliabiliteetit ovat korkeita niillä mittareilla, jotka koostuvat monista osioista ja joissa oppilasvaihtelu on suurta. Näitä ovat keskeiset pääsummat: kokonaisosaaminen (α_R = 0,93, α_U = 0,96) ja matematiikan (α_R = 0,88, α_U = 0,94) sekä äidinkielen kokonaissummat (α_R = 0,93, α_U = 0,96). Myös ala-alueiden summien reliabiliteetit olivat tyydyttäviä vähintään toisella reliabiliteetin indikaattorilla mitattuna: Geometria ja mittaaminen (α_R = 0,65, α_U = 0,85), Luvut ja laskutoimitukset (α_R = 0,83, α_U = 0,91), Matemaattinen ajattelu (α_R = 0,77, α_U = 0,87), Kieli, kirjallisuus ja kulttuuri (α_R = 0,88, α_U = 0,94), Tekstien tulkinta (α_R = 0,90, α_U = 0,95), Kuvien lukeminen (α_R = 0,73, α_U = 0,85), Puhutun ymmärtäminen (α_R = 0,89, α_U = 0,95). Tätäkin tarkemmilla tasoilla reliabiliteetit ovat riittäviä tai jopa korkeita lukuun ottamatta ehkä Tarkan lukemisen kokonaisuutta (α_R = 0,46, α_U = 0,66), joskin tässäkin *MJTR* osoittaa havaintojen olevan varsin loogisessa järjestyksessä (*MJTR* = 0,85).

Kaikkiaan mittarit ovat riittävän tarkkoja erottelemaan oppilaita toisistaan.

2.4 Esimerkkejä eritasoisista tehtävistä

Sekä matematiikan että äidinkielen tehtäväsarjoihin suunniteltiin sellaisia tehtäviä, jotka olisivat todella helppoja lähes kaikille oppilaille. Matematiikan tehtävistä helpoimmat mittasivat esimerkiksi matemaattisten käsitteiden hallintaa (enemmän, vähemmän, suurin, pienin), lukukäsitteen ymmärtämistä (lukusana, lukumäärä, numeromerkki) ja tasokuvioiden tunnistamista (Kuva 4). Keskivaikeissa tehtävissä oli esimerkiksi sarjoittamista, järjestykseen asettamista ja laskutoimituksia lukualueella 1–9 (Kuva 5). Vaikeimpia tehtäviä olivat esimerkiksi murtolukujen alkeisiin liittyvät tehtävät ja pakkaseen eli negatiivisiin lukuihin liittyvät tehtävät, eikä tällaisten tehtävien osaamista tietenkään odotettukaan monilta oppilailta.

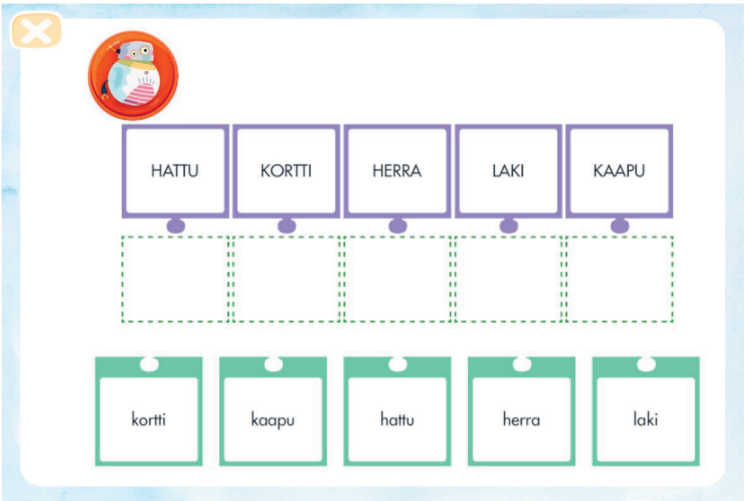
Äidinkielen tehtävistä helpoimmat olivat kirjain-äänne-vastaavuuden tunnistamista ja paikkaa merkitsevien kuultujen sanojen yhdistämistä oikeaan kuvaan (Kuva 6). Keskitasoiset tehtävät mittasivat esimerkiksi lyhyiden yksittäisten sanojen lukemista tai kirjallisuuteen liittyvien käsitteiden tunnistamista kuvista (Kuva 7). Vaikeimmat tehtävät vaativat tarkkaa lukemista tai usean virkkeen pituisten tekstien ymmärtämistä joko kuultuna tai kirjoitettuna.



KUVA 4. Esimerkki matematiikan helposta tunnistamistehtävästä ("*Paina lukua 12.*")

KUVA 5. Esimerkki matematiikan keskivaikeasta laskemistehtävästä ("*Korissa on yhteensä kahdeksan karkkia. Jänis saa niistä kaksi karkkia. Montako jää jäljelle? Paina oikeaa lukua.*")

Kuva 6. Esimerkki äidinkielen helposta alkukirjaintehtävästä ("*Paina sanan ensimmäistä kirjainta. Elefantti.*")



Kuva 7. Esimerkki äidinkielen keskivaikeasta lukemistehtävästä ("*Yhdistä samat sanat pareiksi. Siirrä alhaalla olevat palat ylös.* ")

2.5 Asennemittarit

Kansallisten oppimistulosarviointien yhteydessä Karvissa on järjestelmällisesti koottu tietoa myös oppilaiden asennoitumisesta oppiaineeseen. Vakiotestinä ylemmillä luokilla on vuodesta 1998 lähtien käytetty Fennema–Sherman-asennetestin (Fennema & Sherman, 1978) pohjalta rakennettua testistöä, jossa kartoitetaan asennetta oppiaineeseen kolmella dimensiolla: käsitys itsestä oppiaineen osaajana (OSAA), oppiaineesta pitäminen (PITÄÄ), ja käsitys oppiaineen hyödyllisyydestä (HYÖTY) (ks. Metsämuuronen, 2009a; 2012). Karvin ja aiemmin Opetushallituksen käyttämässä perusmittarissa kutakin dimensiota mitataan viidellä osiolla viisiportaisella Likert-asteikolla. Alkuperäistä Fenneman ja Shermanin mittaria on lyhennetty, ja sen pituus vastaa kansainvälisissä PISA- (Programme for International Student Assessment, OECD, 2014; 2016) ja TIMSS-mittauksissa (Trends in International Mathematics and Science Study, TIMSS, 2012; 2016) käytettyä versiota, joissa kuitenkin käytetään neliportaista asteikkoa. Suomalaisessa versiossa väitteitä on kansainväliseen mittariin nähden selvästi yksinkertaistettu (ks. perusteluja Metsämuuronen, 2012).

Varhaisten kouluvuosien mittauksissa on käytetty perusmittarista lyhennettyä ja yksinkertaistettua mittaria (ks. esimerkiksi Metsämuuronen 2010a). HYÖTY-komponentti on jätetty kokonaan pois ja asteikko on muutettu 5-portaisesta 4-portaiseksi. Lisäksi osioiden sanamuotoja on muokattu konkreettisemmiksi (esimerkiksi ”Matematiikka on helppo oppiaine” on muutettu muotoon ”Matematiikka on helppoa”). Käsillä olevassa ensimmäisen luokan alun mittauksessa on samoja elementtejä kuin aiemmissa varhaisten luokkien mittauksessa siltä osin, että kolmesta peruskomponentista saadaan tietoa suhtautumisesta oppiaineeseen (PITÄÄ) ja käsityksistä omasta osaamisesta (OSAA), mutta osiot ovat erilaisia kuin aiemmassa mittauksessa. Kolmantena asennekokonaisuutena kysyttiin oppilaan käsityksiä sosiaalisiin taitoihin liittyen (Taulukko 4). Lukuun ottamatta kokonaissummaa (k = 10), kaikki osasummat ovat syntyneet vain kolmesta tai neljästä osiosta, mikä johtaa siihen, että perinteisellä alfakertoimella arvioiden summien erottelukyky jää matalaksi tai matalahkoksi (α_R= 0,51–0,69). Toisaalta *MJTR* ja *alfaU* viestivät, että summamuuttujat pystyvät järjestämään havainnot varsin uskottavaan järjestykseen (*MJTR* = 0,81–0,87; α_U= 0,72–0,81). Kokonaisasennetta raportoitaessa erottelukyky on parhaimmillaan (α_R= 0,77, α_U= 0,88, *MJTR* = 0,73).

TAULUKKO 4. Oppilaan käsityksiä koskevien mittareiden reliabiliteetit

	Suhtautuminen kokonaisuutena	Suhtautuminen matematiikkaan ja äidinkieleen	Suhtautuminen äidinkieleen	Suhtautuminen matematiikkaan	Itse koettu osaaminen	Itse koetut sosiaaliset taidot
Lukeminen on...	1	1	1			
Laskeminen on...	1	1		1		
Kirjoittaminen on...	1	1	1			
Tehtävät olivat helppoja	1		1	1	1	
Osaan lukea	1		1		1	
Osaan laskea	1			1	1	
Osaan kirjoittaa	1		1		1	
Osaan leikkiä toisten kanssa	1					1
Minulla on ainakin yksi kaveri	1					1
Osaan odottaa omaa vuoroani	1					1
alfaR	0,766	0,585	0,677	0,519	0,624	0,638
alfaU	0,876	0,727	0,810	0,722	0,789	0,799
MJTR	0,729	0,831	0,790	0,830	0,810	0,867

2.6 Huoltajakysely ja muu tiedonkeruu

Koska koulun aloittavien oppilaiden luku- ja kirjoitustaito on monella vasta alkavassa vaiheessa, valtaosa lapsen taustatietoihin liittyvistä asioista kysyttiin heidän huoltajiltaan. Katoanalyysin yhteydessä (ks. luku 5.2) pohditaan, kuinka edustava huoltajien aineisto on.

Huoltajakysely toteutettiin kahdeksalla Suomessa yleisimmin puhutulla kielellä. Huoltajakyselyssä kartoitettiin mm. lapsen kotikieltä, lähisuvun oppimisvaikeuksia, huoltajien koulutustaustaa, päivähoitohistoriaa, lapsen harrastuneisuutta ja puuhailua, esiopetukseen osallistumista ja esiopetuksen paikkaa sekä huoltajien käsityksiä omasta lapsestaan kuten tämän työskentelytavoista, vertaissuhteista, kiinnostuksen kohteista ja tottuneisuudesta käyttämään tietokonetta tai tablettia. Osaa näistä ja erityisesti näiden yhteyttä osaamiseen on mielekästä kuvata yksittäisten muuttujien lisäksi summamuuttujien avulla. Kolmesta kokonaisuudesta muodostettiin summamuuttujat: oppilaan lähisuvussa ilmenneiden oppimisvaikeuksien määrästä ($k = 5$), oppilaan harrastusten ja puuhailukohteiden määrästä ($k = 9$) sekä huoltajien käsitystä oppilaan persoonasta, joka faktorianalyysissa jakaantuu neljään faktoriin: FA1: *keskittyvä ja loppuun suorittava lapsi*, FA2: *kiinnostunut lapsi*, FA3: *sosiaalisesti taitava lapsi*, FA4: *tietoteknisesti tottunut lapsi*. Taulukkoon 5 on koottu faktorianalyysin avulla löydettyjä summamuuttujia koskevia tietoa. Varsinaiset summamuuttujat laskettiin summaamalla yhteen korkeimmin latautuneet muuttujat.

TAULUKKO 5. Lapsen persoonaa koskeva faktorirakenne (*N* = 4 316)

FA1	FA2	FA3	FA4	
0,813				Lapsi3: tekee aloittamansa tehtävän loppuun asti.
0,656				Lapsi1: tekee annetut tehtävät.
0,591				Lapsi2: keskittyy tekemäänsä.
	0,692			Lapsi8: on kiinnostunut tutkimaan tekstejä, sanoja, kirjaimia ja äänteitä.
	0,572			Lapsi11: on kiinnostunut luokittelemaan, vertailemaan ja laskemaan asioita.
	0,475			Lapsi9: on kiinnostunut musiikista ja rytmeistä.
	0,245			Lapsi10: on kiinnostunut fyysisestä toiminnasta, kuten kiipeilystä ja temppuilusta.
		0,63		Lapsi4: on tottunut työskentelemään ja leikkimään yhdessä toisten lasten kanssa.
		0,585		Lapsi6: on ainakin yksi läheinen ystävä.
		0,312		Lapsi5: käsittelee pettymyksiä vahingoittamatta itseä, toisia tai ympäristöä.
		0,255		Lapsi7käännetty: vertaissuhteissa EI esiintynyt kiusaamista, torjumista tai yksinäisyyttä.
			0,524	Lapsi12: on tottunut käyttämään tablettia.
			0,399	Lapsi13: on tottunut käyttämään tietokonetta.
				Hajotusmenetelmä: Maximum Likelihood. Rotaatiomenetelmä: Promax with Kaiser Normalization. Yksikertainen ratkaisu, jossa alle 0,22 suuruiset lataukset on jätetty näkyvistä

Oppilaan persoonaan liittyvästä faktorirakenteesta (Taulukko 5) huomataan, että faktorirakenne olisi vieläkin tiiviimpi ja ehkä parempi ilman osioita 10 (”Lapsi on kiinnostunut fyysisestä toiminnasta, kuten kiipeilystä ja temppuilusta”) ja 7 (”Lapsen vertaissuhteissa on esiintynyt kiusaamista, torjumista tai yksinäisyyttä”) käännettynä versiona. Lataukset jäävät mataliksi, mikä viittaa siihen, että nämä saattavat edustaa jotain omaa persoonan komponenttia, jota ei kyetty saamaan esiin tällä mittarikonstruktiolla. Ne kuitenkin pidettiin mukana summissa, koska ne loogisesti saattaisivat olla osana faktoria. Muuttujat alentavat perinteisen alfakertoimen arvoa ($\alpha_{R(FA2)} = 0{,}57$ ja $\alpha_{R(FA3)} = 0{,}46$ ks. Taulukko 6), mutta *alfaU* ja *MJTR* indikoivat, että erottelukyky on kuitenkin melko korkea ($\alpha_{U(FA2)} = 0{,}82$, $\mathrm{MJTR}_{(FA2)} = 0{,}86$ ja $\alpha_{U(FA3)} = 0{,}78$, $MJTR_{(FA3)} = 0{,}86$). Tietoteknisesti tottuneita lapsia kuvaava summamuuttuja on lyhyt ($k = 2$) ja tämä heijastuu reliabiliteetin estimaateissa ($\alpha_{R(FA4)} = 0{,}33$ ja $\alpha_{U(FA4)} = 0{,}61$), mutta vastaajien järjestys on looginen ($MJTR_{(FA4)} = 0{,}90$).

Kaksi muuta summaa, oppimisvaikeuksien kasautuneisuus lähisuvussa ja harrastusten ja kiinnostusten kohteiden määrä ovat suoria intensiteetin mittoja: suurempi arvo kuvaa suoraan yhtäältä sitä, kuinka monta erilaista oppimisvaikeustyyppiä lähisuvusta löytyy, ja toisaalta sitä, kuinka monta harrastusten tai puuhailujen kohdetta lapsella on (Taulukko 6). Mittareiden erottelukyvyt ovat riittäviä tai jopa erittäin korkeita ($\alpha_R = 0{,}63$ molemmille; $\alpha_U = 0{,}82–0{,}93$). Se, että oppimisvaikeuksien kasautuneisuuden summa on osioiden tasolla erittäin loogisesti järjestynyt ($MJTR = 0{,}95$) on seurausta siitä, että valtaosalla oppilaista ei ollut huoltajien mukaan lainkaan oppimisvaikeuksia lähisuvussa. Partioharrastukseen tai muihin kerhoihin osallistuminen alensi oleellisesti mittauksen reliabiliteettia johtuen pienestä harrastajien määrästä. Niinpä se jätettiin pois summasta.

TAULUKKO 6. Huoltajakyselystä laskettuja summamuuttuja (*N* = 4 316)

summamuuttuja	osioiden määrä (*k*)	piste määrä	reliabiliteetti *alfaR*	reliabiliteetti *alfaU*	*JTR* keskiarvo
Oppimisvaikeuksien kasautuneisuus lähisuvussa	5	0–5	0,630	0,929	0,955
Harrastusten ja kiinnostusten kohteiden määrä[1]	9	0–9	0,634	0,824	0,829
Lapsen persoona FA1: keskittyvä ja loppuun suorittava lapsi	3	1–4	0,728	0,915	0,951
FA2: kiinnostunut lapsi	4	1–4	0,567	0,818	0,861
FA3: sosiaalisesti taitava lapsi	4	1–4	0,460	0,780	0,863
FA4: tietoteknisesti tottunut lapsi	2	1–4	0,331	0,612	0,896

1) partioharrastus ei mukana

Aineiston keruun yhteydessä koottiin tietoa myös opettajilta ja rehtoreilta. Näitä tietoja hyödynnetään kuitenkin vasta seuraavalla mittauskerralla kolmannen luokan mittauksen yhteydessä, sillä opettajan vaikutus mittausajankohtana aivan koulun alkuvaiheessa on ollut marginaalista eikä tiedoilla ole mielekästä selittää osaamisen tasoa. Tiedoilla on kuitenkin merkitystä, kun arvioidaan opettajan toimien vaikutusta kahden ensimmäisen kouluvuoden aikana.

3

Osioanalyysiin ja vertaistamiseen liittyviä erityis-kysymyksiä

3

3.1 Osioanalyysin vaikutus lopullisiin mittareihin

Ennen kuin mittarin lopullisia summia lasketaan, on mielekästä tehdä tarkentavia osioanalyyseja (ks. Metsämuuronen, 2009a), ettei lopullisiin mittareihin laiteta toimimattomia tehtäviä. Osiot oli jo esitestattu varhaisemmassa vaiheessa vuonna 2017 pienellä otoskoolla, mutta varsinaisen testauksen aikana saavutettava suurempi joukko vastaajia saattaa paljastaa osioista uusia piirteitä, jotka heikentävät niiden käytettävyyttä. Esitestauksen jälkeen oppilaiden tekemään testiin valittiin 122 osiota, joista osa oli erittäin helppoja, jotta mahdollisimman moni olisi voinut saada onnistumiskokemuksia. Osa näistä tehtävistä jäi pois lopullisen osioanalyysin jälkeen, jos ne olivat erottelukyvyltään heikkoja.

Lopulliseen osioanalyysiin käytettiin sekä klassisen että modernin testiteorian mahdollistamia keinoja. Klassisen testiteorian osalta osioista laskettiin perinteisiä vaikeustason ja erottelukyvyn parametreja kuten ratkaisuosuus (p), osio–summa-korrelaatio (ρ_{gX}) ja korjattu osio–summa-korrelaatio (ρ_{gP}). Lisäksi klassisen analyysin piiriin kuuluvat perinteiset ja uudet reliabiliteettikertoimet (ks. luku 2.3.2). Modernin testiteorian osalta analyysissa käytettiin Item Response Theory (IRT)-mallitusta (Birnbaum, 1968) ja tarkemmin yksiparametrista Raschin mallitusta (Rasch 1960) OPLM-ohjelmistolla (Verhelst, Glas, & Verstralen, 1995). IRT-mallituksen vahvuus tässä arvioinnissa tulee yhtäältä sen tuottamista osion ominaiskäyristä (Item Characteristic Curve, ICC), jotka mahdollistavat osioiden tehokkaan graafisen diagnostiikan (ks. Metsämuuronen 2018d), ja toisaalta eri summien vertaistamiseen liittyvistä näkökulmista (ks. luku 3.2).

Osana modernia osioanalyysia tehtiin ns. DIF-analyysi (Differential Item Functioning) sen varalta, onko osioissa vastaamisen peruslogiikkaa erilainen erilaisissa ryhmissä. Kiinnostavina vertailuryhminä olivat sukupuoli ja kieliryhmä. Tehtävä saattaa suosia poikia tai tyttöjä (tai suomen ja ruotsinkielisiä oppilaita) mikä näkyy siinä, että poikien tai tyttöjen (tai suomen ja ruotsinkielisten) käyttäytyminen poikkeaa toisistaan jonkin erityisen osion osalta. Analyysi tehtiin graafisesti osion ominaiskäyrien perusteella. Sukupuolten välillä ei havaittu ilmeisiä poikkeavia vastausmalleja. Sen sijaan suomen- ja ruotsinkielisten versioiden välillä oli oleellisia eroja kahdessa osiossa,

joiden osalta testi suosi selvästi suomenkielisiä oppilaita. Tehtävissä käytettiin ruotsinkielisessä versiossa liian epäfrekventtejä sanoja, joita oppilaat eivät selvästikään tunteneet ja näin juuri kukaan ruotsinkielisistä oppilaista ei saanut tehtäviä oikein. Nämä osiot poistettiin lopullisesta tehtäväsarjasta mittauksen jälkeen ennen summamuuttujien muodostamista.

Kahden DIF-analyysissä poisjääneen osion lisäksi poistettiin viisi tehtävää, joissa erottelukyky oli joko patologinen ($\rho_{gX} < 0$) tai hyvin lähellä nollaa ($\rho_{gX} < 0{,}125$) tai joissa oli patologinen arvauskynnys. Näitä osioita luonnehtii siis se, että osioihin oikein vastaaminen oli satunnaista tai niissä oikean vastauksen arvaaminen oli liian helppoa. Näiden osioiden osalta päätös jättää ne pois lopullisesta tehtäväsarjasta tehtiin yhdistäen klassista ja modernia testiteoriaa. Tehtäväsarjaan kuitenkin jätettiin muutamia erittäin vaikeita tehtäviä, joiden erottelukyky ei ollut kovin korkea klassisilla indikaattoreilla mitattuna. Näitä tehtäviä käytetään mahdollisesti ankkuritehtävinä seuraavaa, 3. luokan, mittausta toteutettaessa.

Kaikkiaan oppilaan tekemässä varsinaisessa tehtäväsarjassa oli siis 122 osiota, joista 115 jäi lopulliseen sarjaan, jonka perusteella summamuuttujat muodostettiin (ks. taulukko 2).

3.2 Vertaistaminen ja muuttujamuunnokset

Eri vuosien testitulosten ja saman mittauskerran eri testiversioiden pistemäärät on saatettava yhteismitalliseksi, ennen kuin vertailua on mielekästä tehdä. Tätä kutsutaan vertaistamiseksi (engl. equating, ks. Béguin, 2000): sopivasti valittujen, eri mittauskerroille tai mittariversioille yhteisten linkkitehtävien avulla voidaan tehtäväsarjojen pistemäärät saattaa vertailukelpoisiksi. Vertaistaminen ei tässä alkumittauksessa ole merkityksellistä, mutta se on oleellista, kun oppilaiden osaamista mitataan uudelleen heidän ollessaan 3. luokan alussa vuoden 2020 syksyllä. Tämän mittauksen yhteydessä käytettiin kuitenkin hyödyksi vertaistamisen taustalla olevaa IRT-mallitusta ja tarkemmin Raschin mallitusta (Rasch 1960), jotta osamittareiden pistemäärät saatettiin samaskaalaisiksi. Käytetty mallitusmetodi on sama kuin aiemmissa pitkittäisvertailuissa (mm. Metsämuuronen 2010a; 2013a; 2017a). Kansainvälisissä PISA- (mm. OECD, 2014, 2016) ja TIMSS-tutkimuksissa (mm. TIMSS, 2012, 2016) käytetään laajempaa mallitusta ja hieman eri metodologia, joskin siellä käytettävä menettely on pitkälti vertailukelpoinen tässä esitettävän menettelyn kanssa.

Summamuuttujat muunnetaan samaskaalaisiksi hyödyntämällä sitä IRT/Raschin mallin ominaisuutta, että oppilaiden taustalla oleva (latentti) osaamisen taso (Theta, θ) ja tehtävän vaikeustaso (Beta, β) ovat identtiset, kun tietyt ehdot täyttyvät (ks. Metsämuuronen, 2009b, 163–165). Kun tiedetään, kuinka vaikeaan tehtävään testattavat pystyvät antamaan oikean vastauksen (50 %:n todennäköisyydellä), eli mikä on tehtävän β-parametrin arvo, tiedetään samalla, mikä on sellainen vastaajan osaamisen taso (θ), jolla tehtävän pystyy ratkaisemaan. Kun siis summapistemäärän perusteella estimoidaan vastaajan osaamisen taso (θ), tiedetään, minkä tasoisiin tehtäviin (β) hän pystyy vastaamaan todennäköisesti oikein. Itse laskenta tehtiin yksiparametrisesti OPLM-ohjelmistolla (Verhelst, Glas & Verstralen, 1995). Yksiparametrisuus tarkoittaa, että vain tehtävien vaikeustasot kalibroidaan samalle asteikolle. Osioiden vaikeustasoparametrin ja oppilaan alustavan

osaamisen tason perusteella kullekin oppilaalle lasketaan ns. latentti osaamisen taso θ, joka on standardoidun normaalijakauman pistemäärä ja vaihtelee välillä −4 ja +4 hajonnan ollessa 1. Tässä vaiheessa prosessia tällä asteikolla osaamiseltaan keskitasoisin oppilas saa arvokseen 0.

Aiemmissa oppimistulosarvioinneissa osaamista ja osaamisen muutosta kuvattiin ratkaisuprosentteina maksimipistemäärästä, joka muodostettiin edellä kuvatun theta-arvon perusteella. Viimeaikaisissa raporteissa osaaminen esitetään kuitenkin samalla skaalalla kuin PISA- ja TIMSS-tutkimuksissa ja aiemmassa pitkittäisarvioinnissa (Metsämuuronen, 2010b; 2013b; 2017a). Asteikko on ns. 10T-muunnos (500+100×muuttuja) edellä lasketusta standardoidusta normaalijakauman pisteestä. Tässä uudessa asteikossa osaamiseltaan keskitasoisin oppilas saa arvokseen 500 ja muuttujan keskihajonta on 100. Aineistossa jokainen osasumma saa keskiarvokseen 500 ja näin eri summamuuttujien tasoja ei pysty siis vertaamaan keskenään. Tähän arvoon 500 verrataan tulososassa eri ryhmiä kuten tyttöjä ja poikia tai suomen- ja ruotsinkielisiä oppilaita. Kun siis poikien keskimääräinen kokonaisosaamisen taso koko tehtäväsarjassa oli 493 ja tyttöjen 507, eroa oli 14 yksikköä 10T- tai PISA-asteikolla mitattuna. Yhtäältä emme tiedä onko tämä ero todellista vai laskettavissa vielä satunnaisvaihtelun piiriin. Toisaalta emme myöskään tiedä onko ero suurta vai pientä. Näiden arvioimiseksi varten tarvitaan tilastollisia apuvälineitä: tilastollista merkitsevyyttä ja efektikokoa, joita käsitellään lyhyesti luvussa 4.

4

Termit ja menetelmät

4

4.1 Käytettävät termit

Analyyseissä käytetään tilastolliseen testaukseen liittyvää termiä ”tilastollisesti merkitsevä” kuvaamaan sitä, kuinka luotettavasti ryhmien välillä on eroa populaatiossa eli muissakin kuin vain otoskouluissa. Otoksessa voi ryhmien välillä olla pieniä eroja, mikä voi olla otoksesta johtuvaa satunnaista vaihtelua. Tilastollisessa testauksessa – kuten esimerkiksi varianssianalyysissa – keskiarvojen eroja testataan aineiston kokoon ja mittaustapaan sopivalla menetelmällä. Kun tekstissä kerrotaan eron kahden tai useamman ryhmän välillä olevan tilastollisesti merkitsevä, se tarkoittaa, että ero tulisi todennäköisesti näkyviin otoksesta riippumatta. Tällöin virhepäätelmän riski on hyvin pieni: riski voi olla esimerkiksi korkeintaan 5 prosenttia, jolloin tekisimme saadusta tuloksesta poikkeavan päätelmän vain viisi kertaa, kun samanlaisia otoksia olisi sata. Tämän indikaattorina tekstissä käytetään merkintää $p = 0{,}05$. 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. Tiedämme siis, että tyttöjen ja poikien kokonaisosaamisen ero on 14 pistettä. Tilastollinen testaus (esimerkiksi t- tai F-testi) kertoo, että eron merkitsevyys on $p < 0{,}001$. Tämä kertoo, että ero poikien ja tyttöjen välillä on populaatiossa todellinen ja että virhepäätelmän riski on pieni. Tilastollinen merkitsevyys ei kuitenkaan kerro onko ero suurta vai pientä – kun otoskoko on suuri, mitätönkin ero voi osoittautua merkitseväksi. Eron suuruuden hahmottamiseksi käytämme toista tilastolukua: efektikokoa.

Ero ryhmien välillä voi olla tilastollisesti merkitsevä – eli ryhmien välillä on melko varmasti eroa riippumatta siitä, millainen satunnainen otos valitaan – mutta ero ei välttämättä ole suurta. Efektikoko kertoo sen, kuinka suurta ryhmien välinen ero on. Jos tyttöjen ja poikien keskiarvot olisivat samat ja jakaumat samanlaiset, efektikoko on nolla. Jos taas esimerkiksi poikien tulos olisi tyttöjen tulosta niin paljon parempi tai heikompi, että 80 % pojista sijoittuu tyttöjen keskiarvon ylä- tai alapuolelle, efektikoko on suuri. Efektikoon mittana käytetään raporteissa ensisijaisesti Cohenin f- ja d-mittoja (Cohen, 1988), koska niitä on helppo verrata eri aineistojen välillä ja koska niille on olemassa karkeita rajoja kuvaamaan efektikoon pienuutta tai suuruutta. Cohenin f on

käytössä, kun vertaillaan kahta tai useampaa keskiarvoa ja Cohenin *d*, kun kuvataan korrelaation suuruutta. Karkeat rajat efektikoon suuruudelle on esitetty Taulukossa 7 (ks. Cohen, 1988, ss. 285–287 *f*:lle ja ss. 25–27 *d*:lle).

TAULUKKO 7. Efektikokojen rajoja

efektikoon indikaattori	pieni efektikoko	keskisuuri efektikoko	suuri efektikoko
Cohenin *f*	noin 0,1	noin 0,2 – 0,3	> 0,4
Cohenin *d*	noin 0,2	noin 0,4 – 0,5	> 0,8

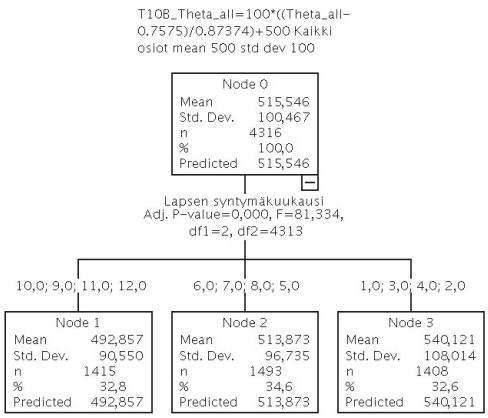
Korrelaatioiden ja regressiomallien yhteydessä käytetään edellisten lisäksi termiä *selitysaste*, joka kertoo, kuinka monta prosenttia muuttujat selittävät toistensa vaihtelusta. Kun kaksi muuttujaa on täydellisessä yhteydessä toisiinsa, korrelaatio (r) muuttujien välillä on $r = 1$. Tällöin riittää, kun tiedetään toinen muuttujista; toinen selittää täydellisesti toisen – selitysaste on 1,00 eli prosentteina ilmaistuna 100 %. Mikäli korrelaatio puolestaan olisi suuruudeltaan $r = 0{,}40$, selitysaste olisi $r^2 = 0{,}4 \times 0{,}4 = 0{,}16$ eli muuttujat selittäisivät toisistaan 16 % ja ilmiöstä jäisi selittymättä 84 %. Varianssianalyysin yhteydessä selitysasteena käytetään Eetan neliötä (η^2, Pearson, 1903; 1905) tai osittaista eetan neliötä (partial eta-squared, η_p^2, ks. Pierce, Block & Aguinis, 2002), kun kyseessä on useita selittäviä tekijöitä. Regressiomallien yhteydessä kuvataan selitysasteena multippelikorrelaatiokertoimen neliö R^2. Kun muuttujia on mallissa useampia kuin yksi, R^2 antaa hieman liian suuren arvion selitysasteesta, koska korrelaatiokertoimen neliö johtaa aina positiiviseen suuntaan menevään satunnaiseen vaihteluun. Tätä korjataan ohjelmistoissa yleisesti Wherryn (1931) korjauksella, jota merkitään tekstissä symbolilla R^2_{Adj}.

4.2 Analyysimenetelmät

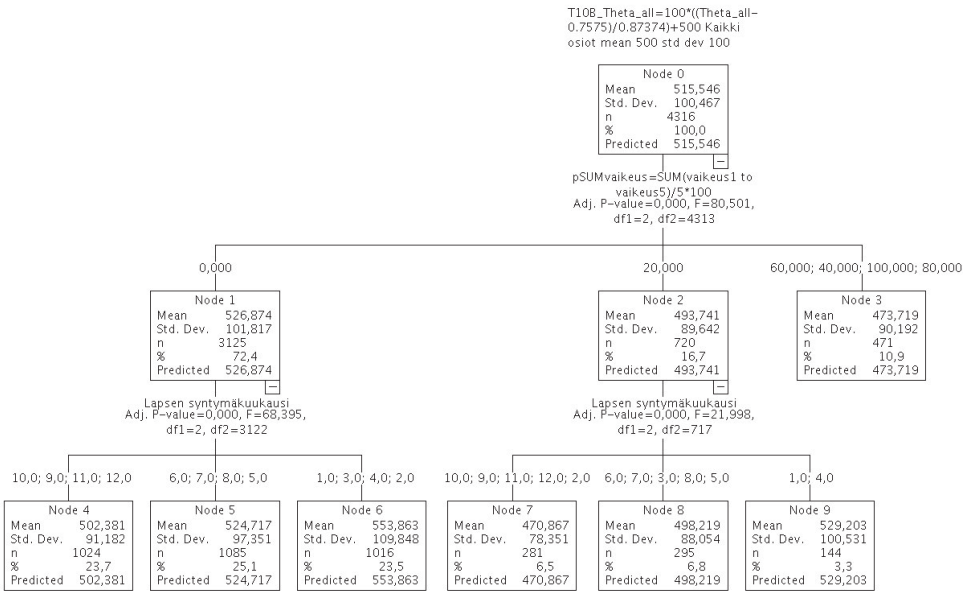
Ryhmien välisiä eroja kuvataan yksinkertaisilla tilastollisilla tunnusluvuilla, kuten osaamisen keskimääräisinä tasoina ja niiden muutoksina. Aineiston analysoinnissa käytetään yleisesti tunnettuja menetelmiä (ks. esimerkiksi Tabachnik & Fidell, 2006; Metsämuuronen, 2017b tai suomeksi esimerkiksi Metsämuuronen, 2009b). Perustunnuslukujen ja korrelaatiokertoimien lisäksi käytetään varianssianalyysin (GLM), regressioanalyysin ja luokitteluanalyysien eri muotoja. GLM on käytössä, kun tarkastellaan keskiarvojen eroja. Ryhmien välisten erojen post hoc -vertailussa käytetään Šidákin tai Tukeyn parittaisvertailua (Šidák, 1967; Tukey, 1953). Näistä Šidákin menettely on tarkempi ja Tukeyn menettely konservatiivisempi (ks. keskustelu Metsämuuronen, 2017b, ss. 340–341, 348–349). Regressioanalyysiä käytetään, kun osaamista selitetään monilla taustamuuttujilla yhtä aikaa. Luokitteluanalyyseista keskeisin on puuanalyysi (Decision Tree Analysis, DTA, mm. Kass, 1980), jota käytetään oleellisten selittävien muuttujien jakokohtien löytämisessä ja ns. numeronmurskaamisessa. Tämä menetelmää kuvataan hieman tarkemmin seuraavassa.

DTA on joukko menetelmiä, joiden avulla analysoidaan laajoja aineistoja ja luokitellaan selittäviä muuttujia (Independent variables) kiinnostavan kohdemuuttujan (Dependent Variable), kuten osaamisen suhteen. Menetelmä on erittäin tehokas tilanteissa, joissa ei välttämättä ole olemassa olevaa teoriaa – tai sitä ei nähdä tarpeelliseksi käyttää – kertomaan, miten selittävät muuttujat pitäisi ryhmitellä, jotta kohdemuuttuja voitaisiin selittää mahdollisimman hyvin. DTA tekee mekaanisesti kaikki mahdolliset muuttujien väliset ryhmittelyt ja valitsee niistä tilastollisin perustein parhaan mahdollisen. Tyypillinen hierarkkinen tulos voisi näyttää esimerkiksi kuvan 8 kaltaiselta. Kuvan 8 mallissa kokonaisosaamista (*10TB_Theta_all*) on selitetty huoltajakyselyistä saatavalla lapsen syntymäkuukaudella (*Lapsen syntymäkuukausi*). Huomataan, että huoltajien aineistoon mukaan tulleiden huoltajien (n = 4 316) lapset ovat hieman parempia kuin koko aineistossa (keskiarvo 515,546, kun se koko aineistossa oli 500). DTA luokittelee syntymäkuukaudet kaikkiin mahdollisiin kombinaatioihin ja huomaa, että merkitsevin ero oppilaiden välillä (F = 81,33; p < 0,001) syntyy kombinaatiolla, että ikäkuukaudet jaetaan kolmeen ryhmään: tammikuusta huhtikuuhun (540), toukokuusta elokuuhun (514) ja syyskuusta joulukuuhun (osaaminen 493). Menetelmä on herkkä muuttujien valinnalle: yhdenkin muuttujan lisääminen tai poistaminen mallista voi muuttaa tulosta oleellisesti. Jos malliin otetaan mukaan muita muuttujia – kuten esimerkiksi tieto siitä, kuinka paljon lähisuvussa on ollut oppimisvaikeuksia – DTA saattaa aistia, että lapsen syntymäkuukausi on merkityksellinen tekijä vain joissain erityisryhmissä (Kuva 9).

Kuvassa 9 huomataan, että DTA:n mukaan syntymäkuukautta parempi selittäjä onkin lähisuvun oppimisvaikeuksien määrä. Mikäli oppimisvaikeudet ovat kasaantuneet niin, että lähisuvussa on 2–5 erilaista oppimisvaikeutta (40–100 % kaikista mittariin mukaan otetuista mahdollisuuksista), syntymäkuukaudella ei ole merkitystä – osaaminen jää merkitsevästi matalammaksi (473) kuin ryhmässä, jossa lähisuvusta löydetään vain yhtä oppimisvaikeutta (492) puhumattakaan siitä, että lähisuvussa ei olisi lainkaan oppimisvaikeuksia (527). Toisaalta syyskuun alun ja joulukuun lopun välisenä aikana syntyneellä oppilaalla on suuri riski saada keskimääräistä heikompi tulos (471), vaikka hänen lähisuvussaan ei olisikaan kuin yhtä oppimisvaikeustyyppiä.



KUVA 8. Esimerkki yksinkertaisesta DTA mallista



KUVA 9. Esimerkki hierarkkisesta DTA-mallista

4.3 Painotus ja ryvästyminen

Aineistossa on kaksi ominaisuutta, joilla on merkitystä analyysituloksiin ja joiden vuoksi tuloksia joudutaan korjaamaan. Yhtäältä aineiston otannasta johtuen tietyt ryhmät ovat yli- tai aliedustettuina. Esimerkiksi ruotsinkielisiä oppilaita on aineistossa enemmän (11 %) kuin heidän suhteellinen osuutensa populaatiossa edellyttäisi (6 %). Kun aineiston analyysin tarkoituksena on kuitenkin kuvata, kuinka osaaminen on jakautunut varsinaisessa populaatiossa eikä vain otoksessa, tätä yli- ja aliedustusta korjataan painotuksen avulla. Painotus voidaan tehdä eri taustamuuttujakombinaatioittain. Tässä aineistossa jokainen oppilas saa painokertoimen sen perusteella, mitä kieliryhmää (suomi/ruotsi), suuraluetta (Helsinki ja Uusimaa/Etelä-Suomi/Länsi-Suomi/Pohjois- ja Itä-Suomi) ja kuntaryhmää (kaupunkimaiset kunnat/taajaan asutut kunnat/maaseutumaiset kunnat) oppilas edustaa.[6] Näin esimerkiksi Helsingissä tai Uudellamaalla kaupunkimaisessa kunnassa koulunsa aloittava ruotsinkielinen oppilas saa painokseen 0,42, kun vastaavasti suomenkielinen oppilas saa painokseen 1,09. Painokertoimen käyttö analyyseissa korjaa keskiarvoja vain hieman – otos edustaa alun perinkin varsin hyvin populaatiota. Esimerkiksi oppilaiden kokonaisosaamisen taso maaseutumaisissa kunnissa oli korjaamattomana 494 pistettä kun se painotettuna oli 492 pistettä. Vaikka painotuksella ei ole välttämättä suurtakaan vaikutusta kokonaiskeskiarvoon, sillä on pieni vaikutus kuitenkin merkitsevyystestaukseen: painotettuna ryhmien välinen ero näyttää hieman

6 Mikäli koulun sisällä olisi tehty otanta – kuten yleensä kansallisessa arvioinnissa on tehty – jokaiselle oppilaalle olisi tullut vielä erikseen koulua koskeva painokerroin. Koska otokseen valituista kouluista mukaan tulivat kaikki oppilaat, kaikilla saman alueen kouluilla on yhtäläinen painokerroin.

selvemmältä kuin ilman painotusta, joskin molemmissa tapauksissa ero on erittäin merkitsevä ($p < 0{,}001$) – ero näkyy varsinaisen testisuureen arvossa ($F = 6{,}544$ vs. $8{,}748$). Merkitsevyystestaus on kuitenkin harhaanjohtava, sillä aineistoa kuvaa toinenkin ilmiö, jolla on tuloksiin suurempi vaikutus kuin painotuksella: ryvästyminen.

TAULUKKO 8. Esimerkki painotuksen ja ryvästymisen vaikutuksesta analyysituloksiin

		kokonaisosaaminen		monitasomallituksella korjattu analyysi	
Kuntaryhmä		**painottamaton[1]**	**painotettu[1]**	**painottamaton[2]**	**painotettu[2]**
Kaupunkimaiset kunnat		502,66	503,29	500,04	500,52
Taajaan asutut kunnat		493,29	494,23	492,39	493,15
Maaseutumaiset kunnat		494,12	491,64	496,22	495,09
yhteensä		500,00	500,13		
		merkitsevyystestaus			
	F	6,544	8,748	1,208	1,311
	p	0,001	< 0,001	0,300	0,271

1) Todellinen oppilaskeskiarvo
2) REML-keskiarvojen keskiarvoina estimoitu marginaalikeskiarvo (estimated marginal means)

Kouluaineistot ovat lähtökohtaisesti ryvästyneitä. Tämä johtuu siitä, että kouluaineistojen yhteydessä käytetään yleensä ryväsotantaa, jossa samasta koulusta valitaan useita oppilaita, joiden keskiarvon avulla kuvataan kyseisen koulun ominaisuuksia (kuten keskimääräisiä oppimistuloksia). Ryväsotannasta – jos oppilaiden valinta on satunnaista – seuraa, että koulun keskiarvo on kyllä tarkka kuvaamaan koulun keskimääräistä osaamista (ja näin kansallista keskiarvoa), mutta kansallisten johtopäätösten näkökulmasta oppilaiden vaihtelu (varianssi) on pienempää kuin tilanteessa, että yhtä monta opiskelijaa olisi valittu satunnaisesti eri puolilta Suomea eri kouluista. Tämä johtuu siitä, että tietyn koulun oppilaita yhdistävät samat olosuhteet, mahdollisesti sama asuinalue ja sama sosioekonominen status, samat opettajat ja samat koulun käytänteet. Tällä ryvästymiseksi kutsutulla ilmiöllä on vaikutusta erityisesti tilastollisten johtopäätösten tekemiseen, jotka usein perustuvat oppilaiden vaihteluun. Tästä syystä ryvästynyttä aineistoa analysoidaan monitasomallituksella (Goldstein 1986; Bryk & Raudenbush 1987; Raudenbush & Bryk 2002; ks. suomeksi esimerkiksi Metsämuuronen 2008; 2009b), kun halutaan saada tarkkoja tuloksia. Monitasomallitus suoritettiin teknisesti SPSS:n Linear Mixed Models toiminnolla REML (Restricted Maximum Likelihood) menetelmällä, joka tuottaa sama tuloksen kuin erityisohjelmistojen (HML ja MLwiN) vastaavat (ks. Metsämuuronen, 2009b).

Monitasomallituksenkin tulos tarkentuu, kun analyysissa käytetään painokerroinmuuttujia. Taulukosta 8 huomataan, että monitasomallitus muuttaa tilastollisen tulkinnan oleellisesti: kun otetaan huomioon aineiston ryvästyminen, ero kuntaryhmien välillä ei ole tilastollisesti merkitsevä. Tämä johtuu teknisesti siitä, että maaseutumaisissa kouluissa oppilasmäärät ovat pienempiä ja näissä yksiköissä ryvästysmiskorjausta tapahtuu pienemmässä määrin ja ryhmien väliset estimoidut keskiarvot ovat lähempänä toisiaan. Monitasomallinnus aistii siis koulujen väliset erot selvästi pienemmiksi kuin mitä aineiston oppilaskeskiarvojen perusteella päätellään.

Lähtökohtaisesti kuitenkin painotetut oppilaskeskiarvot ovat tarkkoja kuvaamaan koulujen keskiarvoa ilman monitasomallitustakin.[7] Niinpä analyyseissä kuvataan kansallisina keskiarvoina aineiston painotetut keskiarvot eikä monitasomallituksen estimoimia marginaalikeskiarvoja, mutta tilastollinen päättely perustuu monitasomallituksen korjaamiin merkitsevyyksiin.

7 Monitasomallituksen tuottamat keskiarvot ovat herkkiä otoskoiden vaihtelulle. Siksi kansainvälisissä vertailuissa kaikista otokseen valituista kouluista pyritään valitsemaan yhtä paljon oppilaita. Jos koulujen keskiarvot poikkeavat voimakkaasti toisistaan, kuten alkumittauksen aineistossa käy, monitasomallituksen tuottamat keskiarvot ovat harhaisia. Laskumekanismi ei kykene erottamaan toisistaan kouluja, joissa on 2 oppilasta tai 100 oppilasta – molempia käsitellään kuin koulussa olisi vain yksi oppilas. Tällä mekanismilla on oleellinen vaikutus siihen, miksi Taulukossa 8 keskiarvot poikkeavat toisistaan näinkin paljon.

Otos ja aineisto

5

5.1 Aineistojen koonti ja yhdistäminen

Merkittävä rajaus otosta valittaessa oli, että muun kuin suomen- ja ruotsinkieliset koulut jätettiin arvioinnin ulkopuolelle. Näin tarkastelun ulkopuolelle jäi automaattisesti kahdeksan englanninkielistä ja muun kielistä koulua. Lisäksi varsinaisen populaation ulkopuolelle jäivät erityiskoulut, mutta näistä tehtiin erillisotos. Kaikkiaan kokonaisotokseen kuului 269 oppilaitosta, joista viisi oli erityiskouluja. Varsinaiseen otokseen kuului siis 264 koulua kattavasti koko manner-Suomen alueelta, mikä vastaa reilua 13 prosenttia kaikista ensimmäisen vuosiluokan perusopetusta antavista kouluista (N = 1 969). Ruotsinkielisiä kouluja oli mukana 53 mikä vastaa 28,2 % ruotsinkielisistä kouluista. Vastaavasti suomenkielisiä kouluja oli 12,2 % kaikista kouluista. Vuosiluokan 1 opetusta antavia erityiskouluja oli mittausajankohtana Vipunen-rekisterin mukaan kaikkiaan 47 ja tältä osin saavutettavuusprosentti oli 10,6 %. Oppilasaineistossa saavutettavuus oli 18,8 %.

Lähtökohtaisesti otokseen valitun koulun kaikki koulunsa aloittavat oppilaat kuuluivat otokseen. Oppilasaineistoa otostettaessa mukaan haluttiin peruspopulaation lisäksi myös erityiskoulujen oppilaita. Arviointiin osallistui myös oppilaita, joille oli jo ennen koulun alkua tehty henkilökohtainen opetuksen järjestämistä koskeva suunnitelma (HOJKS), johon oli kirjattu yksilöllinen oppimäärä joko matematiikassa tai äidinkielessä ja kirjallisuudessa. Kokonaisaineistoon kuului 7 877 oppilasta, joista 65 tuli erityiskouluista ja joiden lisäksi 42:lla oli HOJKS joko matematiikassa tai äidinkielessä. Nämä kaksi erityisryhmää käsitellään aineistoa kuvattaessa erillisinä ryhminä. Siten varsinaiseen otokseen kuului 7 770 oppilasta. Sekä varsinaisen että erityisen oppilasaineiston peruspiirteitä kuvataan luvussa 5.3.

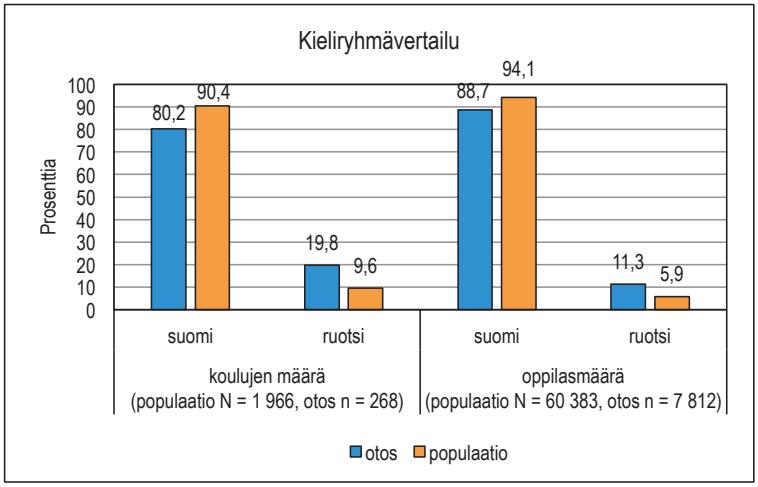
Vaikka lähtökohtaisesti otokseen valituista kouluista kaikki koulun aloittavat oppilaat vastasivat tehtäväsarjaan, oppilaiden huoltajille tehdyn taustakyselyn perusteella tiedetään kuitenkin, että kaikki oppilaat eivät osallistuneet tiedonkeruuseen, mikä oli tietysti odotettavissakin esimerkiksi sairastumisten tai muiden poissaolojen takia. Kaikkiaan 4 564 huoltajaa (56 % kaikista mittaukseen osallistuneista oppilaista) vastasi taustakyselyyn, joskaan huoltajista 196:tta ei pystytty yhdistämään oppilastietoihin, koska oppilas ei ollut osallistunut varsinaiseen arviointiin. Lisäksi osa

huoltajista edusti erityis- ja HOJKS-oppilaiden erityisryhmiä. Varsinaisessa aineistossa huoltajia oli kaikkiaan 4 316, mutta kaikilta ei ole kaikkia vastauksia ja näin muuttujakohtaiset otoskoot saattavat poiketa tästä hieman. Huoltaja-aineiston erityispiirteitä kuvataan luvussa 5.3.

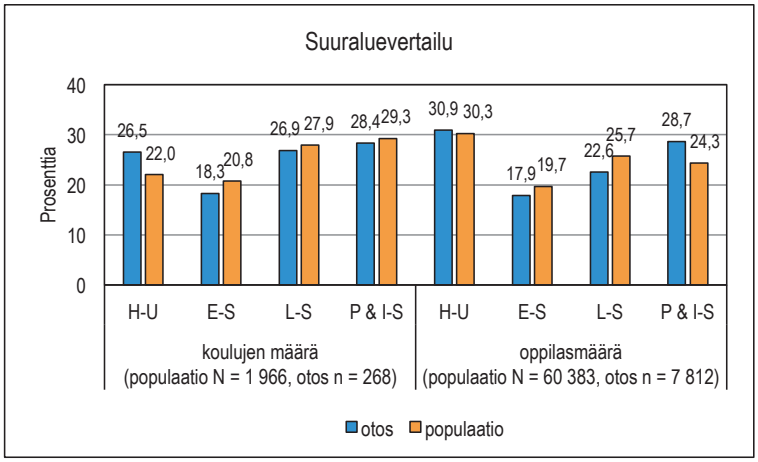
5.2 Otos ja kato

Aineiston harvat otoksesta poisjääneet oppilaat (n > 196, n. 2,5 %) eivät poikkea merkitsevästi muusta joukosta minkään vanhempainkyselyssä mukana olleen muuttujan osalta. Poisjäänti ei siis näytä olevan systemaattisesti tuloksiin vaikuttavaa. Ehkä kiintoisampi vertailu syntyykin siitä, kuinka koko aineisto edustaa populaatiota.

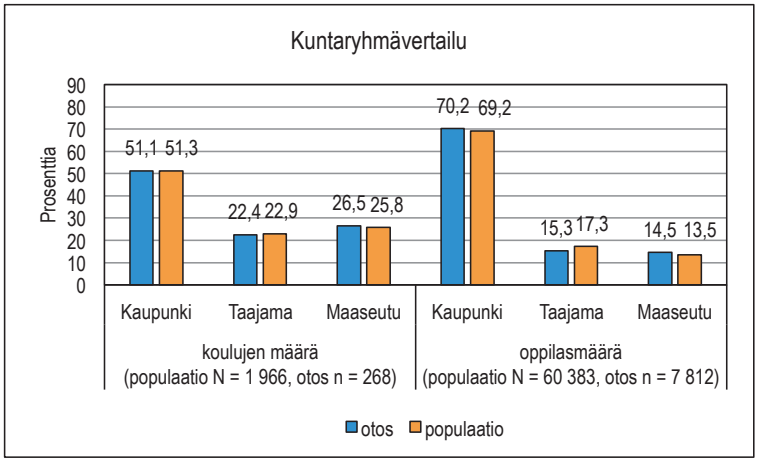
Yhteensä Vipunen-koulurekisterin mukaan (otosta tehtäessä viimeisimmän eli vuonna 2016 vahvistetun tiedon mukaan) ensimmäisen luokan opetusta annettiin 1 966 koulussa yhteensä 60 383 oppilaalle. Otoksen perusteella huomattiin, että Vipusen antama oppilasmäärätieto on saattanut muuttua oleellisestikin. Tätä rekisteritietoa pidetään kuitenkin vertailuaineistona. Kuvassa 10 havainnollistuu se, että ruotsinkielisiä on yliedustettu selvästi (oppilaita 6 % populaatiossa ja 11 % otoksessa). Kuvassa 11 havainnollistuu puolestaan se, että suuralueista Pohjois- ja Itä-Suomen (P-S ja I-S) alueella aineisto on yliedustunut (oppilaita 24 % populaatiossa ja 29 % otoksessa) ja vastaavasti Länsi- ja Etelä-Suomen (L-S ja E-S) alueilla aliedustunut (oppilaita 20 ja 26 % populaatiossa ja 18 ja 23 % otoksessa). Sekä kieliryhmiä että suuralueita koskevat yli- ja aliedustukset tulevat korjattua painokerrointen avulla. Kuvasta 12 havaitaan, että kuntaryhmien suhteen otos vastaa alun perinkin populaatiota varsin tarkasti.



KUVA 10. Kieliryhmien väliset erot otoksessa ja populaatiossa



KUVA 11. Suuralueiden väliset erot otoksessa ja populaatiossa



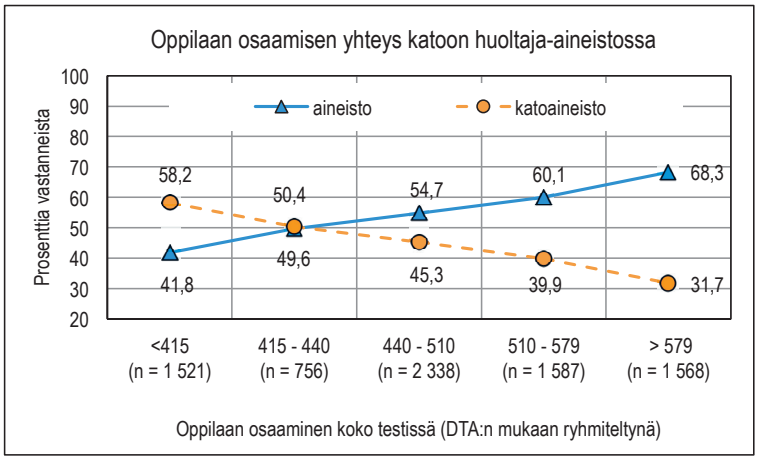
KUVA 12. Kuntaryhmien väliset erot otoksessa ja populaatiossa

Voidaan päätellä, että oppilasaineistosta saatava tieto oppilaiden osaamisesta koulun alkuvaiheessa on hyvin yleistettävissä koko populaatioon erityisesti, kun käsitellään painotettuja tuloksia ja tunnuslukuja. Sen sijaan on syytä olla enemmän varovainen, kun selitetään osaamisen eroja huoltaja-aineiston perusteella. Huoltaja-aineiston vastaajien lapset edustavat nimittäin selvästi paremmin menestyneitä oppilaita kuin vastaamatta jättäneiden huoltajien lapset. Vastanneiden ryhmässä oppilaiden kokonaisosaamisen keskiarvo oli 516 kun se vastaamatta jättäneiden ryhmässä oli 480. Ero on merkitsevä ($p < 0{,}001$) ja efektikooltaan keskisuuri ($f = 0{,}18$). DTA jakaa

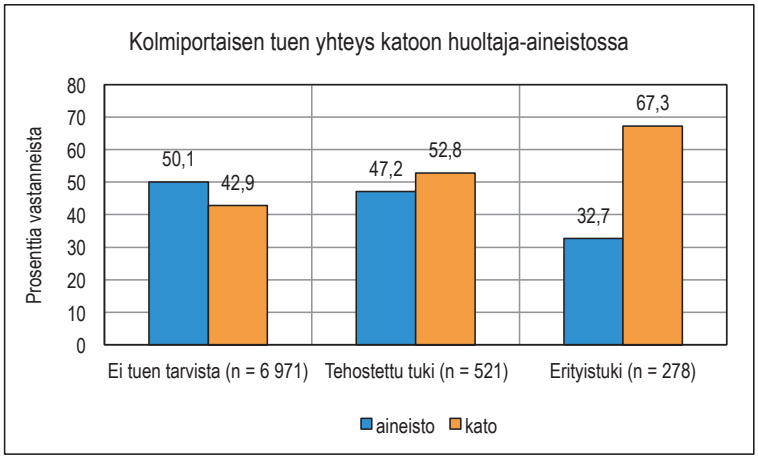
kokonaisosaamisen viiteen ryhmään, joiden välillä ero maksimoituu (Kuva 13). Kun parhaassa osaamisryhmässä (> 579 pistettä) vastaamatta jääneitä oli 32 %, heikommassa osaamisryhmässä (< 415) oppilaista 58 % jäi ilman huoltajien vastauksia.

Asia tarkentuu, kun huomioidaan oppilaiden koulukieli ja suomi tai ruotsi toisena kielenä suomi tai ruotsi toisena kielenä (S2/SV2) -status. Paremmissa osaamisryhmissä (> 440 pistettä) vastaamatta jättäminen oli merkitsevästi todennäköisempää, jos oppilas kävi ruotsinkielistä koulua kuin suomenkielistä koulua. Jos osaaminen oli hyvää (510–579), ruotsinkielisissä kouluissa käyneistä perheistä 52 % jätti vastaamatta kyselyyn, kun suomenkielisessä ryhmässä vastaamatta jätti vain 38 % vanhemmista. Parhaassa osaamisryhmässä (>579) vastaavat luvut olivat 47 % ja 30 %. Heikoimmissa osaamisryhmissä (< 440) vastaamatta jättäminen oli sen sijaan todennäköisempää, jos lapsella oli S2/SV2-status: Erittäin heikosti (< 415) osanneiden S2/SV2-oppilaiden huoltajista 70 % ei vastannut kyselyyn.

Osaamisen kanssa samanaikaisesti – joskin pienemmällä selitysvoimalla – katoon on yhteydessä myös oppilaan saama kolmiportainen tuki. Erityistukea saaneiden ryhmässä katoaineistoon kuului 67% oppilaiden huoltajista ja tehostettua tukea saaneiden ryhmässä 53 % (Kuva 14). Erot ryhmien välillä ovat erittäin merkitseviä ($p < 0{,}001$) ja huomattavia.



KUVA 13. Oppilaan osaamisen yhteys katoon huoltaja-aineistossa



KUVA 14. Kolmiportaisen tuen yhteys katoon huoltaja-aineistossa

Kaikkiaan on siis syytä olla varovainen, kun huoltaja-aineiston tiedoilla selitetään osaamisen eroja. Erityisesti heikoimpien oppilaiden, erityisen tuen päätöksellä ja S2/SV2-statuksella koulunsa aloittaneiden oppilaiden osalta huoltaja-aineistoon liittyy epävarmuutta, vaikka vastanneitakin huoltajia on riittävästi analyysien tekemiseen.

5.3 Lopullinen aineisto ja sen ominaispiirteitä

Yhteenvetona aiemmasta tiedetään, että kokonaisaineistoon kuului 7 877 oppilasta; varsinaiseen otokseen 7 770 oppilasta ja erityisotokseen 107 oppilasta, joista 65 erityiskouluista ja joista 42:lla oli yksilöllistetty matematiikan tai äidinkielen oppimäärä (HOJKS). Huoltajakyselyyn, josta saatiin valtaosa oppilaiden taustatiedoista, vastasi 56 % (n = 4 316) perusotokseen kuuluvien lasten huoltajista. Kaikilta huoltajilta ei kuitenkaan ole vastauksia kaikkiin kysymyksiin, ja näin muuttujakohtaiset otoskoot saattavat poiketa tästä hieman. Taulukkoon 9 on koottu aineistojen keskeisiä jakaumatietoja. Osa tiedoista tulee huoltaja-aineistosta, ja nämä pidetään erillään kokonaisaineistosta.

TAULUKKO 9. Valittuja aineistoa kuvaavia muuttujia ja niiden jakaumatietoja

Muuttuja		Koko aineisto (*n*)	Koko aineisto *n* = 7 770 (%)	Huoltajakyselyyn vastanneet *n* = 4 316 (%)	Erityisaineistot (erityiskoulut + HOJKS) *n* = 107 (%)
Sukupuoli	tyttö	3 875	49,9	49,9	29,9
	poikä	3 895	50,1	50,1	70,1
Oppilaitoksen kieli	suomi	6 902	88,8	91,1	89,7
	ruotsi	868	11,2	8,9	10,3
Oppimäärä	suomi	6 405	82,4		
	ruotsi	834	10,7		
	suomi/ruotsi toisena kielenä	497 + 34	6,8		
Suuralue	Helsinki-Uusimaa	2 409	31	30,2	20,6
	Etelä-Suomi	1 382	17,8	17,8	26,2
	Länsi-Suomi	1 749	22,5	23,5	44,9
	Pohjois- ja Itä-Suomi	2 230	28,7	28,5	8,4
AVI (aluehallintovirasto)	Etelä-Suomen AVI	3 015	38,8	38,2	32,7
	Lounais-Suomen AVI	917	11,8	11,8	48,6
	Itä-Suomen AVI	732	9,4	8,7	10,3
	Länsi- ja Sisä-Suomen AVI	1 672	21,5	22,2	4,7
	Pohjois-Suomen AVI	780	10	10,3	3,7
Kuntaryhmä	kaupunki	5 468	70,4	70,5	76,6
	taajama	1 184	15,2	15,7	7,5
	maaseutu	1 118	14,4	13,8	15,9
Kotikieli	suomi	3 502		81,3	
	ruotsi	183		4,2	
	jokin muu	148		3,4	
	suomi ja ruotsi	195		4,5	
	suomi ja muu	249		5,8	
	ruotsi ja muu	14		0,3	
	suomi, ruotsi ja muu	15		0,3	
S2/SV2-status	Ei	7 239	93,2	95,1	85
	Kyllä	531	6,8	4,9	15
Kolmiportainen tuki	Ei tuen päätöstä	6 971	89,7	92,2	0,0
	Tehostettu tuki	521	6,7	5,7	0,0
	Erityinen tuki	278	3,6	2,1	59,8
	HOJKS	0	0	(0,4)	40,2
Lähisuvussa oppimisvaikeuksia	ei oppimisvaikeuksia	3 125		72,4	
	yhdenlaista oppimisvaikeutta	720		16,7	
	useampia kuin yhdenlaista	471		10,9	

Muuttuja		Koko aineisto (*n*)	Koko aineisto *n* = 7 770 (%)	Huoltajakyselyyn vastanneet *n* = 4 316 (%)	Erityisaineistot (erityiskoulut + HOJKS) *n* = 107 (%)
Lapsen esiopetuspaikka toimi	koulun yhteydessä	1654		38,5	
	päiväkodin yhteydessä	2540		59,1	
	muualla	106		2,5	
Onko lapsi hänelle osoitetussa lähikoulussa?	Kyllä	4032		93,4	
	Ei	284		6,6	
Perheen korkein tutkinto	peruskoulu	48		1,1	
	ammatillinen koulutus	927		21,5	
	ylioppilastutkinto	343		8,0	
	ammattikorkeakoulu	1346		31,2	
	yliopisto	1535		35,6	
	jokin muu	111		2,6	
Osaamisen taso	kokonaisosaaminen	500		516	406
	matematiikka	500		514	403
	äidinkieli	500		515	407

Sukupuolen osalta huomataan, että erityisaineistot ovat selvästi sukupuolittuneet: 70 % erityisaineistoihin kuuluneista on poikia. Samoin alueellisen jakautumisen suhteen tämä ryhmä poikkeaa perusjoukosta: yli 40 % näistä oppilaista tulee Länsi-Suomen suuralueelta ja tarkemmin Lounais-Suomen AVI-alueelta, kun koko aineistossa oppilaita näillä alueilla on 22 % ja 12 %. Tältä alueelta oli kaksi erityiskoulua mukana otoksessa.

Aineistossa on selkeä yliedustus ruotsinkielisten koulutuksen järjestäjien opiskelijoita (11 %); 6 prosentin edustus oli ollut lähempänä demografista edustavuutta. Tämä johtuu siitä, että kansallisissa arvioinneissa on perinteisesti yliotostettu ruotsinkielisten koulujen oppilaita uskottavien tulosten raportoimiseksi myös tässä ryhmässä. Kotikielen suhteen asia on monimutkaisempi kuin oppilaitoksen opetuskielen suhteen. Puhtaasti ruotsinkielisiä opiskelijoita on huoltaja-aineiston mukaan noin 4 prosenttia, sillä 4,5 prosenttia oppilaista osoittautui kotikieleltään sekä suomen- että ruotsinkielisiksi eli kaksikielisiksi. Muistetaan myös, että ruotsinkieliset vanhemmat eivät olleet yhtä innokkaita vastaamaan huoltajakyselyyn kuin suomen- ja muunkielisten oppilaiden huoltajat (ks. luku 5.2). Vajaalla 10 prosentilla opiskelijoista on jokin muu kuin suomen kieli kotikielenään – useimmiten suomen kielen rinnalla (6 %). Suomi tai ruotsi toisena kielenä (S2/SV2) -statuksen saaneita oppilaita oli 7 %.

Huoltajien koulutustaustaa on OPH:n ja Karvin oppimistulosarvioinneissa vuodesta 2011 lähtien kartoitettu yksinkertaisella tiedolla siitä, ovatko vanhemmat ylioppilaita vai eivät (Kuusela, 2011). Huoltajien ylioppilastausta on selittänyt selvästi osaamisen eroja (esimerkiksi Metsämuuronen,

2013b; 2017a).[8] Tässä aineistossa huoltajien koulutustaustaa kysyttiin kuitenkin tarkemmin kuin aiemmissa mittauksissa. Lapsen yhden tai kahden huoltajan – useimmin äidin ja isän – korkeinta koulutusta kysyttiin kuuden portaan tarkkuudella: peruskoulu, ammatillinen koulutus, ylioppilastutkinto, ammattikorkeakoulu, yliopisto ja jokin muu. Huoltaja-aineiston mukaan huoltajien koulutusjakaumat poikkeavat toisistaan selvästi erityisesti ammatillisen koulutuksen suhteen. Äideistä 27 % oli ammattikoulutuksen saaneita ja isistä 41 %. Vastaavasti äideistä 30 % oli ammattikorkeakoulutaustaisia, kun isistä heitä oli 24 %. Ammatillinen koulutus oli kuitenkin melko harvoin perheen korkein koulutus (21 %). Useammin korkein koulutus vastaajaperheissä oli ammattikorkeakoulutus (31 %) tai yliopistokoulutus (36 %).

8 ks. myös esimerkiksi Hildén & Rautopuro (2014, s. 80), Härmälä & Huhtanen (2014, s. 197) ja Härmälä, Huhtanen & Puukko (2014, s. 80), Kärnä, Hakonen & Kuusela (2012, ss. 142–144), Ouakrim-Soivio & Kuusela (2012; ss. 116–124), Summanen, (2014, ss. 107–110) ja Venäläinen (2014, ss. 138–142).

6

Yhteenveto

6

Kaikkiaan koulun alkuvaiheen osaamista kartoittava tiedonkeruu on validiteetiltaan ja reliabiliteetiltaan riittävä luotettavien johtopäätösten tekemiseen kansallisella tasolla. Oppilasaineistossa kadolla ei ole juuri merkitystä, vaan tuloksen voidaan painotusten jälkeen suoraan yleistää koskemaan koko populaatiota. Sen sijaan varovaisuutta on noudatettava, kun kuvataan huoltajaaineiston perusteella oppilaiden taustatietojen yhteyttä osaamiseen. Vaikka tietoa on runsaasti kaikista osaamisryhmistä, huoltajakyselyn vastaajat edustivat selvästi hieman keskimääräistä paremmin suoriutuneita oppilaita. Aineisto luo hyvän pohjan seurata kuinka varsin eritasoisella osaamisella kouluun tulleet lapset kehittyvät tulevien vuosien aikana.

7

Viitteet

7

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*. New York: Pearson.

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*. New York: David McKay Company.

Bryk, A. S., & Raudenbush, S. W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin*, *101*(1), 147–158. DOI: http://dx.doi.org/10.1037/0033-2909.101.1.147.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. 2nd edition. Hillsdale, NJ: Erlbaum.

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

Goldstein, H. (1986). Multilevel mixed linear model analysis using iterative generalised least squares. *Biometrika*, *73*(1), 43–56. DOI: http://dx.doi.org/10.1093/biomet/73.1.43.

Gulliksen, H. (1950). *Theory of mental tests*. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Hildén, R., & Rautopuro, J. (2014). *Ruotsin kielen A-oppimäärän oppimistulokset perusopetuksen päättövaiheessa 2013*. Julkaisut 2014:1. Tampere: Kansallinen koulutuksen arviointikeskus.

Härmälä, M., & Huhtanen, M. (2014). *Ranskan kielen A- ja B-oppimäärän oppimistulokset perusopetuksen päättövaiheessa 2013*. Julkaisut 2014:3. Helsinki: Kansallinen koulutuksen arviointikeskus.

Härmälä, M., Huhtanen, M., & Puukko, M. (2014). *Englannin kielen A-oppimäärän oppimistulokset perusopetuksen päättövaiheessa 2013*. Julkaisut 2014:2. Tampere: Kansallinen koulutuksen arviointikeskus.

Jonckheere, A. R. (1954). A distribution-free k–sample test against ordered alternatives. *Biometrika, 41*(1–2), 133–145. DOI: http://dx.doi.org/10.1093/biomet/41.1-2.133.

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

Kuder, G. F., & Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika*, *2*(3), 151–160. DOI: http://dx.doi.org/10.1007/BF02288391.

Kärnä, P., Hakonen, R., & Kuusela, J. (2012). *Luonnontieteellinen osaaminen perusopetuksen 9. luokalla 2011*. Koulutuksen seurantaraportit 2012:2. Helsinki: Opetushallitus.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, Mass: Addison –Wesley Publishing Company.

Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics*, *18*(1), 50–60. DOI: http://dx.doi.org/10.1214/aoms/1177730491.

Metsämuuronen, J. (2008). *Monitasomallituksen perusteet*. Metodologia-sarja 11. International Methelp Ky. Jyväskylä: Gummeruksen kirjapaino Oy.

Metsämuuronen, J. (2009a). *Metodit arvioinnin apuna*. Oppimistulosten arviointi 1/2009. Opetushallitus. Helsinki: Yliopistopaino.

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

Metsämuuronen, J. (2010a). Pitkittäisaineistoon liittyviä menetelmäratkaisuja. Teoksessa E. K. Niemi, & J. Metsämuuronen (toim.), *Miten matematiikan taidot kehittyvät? Matematiikan oppimistulokset peruskoulun viidennen vuosiluokan jälkeen vuonna 2008*. (ss. 71–92.) Koulutuksen seurantaraportit 2010:2. Helsinki: Opetushallitus.

Metsämuuronen, J. (2010b). Osaamisen ja asenteiden muutos perusopetuksen 3. – 5. luokilla. Teoksessa E. K. Niemi & J. Metsämuuronen (toim.), *Miten matematiikan taidot kehittyvät? Matematiikan oppimistulokset peruskoulun viidennen vuosiluokan jälkeen vuonna 2008*. (ss. 93–136.) Koulutuksen seurantaraportit 2010:2. Helsinki: Opetushallitus.

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). Pitkittäisaineistoon liittyviä menetelmäratkaisuja. Teoksessa J. Metsämuuronen (toim.), *Perusopetuksen matematiikan oppimistulosten pitkittäisarviointi vuosina 2005–2012*. (ss. 31–64.) Koulutuksen seurantaraportit 2013:4. Helsinki: Opetushallitus.

Metsämuuronen, J. (2013b). Matemaattisen osaamisen muutos perusopetuksen luokilla 3–9. Teoksessa J. Metsämuuronen (toim.), *Perusopetuksen matematiikan oppimistulosten pitkittäisarviointi vuosina 2005–2012*. (ss. 65–172.) Koulutuksen seurantaraportit 2013:4. Helsinki: Opetushallitus.

Metsämuuronen, J. (2016). Item–total correlation as the cause for the underestimation of the alpha estimate for the reliability of the scale. *GJRA – Global Journal for Research Analysis*, *5*(1), 471–477. Osoitteessa https://www.worldwidejournals.com/global-journal-for-research-analysis-GJRA/file.php?val=November_2016_1478701072__159.pdf (Luettu 20.8.2019).

Metsämuuronen, J. (2017a). *Oppia ikä kaikki – Matemaattinen osaaminen toisen asteen koulutuksen lopussa 2015*. Kansallinen koulutuksen arviointikeskus. Julkaisuja 1:2017. Tampere: Karvi. Osoitteessa https://karvi.fi/app/uploads/2017/03/KARVI_0117.pdf (Luettu 25.3.2019).

Metsämuuronen, J. (2017b). *Essentials of research methods in human sciences. Vol 2: Multivariate methods*. New Delhi, India: SAGE Publications.

Metsämuuronen, J. (2017c). *Essentials of research methods in human sciences. Vol 3: Advanced Analysis*. New Delhi, India: SAGE Publications.

Metsämuuronen, J. (2018a). Seeking the true reliability – The mechanics of underestimation in some known estimators of reliability. Preprint. DOI: http://dx.doi.org/10.13140/RG.2.2.34100.09603/2.

Metsämuuronen, J. (2018b). Seeking the true reliability II – AlphaU, a new estimator of reliability to correct the underestimation in coefficient alpha and maximal reliability. Preprint. DOI: http://dx.doi.org/10.13140/RG.2.2.23964.95369/1.

Metsämuuronen, J. (2018c). New directional coefficient of correlation U_{gX} for the item analysis and measurement modelling settings. Preprint. DOI: http://dx.doi.org/10.13140/RG.2.2.11142.83529.

Metsämuuronen, J. (2018d). Essentials of visual diagnosis of test items – Logical and pathological patterns in items to be detected. Preprint. DOI: http://dx.doi.org/10.13140/RG.2.2.13950.23364.

Metsämuuronen, J. (2019a). Proportion of logically ordered cases as an index to assess the discrimination power of the score. Preprint. DOI: http://dx.doi.org/10.13140/RG.2.2.23350.75849.

Metsämuuronen, J. (2019b). Rank polyserial correlation for the measurement modelling settings. Preprint. DOI: http://dx.doi.org/10.13140/RG.2.2.35865.67682.

Metsämuuronen, J. (2019c). Dimension-corrected rank polyserial correlation for the measurement modelling settings. Preprint. DOI: http://dx.doi.org/10.13140/RG.2.2.31739.36644.

Metsämuuronen, J., & Räsänen, P. (2018). Opettaja mittaajana. Teoksessa J. Joutsenlahti, H. Silfverberg, & P. Räsänen (toim.), *Matematiikan opetus ja oppiminen* (ss. 321–343). Jyväskylä: Niilo Mäki Instituutti.

Metsämuuronen, J., & Salonen, V. (2017). *Matematiikan osaamisen piirteitä ammatillisessa koulutuksessa 2015 ja pitkän ajan muutoksia.* Kansallinen koulutuksen arviointikeskus. Publications 2:2017. Helsinki: Kansallinen koulutuksen arviointikeskus. Osoitteessa https://karvi.fi/app/uploads/2017/03/KARVI_0217.pdf (Luettu 20.8.2019).

OECD (2014). *PISA 2012 Results: What students know and can do* (Volume I, Revised edition, February 2014). Student Performance in Mathematics, Reading and Science. OECD Publishing. http://dx.doi.org/10.1787/9789264208780-en.

OECD (2016). *PISA 2015 Results (Volume I). Excellence and equity in education.* OECD Publishing. DOI: http://dx.doi.org/10.1787/9789264266490-en.

OPH (2014). *Perusopetuksen opetussuunnitelman perusteet 2014.* Määräykset ja ohjeet 2014:96. Next Print Oy, Helsinki: Opetushallitus.

OPH (2016). *Alkuopetuksen opetussuunnitelman perusteet 2014.* Määräykset ja ohjeet 2016:1. Helsinki: Opetushallitus.

Ouakrim-Soivio, N., & Kuusela. J, (2012). *Historian ja yhteiskuntaopin oppimistulokset perusopetuksen päättövaiheessa 2011.* Koulutuksen seurantaraportit 2012:3. Helsinki: Opetushallitus.

Pearson, K. (1903). I. Mathematical contributions to the theory of evolution. – XI. On the influence of natural selection on the variability and correlation of organs. *Philosophical Transactions of the Royal Society A. Mathematical, Physical and Engineering Sciences, 200*(321–330), 1–66. DOI: https://doi.org/10.1098/rsta.1903.0001.

Pearson, K. (1905). *On the general theory of skew correlation and non-linear regression.* London. Dulau & Co. Osoitteessa https://archive.org/details/ ongeneraltheory00peargoog/page/n3 (Luettu 20.8.2019).

Pierce, C. A., Block, R. A., & Aguinis, H. (2004). Cautionary note on reporting eta-squared values from multifactor ANOVA designs. *Educational and Psychological Measurement*, *64*(6), 916–924. DOI: http://dx.doi.org/10.1177/0013164404264848.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Application and data analysis methods.* 2nd edition. Advanced Quantitative Techniques in the Social Sciences Series. Thousands Oaks: Sage Publications.

Šidák, Z. (1967). Rectangular confidence region for the means of multivariate normal distributions. *Journal of the American Statistical Association*, *62*(318), 626–633. DOI: http://dx.doi.org/10.1080/01621459.1967.10482935.

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

Summanen, A.-M. (2014). *Terveystiedon oppimistulokset perusopetuksen päättövaiheessa 2013.* Koulutuksen seurantaraportit 2014:1. Helsinki: Opetushallitus.

Tabachnick, B. G., & Fidell, L. S. (2006). *Using multivariate statistics.* 5th edition. Boston: Allyn and Bacon.

Terpstra, T. J. (1952). The asymptotic normality and consistency of Kendall's test against trend, when ties are present in one ranking. *Indagationes Mathematicae, 14*(3), 327–333. DOI: http://dx.doi.org/10.1016/S1385-7258(52)50043-X.

TIMSS (2012). Mullis, I. V. S., Martin, M. O., Foy, P., & Arora, A. (2012). *TIMSS 2011 International results in mathematics*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. Osoitteessa https://timssandpirls.bc.edu/isc/publications.html (Luettu 20.8.2019).

TIMSS (2016). Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2016). *TIMSS 2015 International results in mathematics*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. Osoitteessa https://timssandpirls.bc.edu/isc/publications.html (Luettu 20.8.2019).

Tukey, J. W. (1953). The problem of multiple comparisons. Teoksessa H. I. Braun (toim.), *The collected works of John W. Tukey VIII* (Multiple Comparisons: 1948–1983 1–300). New York, NY: Chapman and Hall.

Verhelst, N. G., Glas, C. A. W., & Verstralen, H. H. F. M. (1995). *One-Parameter Logistic Model OPLM*. Arnhem: Cito.

Venäläinen, S. (2015). *Arjen tiedot ja taidot hyvinvoinnin perustana. Kotitalouden oppimistulokset perusopetuksen päättövaiheessa 2014*. Julkaisut 2015:5. Helsinki: Kansallinen koulutuksen arviointikeskus.

Wherry, R. J. Sr. (1931). A new formula for predicting the shrinkage of the coefficient of multiple correlation. *Annals of Mathematical Statistics, 2*, 440–457.

Åkerlund, C., Marjanen, J., & Lepola, L. (2019) *Lärarresultat i Finska i åk 6. Resultat av en utvärdering våren 2018*. Publikationer 6:2019. Helsinki: Kansallinen koulutuksen arviointikeskus.

Kansallinen koulutuksen arviointikeskus (Karvi) on itsenäinen koulutuksen arviointivirasto. 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.

Raportissa kuvataan lukuvuoden alussa syksyllä 2018 suoritetun ensimmäisen luokan oppilaiden osaamisen arvioinnin menetelmällisiä ratkaisuja. Arvioinnissa mitattiin oppilaiden osaamista matematiikassa ja äidinkielessä ja kirjallisuudessa. Raportti kuvaa mittareiden validiteetti- ja reliabiliteettitarkastelut, osioanalyysin, otoksen, menetelmävalinnat ja aineistojen perustunnusluvut.

ISBN 978-952-206-546-9 nid.
ISBN 978-952-206-547-6 pdf

ISSN 2342-4176 (painettu)
ISSN 2342-4184 (verkkojulkaisu)
ISSN 2342-4176

9 789522 065469

Kansallinen
koulutuksen arviointikeskus
PL 28 (Mannerheiminaukio 1 A)
00101 HELSINKI
Puhelinvaihde: 029 533 5500
Faksi: 029 533 5501

karvi.fi