



TAMPEREEN TEKNILLINEN YLIOPISTO
TAMPERE UNIVERSITY OF TECHNOLOGY

ALEKSI ROIMA
TAPAUSTUTKIMUS DATA-ANALYTIIKAN ESIIN TUOMISTA PIIR-
TEISTÄ OPPIMISEN ARVIOINNISSA

Kandidaatintutkielma

Tarkastaja: lehtori Pasi Hellsten

16.8.2017

TIIVISTELMÄ

Alexi Roima: Tapaustutkimus data-analytiikan esiin tuomista piirteistä oppimisen arvioinnissa

Tampereen teknillinen yliopisto

Kandidaatintyö, 30 sivua, 7 liitesivua

Elokuu 2017

Tietojohdamisen kandidaatin tutkinto-ohjelma

Pääaine: Tietojohdaminen

Tarkastaja: lehtori Pasi Hellsten

Avainsanat: Data-analytiikka, oppimisen arviointi, itsearviointi, korrelaatiotutkimus

Tässä kandidaatintyössä tutkitaan data-analytiikan esiin tuomia piirteitä oppilaan itsearviointeihin liittyen digitaalisesta arviointiohjelmistosta kerätyssä aineistossa. Tutkimus on rajattu aineiston osalta koskemaan suomalaista peruskoulua ja metodologian osalta korrelaatioiden tutkimista ja visualisointia. Korrelaatioiden tutkimisen lisäksi, työssä avataan aineiston käsittelyä ja tutkimuksen teknistä toteutusta. Tutkimusta varten tehdyssä selvitystyössä ei löydetty vastaavaa aiempaa suomalaista tutkimusta, joten taustateoria ja testattavat hypoteesit luodaan kansainvälisten tutkimustulosten perusteella. Teorian pohjalta muodostetaan neljä oppimisen itsearviointiin liittyvää hypoteesia, joiden paikkansapitävyyttä testataan suomalaisista peruskouluista kerätyssä data-aineistossa.

Ensimmäisen hypoteesin mukaan itse- ja opettaja-arvioinnit korreloivat keskenään positiivisesti. Hypoteesia testattiin Spearmanin järjestyskorrelaatiokertoimella, jonka laskennalliseksi arvoksi saatiin 0,828, joka viittaa voimakkaaseen positiiviseen korrelaatioon eli korrelaatiokerroin tuki hypoteesia. Toinen hypoteesi väitti arviointimäärien parantavan oppilaan itsearviointikykyä. Väitteelle ei kuitenkaan löytynyt aineistosta tukea, sillä arviointimäärien mukaan luokitellut oppilasryhmäkohtaiset korrelaatiokertoimet eivät olleet monotonisessa järjestyksessä. Kolmannen hypoteesin mukaan korkea vuosiluokka parantaa itsearviointikykyä. Vuosiluokan ja itsearviointikyvyn korrelaatiokertoimien väliltä löydettiin lievää positiivista trendiä, joka tuki hypoteesia. Neljäs hypoteesi esitti paremmin suoriutuvien oppilaiden aliarvioivan suoriutumistaan ja heikommin suoriutuvien yliarvioivan suoriutumistaan itsearvioinnissa. Hypoteesi osoittautui paikkansapitäväksi, sillä 53 % heikommin suoriutuneista oppilaista yliarvioi ja samaan aikaan 16 % paremmin suoriutuneista aliarvioi suoriutumistaan itsearvioinneissaan.

Saavutetut tulokset olivat tilastollisesti erittäin merkitseviä, mutta siitä huolimatta, tutkimuksen tuloksissa on otettava huomioon esimerkiksi kapea otanta kahdesta suomalaisesta peruskoulusta, oppilaan ja opettajan mahdollisuus nähdä toistensa antamat arvioinnit sekä opettaja-arvioinnin luotettavuus puolueettomana vertailukohteena. Saavutetut tulokset luovat lähtökohdan oppilaan arvosanoja ennustavien mallien luomiselle, ja niiden pohjalta voidaan tunnistaa tutkimuksen tuloksia syventäviä ja tarkentavia jatkotutkimuskohteita.

ALKUSANAT

Tämä kandidaatintyö on tehty opinnäytteenä tietojohdamisen koulutusohjelman kandidaattitutkintoa varten. Työni aiheen halusin liittää kiinnostuksen kohteenani olevaan data-analytiikkaan ja tähän sainkin erinomaisen mahdollisuuden DEEVA-tutkimushankkeen kautta. Mahdollisuuden tarjoamisesta kiitoksen ansaitsee associate professor Nina Helander ja tuesta työn alkuvaiheessa tutkijatohtori Heli Väättäjä. Kandidaatintyöni lopullinen aihe valikoitui yhteistyössä SkillzzUp Oy:n Kimmo Kumpulaisen sekä Jani Mikkolan kanssa. Haluan esittää kiitokseni heille kommenttien antamisesta työnteon aikana ja mahdollisuudesta suorittaa kandidaatintyö yritysyhteistyössä.

Suuret kiitokset ansaitsee työni ohjaaja lehtori Pasi Hellsten erinomaisista kommentteistaan ja ajatuksien herättämisestä tutkimukseen liittyen. Haluan kiittää myös tietojohdamisen kandidaatintyöryhmää kommentteista ja rakentavasta palautteesta työn aikana sekä Tampereen teknillisen yliopiston Novi-tutkimusryhmää saaduista neuvoista ja ohjeista. Kiitokset tutkimusaihetta koskeneista asiantuntevista kommentteista työhön liittyen ansaitsee Kansallisen koulutuksen arviointikeskuksen projektipäällikkö Laura Pihala. Lopuksi haluan vielä esittää kiitokset puolisolleni Annika Lindellille tuesta ja kieliopillisista ohjeista työn viimeistelyssä.

Tampereella 16.8.2017

Aleksi Roima

SISÄLLYSLUETTELO

1.	JOHDANTO	1
1.1	Tutkimuksen taustat ja merkitys	2
1.2	Tutkimusongelma ja rajaus	2
1.3	Tutkimuksen rakenne	3
2.	TUTKIMUSMENETELMÄ JA -AINEISTO.....	4
2.1	Tutkimusmenetelmän kuvaus.....	4
2.2	Korrelaation tutkiminen tilastollisilla menetelmillä.....	4
2.2.1	Korrelaatiokertoimet	5
2.3	Tutkimusaineiston esittely.....	7
2.3.1	Aineiston käsittely.....	9
3.	TEOREETTINEN TAUSTA	10
3.1	Itsearviointi.....	10
3.2	Data-analytiikka opetuksessa	11
3.3	EDM:n hyödyntäminen oppimisen arvioinnissa	12
4.	KORRELAATIOT DIGITAALISESSA ARVIOINTIAINEISTOSSA	15
4.1	Oppilaan itsearviointikyky	15
4.1.1	Itse- ja opettaja-arviointien korrelaatio	15
4.1.2	Arviointien määrän vaikutus itsearviointikykyyn.....	16
4.1.3	Vuosiluokan vaikutus itsearviointikykyyn.....	17
4.2	Opiskelumenesytksen vaikutus itsearviointiin	18
5.	TULOSTEN ANALYSOINTI.....	20
5.1	Tutkimuksen tulokset	20
5.2	Tutkimuksen arviointi	22
6.	PÄÄTELMÄT	24
6.1	Tulosten yhteenveto	24
6.2	Jatkotutkimus ja rajoitteet	25
	LÄHTEET.....	26

LIITE A: datan_putsaus.py

LIITE B: yleinen_korrelaatio.py

LIITE C: arviointimaarien_vaikutus_korrelaatioon.py

LIITE D: vuosiluokan_vaikutus_korrelaatioon.py

LIITE E: suoriutumisen_vaikutus_itsearviointiin.py

LYHENTEET JA KESKEISET KÄSITTEET

Arviointi	Kasvatuksellisesti prosessien, edellytysten ja tuloksena syntyvän arvon tai ansion määrittelemistä. Arviointi luodaan vertaamalla tuloksia ja sitä edistäneitä prosesseja ennalta asetettuihin tavoitteisiin ja edellytyksiin. (Atjonen 2007, s. 19)
Arviointiohjelmisto	Opetuksessa käytettävä digitaalinen arviointiväline, johon opettaja luo arviointikohteita. Oppijat tekevät arviointikohteisiin itsearviointeja ja opettajat antavat omia arviointejaan oppilaskohteisesti.
Data-analytiikka	Tietoteknisten välineiden avulla merkittävien mallien tunnistamista suurista data-massoista. Mallien tunnistamisella voidaan ymmärtää käsiteltävää asiaa paremmin ja kehittää organisatorista päätöksentekoa. (Chandler & Rod 2016)
EDM	(engl. <i>Educational data mining</i>) Tietoteknisten metodien joukkoa hyödyntävä tutkimussuunta, jossa tutkitaan opetukseen ja oppimiseen liittyvää dataa, jonka avulla oppijoita ja oppimisympäristöjä voidaan ymmärtää paremmin (Baker & Yacef 2009).
Itsearviointi	Tavoitteiden ymmärtämisestä, suorituksen tekemiseen ja oman suoriutumisen arviointiin etenevä prosessi, jonka tavoitteena on kehittää oppijan kykyä oppia ja löytää omia vahvuuksia sekä heikkouksia.
Korrelaatiokerroin	Muuttujien välistä riippuvuutta kuvaava lukuarvo, joka voi saada lukuarvoja välillä [-1; 1]. Korrelaatiokerroin lähellä arvoa 0 tarkoittaa ettei riippuvuutta ole ja lähellä asteikon päätepisteitä oleva arvo viittaa voimakkaaseen positiiviseen tai negatiiviseen riippuvuuteen. (Heikkilä 2005, ss. 90–91)
LA	(engl. <i>Learning analytics</i>) Oppijoiden ja oppimisen kontekstiin liittyvän datan keräämiseen, mittaamiseen, analysoimiseen ja raportoimiseen keskittyvä tutkimussuunta, joka pyrkii ymmärtämään ja kehittämään oppimista (Siemens 2012).
Opettaja-arviointi	Opettajan antama arviointi tietystä oppijasta tiettyyn arviointikohteeseen. Opettaja-arvioinnin tarkoitus on tasapuolisesti ja mahdollisimman tarkasti arvioida eri oppijoiden taitotasoa.

P-arvo

Todennäköisyyttä kuvaava lukuarvo sille, että nollassa nollahypoteesi voi lasketulle tulokselle käytetyssä otoskoossa pitää paikkansa. P-arvoa käytetään yleisesti kuvaamaan tilastollista merkitsevyyttä. P-arvo voi saada arvoja välillä $[0; 1]$ ja alhainen p-arvo tarkoittaa pientä riskiä tuloksen satunnaisuudesta. (Ruohonen 2011, ss. 34–35) Esimerkiksi korrelaatiokerrointa laskettaessa verrataan saatua kerrointa olettamukseen, ettei muuttujien välillä ole riippuvuutta.

Python

Avoimeen lähdekoodiin perustuva tulkattava ohjelmointikieli (Python Software Foundation 2017a). Käytetään yleisesti verkkosovellusten ohjelmoinnissa sekä datan käsittelyssä.

1. JOHDANTO

Suomen valtioneuvosto on vuonna 2015 määrittänyt toimintasuunnitelmassaan osaamiseen ja koulutukseen liittyväksi ensimmäiseksi kärkihankkeekseen ”Uudet oppimisympäristöt ja digitaaliset materiaalit peruskouluihin”, jonka tavoitteena on nostaa Suomi innostavan ja modernin oppimisen kärkimaaksi (Valtioneuvoston kanslia 2016). Tavoite on asetettu korkealle ja sen saavuttaminen vaatii merkittäviä investointeja digitaalisiin oppimisoloihin ja näitä tukeviin työkaluihin. Peruskoulun digitalisaatio ei ole tähän mennessä sujunut ongelmitta, vaan mediassakin esillä olleita haasteita ovat aiheuttaneet niin opettajien muutosvastarinta, teknologisten laitteiden puuttuminen, kuin myös ohjelmistojen soveltumattomuus opetuskäyttöön (Laitinen 2015; Nironen 2016; Saari & Sääntti 2016; Salumäki 2016). Digitaaliset oppimisvälineet ja -ohjelmistot ovatkin saaneet viime vuosina melko negatiivista huomiota mediassa, vaikka niiden oikeaoppisella hyödyntämisellä voidaan edesauttaa yksilöllistä oppimista (e.g. Warren et al. 2008; McClarty et al. 2012; Hung et al. 2014). Aiheeseen liittyvä negatiivisuus johtuu osittain siitä, ettei tutkimustietoa digitalisaation vaikutuksista suomalaisten peruskoululaisten oppimistuloksiin ole juurikaan tarjolla.

Yleisellä tasolla oppimistulosten arvioinnin tarkoituksena on kehittää koulutusta pysymään ajankohtaisena ja tasalaatuisena sekä varmistaa koulutettavien tasa-arvo. Oppimisen arviointi parantaa oppimisen edellytyksiä, tukee oppimista, opettajan työtä ja oppimisympäristön kehittymistä. (Räisänen 2013) Vaikka arviointia voikin nopeasti ajatella vain suoriutumisen ja oppimisen mittarina, on sillä merkittävä rooli myös arvioitavan motivaatioon. Newstead (2002) korostaakin, että oppilaiden motivaatio on niin riippuvainen oppilaan saamista arvioinneista, että todelliset oppimistulokset jäävät yleensä toissijaiseksi. Opettajien tekemä oppimisen arviointi vaikuttaakin usein negatiivisesti oppimistuloksiin. (Newstead 2002) Tämä herättääkin ihmettelemään, miksi opettajien antamalla arvioinneilla on edelleen niin keskeinen rooli koko suomalaisessa koulujärjestelmässä?

Arvioinnin vaikutukset oppilaan motivaatioon ja tätä kautta oppimistuloksiin eivät rajoitu kuitenkaan pelkästään opettajien antamiin arviointeihin. Paris et al. (1991) sekä Ross et al. (1999) huomauttavat tutkimuksissaan, että oppilaat kiinnittävät enemmän huomiota itsearviointeihin kuin muiden antamiin arviointeihin. Asia korostuu erityisesti negatiivisen palautteen kohdalla ja pidemmälle edenneillä oppilailta. (Paris et al. 1991; Ross et al. 1999) Opetushallituksen opetusneuvos Räisänen (2013) kertoo, että Suomen oppimistulosten arviointi on vakiinnuttanut asemansa melko muuttumattomaksi, vaikka kansainvälinen yhteistyö ja kerätyt kokemukset nostavatkin kehittämistarpeita esille. Vielä tois-taiseksi on epäselvää, miten arviointien siirtyminen sähköisiin ohjelmistoihin tulee muuttamaan tätä tilannetta.

1.1 Tutkimuksen taustat ja merkitys

Tässä kandidaatintyössä tutkitaan digitaalisen arviointiohjelmiston vaikutuksia peruskoulun oppilaiden itsearviointeihin. Tutkimuksen taustalla vaikuttaa vahvasti syksyllä 2016 voimaan astunut Perusopetuksen opetussuunnitelman perusteet 2014 (Opetushallitus 2014a). Uuden opetussuunnitelman voimaan astumisen jälkeen peruskoulu on merkittävässä murroksessa, jonka myötä oppilaille pyritään antamaan entistä ajankohtaisemmat tietotaidot kansainvälisestä näkökulmasta (Opetushallitus 2014b). Itsearviointiin rooliin opetussuunnitelmassa palataan luvussa 3.

Kandidaatintyö tehdään osana Tampereen teknillisen yliopiston, Tampereen ammattikorkeakoulun ja Turun ammattikorkeakoulun yhteistä tutkimushanke DEEVA:a. Tutkimushankkeessa selvitetään digitalisaation myötä syntyviä uusia arvionluontitapoja hyödyntäen dataa, tunteita ja kokemuksia (Jalonen et al. 2017). Kandidaatintyö tehdään yhteistyössä tamperelaisen ohjelmisto- ja valmennusyrityksen SkillzzUp Oy:n kanssa, jonka arviointiohjelmiston dataa tutkimuksessa hyödynnetään. SkillzzUp Oy tarjoaa pedagogisen täydennyskoulutuksen lisäksi tuotteenaan arviointiohjelmistoa, jolla voidaan tukea yksilöllistä oppilaan ohjausta arviointien kautta.

Tutkimusta voidaan pitää merkittävänä niin yhteiskunnallisesti kuin myös kohdeyrityksen kannalta. Vaikka oppimisen itsearviointiin liittyen onkin tehty runsaasti kansainvälistä tutkimusta (e.g. Dochy et al. 1999; Fitzgerald et al. 2003; Lew et al. 2010; Taras 2010; Hatami 2015), ei tehdyssä selvitystyössä noussut esille juurikaan aiheeseen liittyvää suomalaista tutkimusta. Suomen peruskoulun uniikin luonteen takia ei voida olla varmoja kansainvälisten tulosten sovellettavuudesta Suomen koulumaailmaan. Tämä lisää kandidaatintyön tutkimuksen merkittävyyttä uutuusarvolla. SkillzzUp Oy:lle tutkimus tarjoaa parempaa ymmärrystä arviointiohjelmistonsa vaikutuksista koulutukselle, esimerkiksi ohjelmiston käytön ja oppilaiden arviointitulosten välisistä yhteyksistä.

1.2 Tutkimusongelma ja rajaus

Tässä empiirisessä tutkimuksessa selvitetään data-analytiikan esiin tuomia piirteitä oppimisen arvioinnissa. Aihe itsessään on melko laaja, joten sitä rajataan sekä käytettävän aineiston että metodologian osalta. Aineiston osalta työssä perehdytään ainoastaan suomalaiseen peruskouluun eli vuosiluokkiin 1–9. Peruskoulun oppilaat ovat pääsääntöisesti iältään noin 6–16-vuotiaita oppilaita. Käytettävä aineisto pitää sisällään oppilaiden itsearviointeja sekä opettaja-arvioita kahdesta eri suomalaisesta peruskoulusta. Opettaja-arvioinneilla viitataan työssä opettajien tekemiin arviointeihin oppilaista.

Metodologian osalta tutkimus rajataan käsittelemään ainoastaan korrelaatioiden tutkimista. Tutkimusta ei rajata kuitenkaan pelkästään tilastollisen korrelaation laskemiseen aineistosta, vaan se pitää sisällään myös tulosten visualisoinnin ja analyysin. Kandidaatintyössä vastataan seuraavaan päätutkimuskysymykseen:

Miten data-analytiikkaa voidaan hyödyntää peruskoulun oppilaiden itsearviointien korrelaatioiden tutkimisessa arviointiohjelmistosta kerätystä aineistosta?

Tähän päätutkimuskysymykseen haetaan ratkaisua seuraavien alatutkimuskysymysten avulla:

- Millä tavoilla korrelaatiota voidaan tutkia data-aineistosta?
- Mitkä tekijät vaikuttavat oppilaan kykyyn arvioida omaa oppimistaan?
- Mitä korrelaatioita SkillzzUp Oy:n suomalaisista peruskouluista kerätystä datasta voidaan tunnistaa luotujen hypoteesien perusteella?

Tutkimuksen pääpaino on kirjallisuuden pohjalta tunnistettavien hypoteesien testaamisessa, mutta samalla halutaan tarjota aineiston käsittelyyn ja testien suorittamiseen vaa-dittava teknologinen tausta. Tutkimuksessa avataan sanallisesti keskeiset ohjelmakoo-dissa suoritettavat toimenpiteet, ja yksityiskohtaisemmat Python-koodit ovat nähtävillä työn liitteissä.

1.3 Tutkimuksen rakenne

Tutkimus jakautuu kuuteen lukuun. Johdannon jälkeisessä toisessa luvussa esitellään tutkimusmenetelmä, johon liittyy sekä tilastollisten menetelmien että käytettävien ohjelmis-tojen kuvaaminen. Toisessa luvussa esitellään myös empiirisessä tutkimuksessa käytet-tävä data-aineisto ja sen käsittely. Aineiston ja menetelmän esittely heti alkuvaiheessa on tämän tutkimuksen kannalta keskeistä, jotta lukijalle selviää, minkä tyyllisiä korrelaatioita teorian pohjalta pyritään tunnistamaan. Kolmannessa luvussa käsitellään oppilaan itsearvioinnin merkitystä oppimisessa, joka osaltaan perustelee tutkimuksen merkittävyyttä ja nostaa esille oppimiseen vaikuttavia tekijöitä, joiden esiintymistä data-aineistossa tutki-taan. Samassa luvussa käsitellään myös aihepiiriin liittyvää aiempaa tutkimusta eli, mitä data-analytiikan hyödyntämisestä oppimisen arvioinnissa on tutkittu.

Kolmannen luvun taustateorian pohjalta luodaan ja esitellään empiirisessä osiossa testat-tavat hypoteesit. Neljännessä luvussa selvitetään edeltävässä luvussa luotujen hypotee-sien paikkansapitävyyttä suomalaisista peruskouluista kerätystä data-aineistossa lasken-nallisten arvojen ja visualisointien pohjalta. Viidennessä luvussa esitellään aiemmassa luvussa lasketut tulokset, analysoidaan niiden merkitystä sekä arvioidaan tutkimukseen vaikuttaneita tekijöitä. Kuudennessa eli viimeisessä luvussa tiivistetään keskeiset tulokset ja esitetään jatkotutkimusehdotuksia.

2. TUTKIMUSMENETELMÄ JA -AINEISTO

Luvussa kuvataan tutkimuksen tekemiseen käytettävä tutkimusmenetelmä sekä aineiston tulkintaan käytettävät tilastolliset menetelmät, joilla korrelaatiota voidaan tutkia. Tämän lisäksi luvussa esitellään tutkimuksessa käytettävä tutkimusaineisto ja sen käsittelyssä tehdyt toimenpiteet.

2.1 Tutkimusmenetelmän kuvaus

Tutkimus toteutetaan kvantitatiivisena empiirisenä tutkimuksena, jossa selvitetään peruskoulun oppilaiden oppimisen arvioinnissa käytettävän digitaalisen arviointiohjelmiston datasta löytyviä riippuvuuksia eli korrelaatioita. Heikkilän (2005) mukaan kvantitatiivinen tutkimusprosessi pitää sisällään muun muassa seuraavat vaiheet: aikaisempiin tutkimuksiin ja kirjallisuuteen perehtyminen, mahdollisten hypoteesien laatiminen, tietojen käsitteleminen sekä analysoiminen, tulosten raportoiminen ja johtopäätökset. Samaa kvantitatiivisen tutkimuksen rakennetta noudatetaan myös tässä kandidaatintyössä.

Datan käsittely tapahtuu Python-ohjelmointikielellä pandas-kirjastoa hyödyntäen. Pandas on joustava strukturoidun datan tehokkaaseen käsittelyyn tarkoitettu analyyttikatyökaluja tarjoava ohjelmointikirjasto. Pandas-kirjasto tarjoaa työkaluja myös korrelaatiokertoimien tutkimiseen. Korrelaatiokertoimista valittavina matemaattisina metodeina ovat Pearson, Kendall ja Spearman (Python Software Foundation 2017b). Näistä keskeisimmät metodit esitellään työn luvussa 2.2.1. Hauke ja Kossowski (2011) esittävät tutkimuksessaan, että korkeat korrelaatiokertoimet eivät itsessään ole todiste kahden asian välisestä riippuvuudesta, vaikka korrelaatiokerrointa usein käytetäänkin riippuvuuden tieteelliseen perusteluun. Tämän takia löydetyistä korrelaatiokertoimista onkin syytä laskea tilastollista merkitsevyyttä kuvaava p-arvo ja visualisoida löydökset riippuvuuksien tulkitsemiseksi.

Aineiston tulokset visualisoidaan Pythonin matplotlib-kirjaston päälle rakennetun seaborn-kirjaston avulla. Matplotlib-kirjaston avulla voidaan visualisoida 2D-kuvia suoraan Pythonin ohjelmakoodin avulla (Hunter 2007). Seaborn puolestaan helpottaa informatiivisten visualisaatioiden tekemistä suoraan pandas-kirjaston datataulujen avulla (Waskom 2015). Tutkimuksessa hyödynnettävät työkalut on valittu paitsi niiden soveltuvuuden takia, myös ohjelmien avoimen lähdekoodin takia. Tämä mahdollistaa tutkimusmenetelmän toistamisen ilman kaupallisia ohjelmistoja.

2.2 Korrelaation tutkiminen tilastollisilla menetelmillä

Tutkittaessa asioiden välistä riippuvuussuhdetta eli korrelaatiota, tarkastellaan kahden tekijän tai muuttujan vaikutusta toisiinsa. Tavoitteena korrelaatioiden tutkimisessa on usein

löytää tilastollista riippuvuutta, jolloin toisen muuttujan arvoja on mahdollista hyödyntää toisen muuttujan arvojen ennustamisessa. (Mellin 2006, ss. 240–241) Muuttujien välisen korrelaation laskemiseen on tarjolla useita erilaisia korrelaatiokertoimia, jotka sopivat erityyppisten muuttujaparien vertailuun. Yleisimmin käytettyjä korrelaatiokertoimia ovat Pearsonin korrelaatiokerroin, Spearmanin järjestyskorrelaatiokerroin ja Kendallin järjestyskorrelaatiokerroin. (Heikkilä 2005, ss. 203–204) Spearmanin ja Kendallin järjestyskorrelaatiokertoimet eivät eroa merkittävästi toisistaan ja niitä voidaan käyttää samanlaisiin aineistoihin. (Colwell & Gillett 1982) Spearmanin järjestyskorrelaatiokerroin on kuitenkin näistä kahdesta yleisemmin käytetty yksinkertaisuutensa takia (Colwell & Gillett 1982) ja tämän takia kahden järjestyskorrelaatiokertoimen osalta tässä tutkimuksessa tutustutaan vain Spearmaniin.

2.2.1 Korrelaatiokertoimet

Vertailtaessa aineiston otoksen kahta muuttujaa, lasketaan muuttujien välinen korrelaatiokerroin yleensä hyödyntäen Pearsonin otoskorrelaatiokerrointa, joka saadaan kaavalla (Mellin 2006, ss. 249–250; Ruuhonen 2011, ss. 76–78; Xiao et al. 2016)

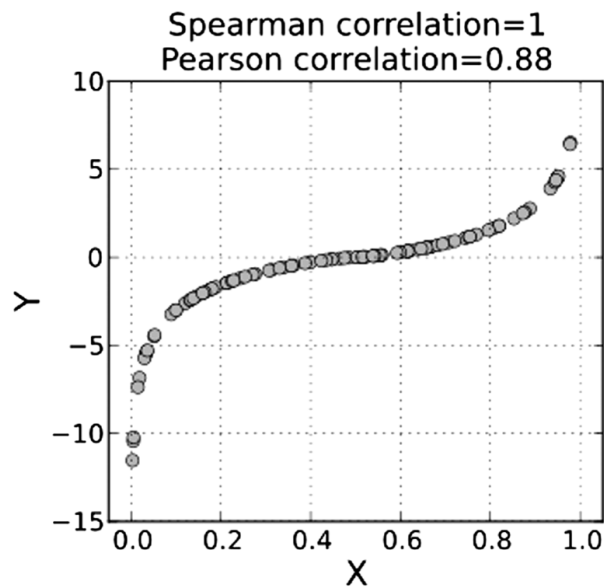
$$r_p = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}, \quad (1)$$

jossa n tarkoittaa lukuparien määrää, x_i muuttujan x järjestyksessä i :ttä muuttujaa, y_i muuttujan y järjestyksessä i :ttä muuttujaa, \bar{x} muuttujan x keskiarvoa ja \bar{y} muuttujan y keskiarvoa. Pearsonin otoskorrelaatiokerroin ei kuitenkaan sovellu järjestysasteikollisten muuttujien vertailuun. Xiao et al. (2016) huomauttavat Pearsonin otoskorrelaatiokertoimen heikkoutena olevan myös sen keskittymisen lineaariseen korrelaatioon, jolloin esimerkiksi potenssifunktioiden muotoon asettuvat havaintopisteet eivät osoita todellisuutta vastaavaa voimakasta korrelaatiota. Tämän tutkimuksen aineistossa on useita jaettuja sija-lukuja, eikä sen lineaarisuudesta voida olla varmoja, joten analyysissä on käytettävä jotain muuta korrelaatiokertoimen laskentakaavaa kuin Pearsonin.

Järjestysasteikollisille muuttujille, kuten tutkimuksemme arvosana-asteikolle, voidaan laskea korrelaatiokerroin hyödyntäen Spearmanin järjestyskorrelaatiokerrointa. Spearmanin järjestyskorrelaatiokerroin keskittyy muuttuja-arvojen suuruusjärjestyksen yhteensopivuuteen (Mellin 2006, ss. 262–264). Kyseessä on Pearsonin korrelaatiokertoimen erikoistapaus, jossa molemmat muuttujajoukot järjestetään suuruusjärjestykseen pienimmästä suurimpaan ja korvataan muuttujien arvot järjestyslukuilla (Xiao et al. 2016). Tämän jälkeen Spearmanin järjestyskorrelaatiokerroin saadaan laskettua kaavalla (Mellin 2006, ss. 262–264; Xiao et al. 2016)

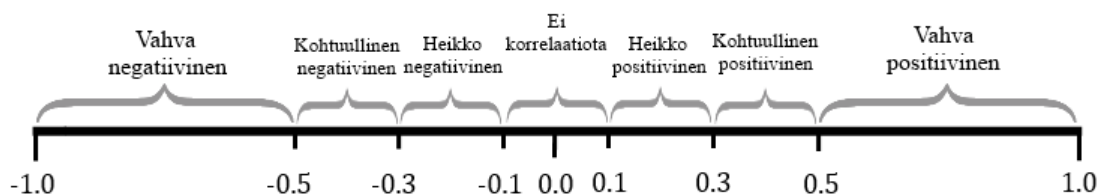
$$r_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)}, \quad (2)$$

jossa $d_i = X_i - Y_i$, eli muuttujaparin järjestysnumeroiden erotus, ja n tarkoittaa lukuparien määrää. Kuvassa 2.1 esitetään, mitä Pearsonin ja Spearmanin korrelaatiokerroimen käyttäminen korrelaation kuvaajana tarkoittaa epälineaarisisessa aineistossa. Kuvasta 2.1 voidaan nähdä, että tilanteessa Spearmanin korrelaatiokerroin saa arvon 1, eli se kuvaa täydellistä positiivista korrelaatiota, vaikka aineisto ei asetukkaan suoralle. Pearsonin korrelaatiokerroin puolestaan yrittää sovittaa aineistoon suoraa, jonka takia korrelaatiokerroin jää selkeästi alhaisemmaksi. (Xiao et al. 2016)



Kuva 2.1: Epälinearisuuden vaikutus korrelaatiokerroimiin (Xiao et al. 2016)

Korrelaatiokerroin voi saada lukuarvoja väliltä $[-1; 1]$. Lukuarvo -1 tarkoittaa täydellistä negatiivista korrelaatiota, jolloin muuttujan x kasvaessa muuttuja y vähenee. Lukuarvo 1 puolestaan kuvaa täydellistä positiivista korrelaatiota, jolloin muuttujan x kasvaessa muuttuja y kasvaa myös. Täydelliset korrelaatiot ovat hyvin harvinaisia ja ne usein viittaavatkin saman asian mittaamiseen eri muuttujilla. (Drosg 2007, ss. 91–120) Xiaon et al. (2016) näkemys laskennallisten korrelaatiokerroimien merkityksestä muuttujien korrelaatioon on visualisoitu kuvassa 2.2.



Kuva 2.2: Korrelaatiokerroimen merkittävyys arvoalueella (mukailien lähteestä Xiao et al. 2016)

Korkea korrelaatiokerroin ei vielä itsessään kerro tilastollisesta merkittävydestä eli siitä, että tuloksen voisi yleistää pitävän paikkansa myös suuremmassa joukossa. Tilastollista

merkitsevyyttä testataan p-arvolla, jolloin saatua korrelaatiokerrointa verrataan hypoteettiseen tilanteeseen, jossa korrelaatiota ei ole, eli korrelaatiokerroin saa lukuarvon 0. P-arvo saadaan laskettua kaavalla (Mellin 2006, ss. 256–259)

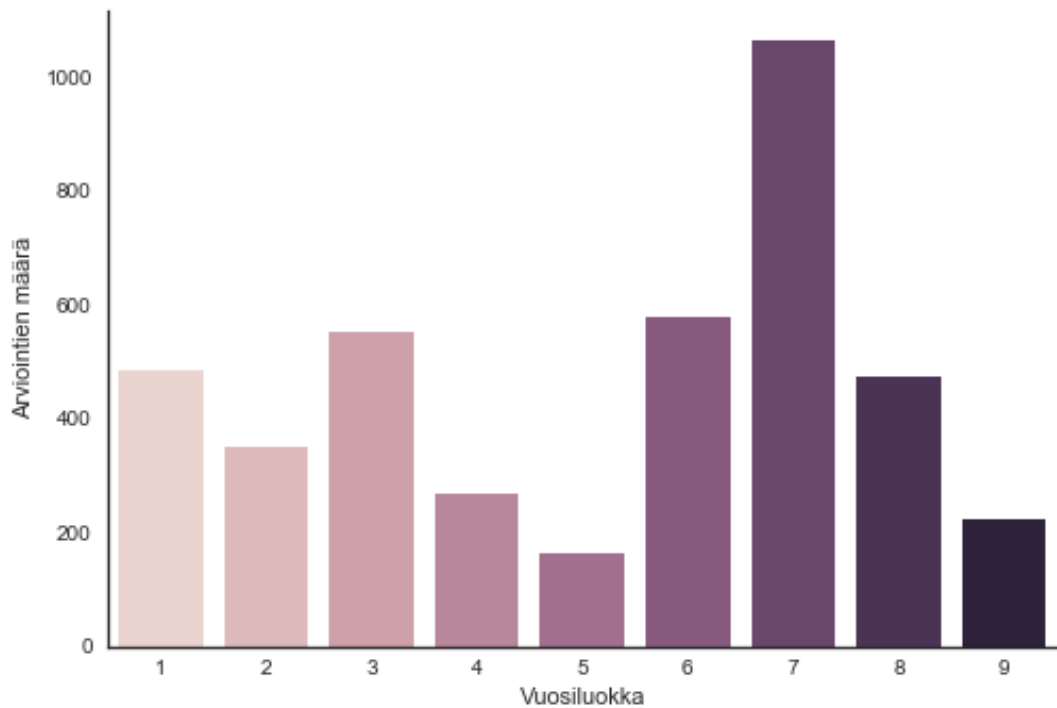
$$t = r \sqrt{\frac{n-2}{1-r^2}}, \quad (3)$$

jossa r on korrelaatiokerroin ja n otoskoko. Tuloksena saatu t noudattaa Studentin t -jakaumaa vapausastein $n-2$. Studentin t -jakauma on todennäköisyysjakauma, jota voidaan hyödyntää keskiarvon tarkasteluun pienilläkin otosko'oilla ilman tulosten luotettavuuden kärsimistä (Heikkilä 2005, s. 106). Yleisesti ottaen alle 5 % suuruista p-arvoa voidaan pitää tilastollisena todisteena korrelaation esiintymisestä perusjoukossa, sillä esimerkiksi Holopainen ja Pulkkinen (2002) sekä Heikkilä (2014) esittävät p-arvon välillä 0,01–0,05 olevan tilastollisesti melkein merkitsevää, p-arvon välillä 0,001–0,01 olevan tilastollisesti merkitsevää ja p-arvon ollessa alle 0,001 sen olevan tilastollisesti erittäin merkitsevää.

Heikkilä (2005, ss. 203–207) ja Ketokivi (2009, ss. 47–52) huomauttavat molemmat, että tilastollisesti merkitseväkään korrelaatio ei vielä tarkoita, että kahden asian välillä vallitsisi kausaalisuhde. Kausaalisuhde tarkoittaa, että muuttujien välillä vallitsee syy-seuraussuhde. Muuttujien korrelaatiokertoimen epäloogisuutta voivat aiheuttaa esimerkiksi epälineaarinen riippuvuussuhde tai molempien muuttujien taustalla vaikuttava kolmas muuttuja, jota ei tarkastelussa ole huomioitu. (Heikkilä 2005, ss. 203–207; Ketokivi 2009, ss. 47–52) Tämän huomioiminen on tärkeää tutkimuksen johtopäätöksissä ja edellyttää myös korrelaatiokertoimien matemaattisten kaavojen ymmärrystä.

2.3 Tutkimusaineiston esittely

Empiirisenä aineistona käytetään SkillzzUp Oy:n keräämää data-aineistoa. Käsiteltävä aineisto on .csv-tiedostossa (comma-separated values), johon on kirjattuna 4 159 oppilaan suoritumista kuvaavaa itse- tai opettaja-arviointia suomalaisista peruskouluista vuosiluokilta 1–9. Näiden vuosiluokkien oppijoista käytetään tässä työssä jatkossa epäselvyyksien välttämiseksi nimitystä oppilas. Arviointeihin liittyvän henkilödatan poisto yksityisyydensuojan takaamiseksi suoritettiin kohdeyrityksessä ennen datan avaamista tutkimuskäyttöön. Aineiston jakautuminen eri vuosiluokille on esitetty kuvassa 2.3.



Kuva 2.3: Aineiston sisältämien arviointien jakautuminen peruskoulun vuosiluokille

Kuvasta 2.3 nähdään, että aineisto ei jakaudu tasaisesti eri vuosiluokille, vaan vuosiluokkakohtainen aineisto vaihtelee 160:n (viides vuosiluokka) ja 1 060:n (seitsemäs vuosiluokka) arvioinnin välillä. Kuvan 2.3 arviointimäärissä ei ole huomioitu kaikkien aineiston sisältämien arviointien sopivuutta tehtäviin analyyseihin, joten analyyseissä käytettävä aineisto tulee olemaan suppeampi. Kuvasta huomataan kuitenkin, että korkeammilla vuosiluokilla on hieman enemmän tehtyjä arviointeja, joka selviää myös vuosiluokkien mediaanin lukuarvosta 6. Tämä tarkoittaa, että aineisto kuvaa hieman keskiverto peruskoululaista korkeammalla vuosiluokalla olevaa oppilasta. Vaikka opintojen eteneminen ja korkeampi vuosiluokka yleisesti yhdistetäänkin tarkempaan itsearviointikykyyn (Falchikov & Boud 1989; Butler 1990), on poikkeavuus kuitenkin melko vähäinen, joten aineiston peilautuvuutta koko peruskouluun pidetään tutkimuksen kannalta riittävänä. Taulukossa 2.1 on kirjattuna aineiston sarakkeet ja selitetty, mitä ne pitävät sisällään.

Taulukko 2.1: Data-aineiston alkuperäiset sarakkeet ja selitteet

Sarake	Selite
<i>ratings_id</i>	Arvioinnin yksilöivä numerotunniste
<i>ratings_value</i>	Arvioinnissa annettu lukuarvo välillä [1; 5], joka kuvaa oppilaan suoriutumista
<i>ratings_date</i>	Ajankohta, jolloin arviointi on luotu ohjelmistoon
<i>rating_is_self_evaluation</i>	Totuusarvo t (engl. <i>true</i> , tosi) tai f (engl. <i>false</i> , epätosi), joka kertoo, onko arviointi itsearvio
<i>ratings_updated_at</i>	Ajankohta, jolloin arviointia on päivitetty
<i>skill_id</i>	Arviointiin liittyvän taidon yksilöivä numerotunniste
<i>user_id</i>	Arvioitavan oppilaan yksilöivä numerotunniste
<i>is_latest</i>	Totuusarvo t tai f, joka kertoo, onko arviointi viimeisin arvio oppilaan arviointikohteesta
<i>course_id</i>	Vuosiluokan ja oppiaineen yksilöivä numerotunniste
<i>subject_name</i>	Arvioitavan oppiaineen nimi
<i>year_name</i>	Arvioitavan oppiaineen vuosiluokka

2.3.1 Aineiston käsittely

Alkuperäisten sarakkeiden perusteella ei ole mahdollista suoraan analysoida itsearviointien tarkkuutta, vaan tämä vaatii datan käsittelyä. Käsittelyn helpottamiseksi luotiin aineistoon uusi sarake *skill_user_id*, joka on yhdiste sarakkeista *skill_id* ja *user_id*. Tämän sarakkeen avulla pystyttiin yksilöimään yksittäiseen arviointikohteeseen yksittäiselle oppilaalle tehdyt arviot. Datasta on ensin jätettävä jäljelle vain sellaiset arvioinnit, joissa tiettyä taitoa ja oppilasta kohtaan löytyy sekä itse- että opettaja-arviointi. Tämä suoritettiin käymällä aineisto rivi kerrallaan läpi ja aina uuden rivin kohdalla tarkastamalla ensimmäiseksi, onko kyseessä viimeisin arvostelu *is_latest*-solun avulla. Tarkasteltavan arvostelun ollessa viimeisin arvostelu kyseiseen arviointikohteeseen, ja *skill_user_id*:n ollessa ennalta tuntematon, eli sitä ei aiemmin ollut esiintynyt, alustettiin tunnisteelle arvo 0 sekä itse- että opettaja-arviointiin. Tämän jälkeen, riippuen siitä kumpi arviointi oli kyseessä, kirjattiin rivin arvio joko itse- tai opettaja-arvioinniksi.

Kaikkien rivien läpikäymisen jälkeen tarkistettiin tehdyt kirjaukset, ja mikäli toisen arvioinnin arvo oli 0, eli itse- tai opettaja-arvio puuttuu, poistettiin rivi käytettävästä aineistosta. Tämän jälkeen aineisto on tutkimuksen kannalta käyttökelpoinen ja siinä on jäljellä vain hyödynnettäviä rivejä. Kaiken kaikkiaan lopullinen aineisto piti sisällään 2 067 riviä, jotka käsittävät 1 032 arviointikohdetta, eli yksilöllistä *skill_user_id*-sarakkeen arvoa, joille löytyy sekä opettaja- että itsearvio. Analyysia ja visualisointia varten aineiston käsittelyssä luotiin karsittu datataulu, jossa on kaikki hyväksyttävät rivit ja niiden sarakkeet. Lisäksi samasta datataulusta luotiin dict-kirjasto, jossa avaimena on arvioinnin yksilöivä *skill_user_id*-tunniste ja arvona sille annettu itse- ja opettaja-arvio kahden alkion mittaisessa listassa. Yksityiskohtaisempi datan käsittelyyn käytetty ohjelmakoodi on esitetty liitteessä A.

3. TEOREETTINEN TAUSTA

Tässä luvussa esitellään tutkimuksen aihepiiriin liittyvä teoreettinen tausta. Luvussa käsitellään aluksi itsearviointin määritelmiä ja sen roolia sekä yleisellä tasolla että Suomen peruskoulujen tasolla. Luvussa esitellään myös aihepiiristä aiemmin tehtyä tutkimusta, joiden perusteella koostetaan data-aineistosta testattavat hypoteesit.

Taulukko 3.1: Käytetyt hakulausekkeet ja hakutulokset

Haettavan kirjallisuuden aihepiiri	Käytetty hakulauseke	Andor	Scopus
Itsearviointikykyyn vaikuttavat tekijät	("student self-assessment") AND ("accuracy") AND ("education")	1310	37
Data-analytiikan hyödyntäminen opetuksessa	("EDM" OR "educational data mining") AND ("LA" OR "learning analytics") AND ("data science")	103	6
EDM:n hyödyntäminen oppimisen arvioinnissa	("EDM" OR "educational data mining") AND ("assessment") AND ("learning")	5	89

Taulukossa 3.1 on esitetty tutkimuksen teoriataustan kirjallisuusmateriaalien etsintään käytetyt hakulausekkeet ja niiden perusteella saadut tulokset. Teoriaosuuden lähdemateriaalia haettiin ensisijaisesti hakulausekkeilla löydettyistä teoksista, joiden soveltuvuutta arvioitiin tiivistelmien pohjalta. Hakulausekkeiden lisäksi hyödynnettiin hyväksi todettujen tutkimuksien lähdeluetteloita ja niiden perusteella löydettyä lisätutkimusta sekä saatuja artikkelisuosituksia. Suomen koulujärjestelmän uniikin luonteen takia teoriakirjallisuuden osalta pyritään hyödyntämään myös kotimaisia julkaisuja aihepiiriin liittyen.

3.1 Itsearviointi

Itsearvioinnissa on kyse siitä, että oppija ymmärtää arvioitavan kohteen vaatimustason sekä kriteerit, ja ymmärtämisen jälkeen osaa verrata omaa suoriutumistaan näihin asetettuihin tavoitteisiin (Rust et al. 2003). Taras (2010) puolestaan korostaa, että itsearvioinnissa on huomioitava lopputulosten ja tavoitteiden yhteensopivuuden lisäksi myös työskentelyn taso suorituksen aikana. Anttila (2013, s.107) näkee itsearvioinnin prosessina, jonka eri vaiheita pohditaan osana oppimista. Itsearvioinnin prosessissa korostetaan oman työskentelyn tarkastelua, oppimista sekä omien vahvuuksien ja heikkouksien havaitsemista. Itsearvioinnin tarkoitus on muodostaa oppijalle käsitys omista taidoistaan ja kehittämiskohteistaan. (Anttila 2013, s. 107) Boud (1995, s. 17) tiivistääkin, että itsearvioinnissa on kyse oppijan oppimistaitojen kehittämisestä, ei arvosanojen antamisesta tai opettajan vastuun siirtämisestä. Määritelmien pohjalta voidaankin yleistää itsearvioinnin olevan tavoitteiden ymmärtämisestä, suorituksen tekemiseen ja oman suoriutumisen arviointiin etenevä prosessi, jonka tavoitteena on kehittää oppijan kykyä oppia ja löytää omia

vahvuuksiaan sekä heikkouksiaan. Yleisen tason tavoitteena itsearvioinnille voidaan pitää kykyä toimia jatkossa vastaavassa tilanteessa tehokkaammin ja suoriutua paremmin.

Itsearvioinnin rooli suomalaisessa peruskoulussa perustuu perusopetuslakiin, jonka mukaan koko oppilaan arvioinnin tarkoituksena on kehittää, kannustaa sekä ohjata yksilön oppimista ja antaa oppilaille edellytykset itsearviointiin (Perusopetuslaki, 1998\628 § 22). Tällä hetkellä voimassa oleva perusopetuksen opetussuunnitelman perusteet 2014, johon kaikkien paikallisten opetussuunnitelmien tulee lain mukaan pohjautua, täydentää, että oppilaita on ohjattava havainnoimaan työskentelyään itsearvioinnin taitojen kehittämiseksi. Perusopetuksen opetussuunnitelmassa määritetään myös karkeat ohjeistukset sille, miten oppijan itsearviointitaitoja tulee kehittää eri vuosiluokilla. (Opetushallitus 2014a, s. 47-49)

Itsearvioinnin keskeistä merkitystä osana oppimisprosessia tukevat useat tutkimustulokset (e.g. Dochy et al. 1999; Fitzgerald et al. 2003; Lew et al. 2010; Hatami 2015). Dochyn et al. (1999) tutkimuksessa käydään läpi 63 aihepiiriin liittyvää artikkelia, joiden perusteella löytyi positiivinen yhteys itsearvioinnin ja oppimisen välillä. Oppijat, jotka hyödyntävät itsearviointia osana oppimista, pärjäsivät paremmin koetilanteissa ja oppivat myös paremmin arvioimaan omaa suoriutumistaan. (Dochy et al. 1999) Hatami (2015) puolestaan löytää tutkimuksissaan positiivisen yhteyden itsearviointien tekemisen ja ajattelemaan oppimisen välillä tutkittuaan eri opetustekniikoiden vaikutuksia itsearviointeihin ja opettaja-arviointeihin. Tämän yhteyden kautta itsearvioinnin voitiin nähdä tukevan oppimista. Fitzgerald et al. (2003) huomaavat tutkimuksessaan itsearviointitarkkuuden ja opiskelumenestyksen välillä olevan yhteyden. Vaikka pääsääntöisesti itsearviointi ja oppiminen yhdistetäänkin toisiinsa parempien oppimistulosten välityksellä, eivät kaikki tutkimukset ole samaa mieltä näiden välisestä yhteydestä (e.g. Maguire et al. 2001; Eva et al. 2004). Esimerkiksi Eva et al. (2004) väittävät tutkimuksessaan, että aiemman suoriutumistason perusteella pystyi tarkemmin ennakoimaan tulevia koearvosanoja kuin oppijan itse tekemillä arvioinneilla.

3.2 Data-analytiikka opetuksessa

Data-analytiikan hyödyntämistä opetuksessa on tutkittu runsaasti (e.g. Baker & Yacef 2009; Alshammari et al. 2013; Romero & Ventura 2013; Papamitsiou & Economides 2014) ja siihen liittyvä tutkimus onkin jakaantunut pitkälti kahteen eri tutkimussuuntaan, jotka ovat EDM (engl. *Educational Data Mining*) ja LA (engl. *Learning Analytics*) (Papamitsiou & Economides 2014). Romero ja Ventura (2013) määrittelevät EDM:n tietotekniikkaa apuna käytävien metodien tutkimiseksi, kehittämiseksi ja hyödyntämiseksi opetukseen liittyvissä suurissa datamassoissa, joita olisi vaikeaa tai mahdotonta analysoida ilman tietoteknisiä välineitä. Tutkimuskentän yhtenä perustajana tunnettu Baker yhdessä Yacefin (2009) kanssa esittävät EDM:n olevan joukko metodeja, joilla tutkitaan opetuksesta syntyvää dataa, jotta oppijoita ja oppimisympäristöjä voidaan ymmärtää pa-

remmin. Kaksi vuotta tämän jälkeen Baker (2011) puolestaan esitti tiiviimmän määritelmän, jonka mukaan EDM on työkalu, jolla etsitään hyödyllisiä tietoja opetuksen tarjoajan sähköisestä datasta.

Siemens (2012) määrittää LA:n oppijoiden ja heidän konteksteihinsa liittyvän datan mitaamiseksi, keräämiseksi, analysoimiseksi ja raportoimiseksi, jotta oppimista ja oppimisympäristöä voidaan paremmin ymmärtää ja kehittää. Elias (2011) puolestaan määrittelee LA:n nousevaksi tutkimuskentäksi, joka hyödyntää moderneja analytiikkatyökaluja opetuksen ja oppimisen edistämiseksi. Esitettyjen määritelmien pohjalta voidaan tunnistaa, että molemmat tutkimussuunnat pyrkivät kehittämään opetusta hyödyntäen tietotekniikan mahdollistamia tiedon käsittelytapoja. Saman johtopäätöksen esittävät Romero ja Ventura (2013), Papamitisiou ja Economides (2014) sekä Chatti et al. (2013), mutta lisäävät kuitenkin huomion, että tutkimussuunnat eroavat tekniikoiltaan, painotuksiltaan ja löydöksiltään.

Siemens ja Baker (2012) selvittivät tutkimuksessaan LA:n ja EDM:n eroja sekä yhtäläisyyksiä ja tunnistivat seuraavat keskeiset erot: holismin osalta EDM keskittyy enemmän yksittäisten komponenttien yhdistämiseen ja tulkitsemiseen, kun taas LA pyrkii hahmottamaan aihepiirin kokonaisuutta. EDM:ssä korostuu opetusohjelmistojen käyttö ja oppijan suoriutumisen ennakoiminen. Tekniikoiden ja metodien osalta LA:han liitetään usein mm. verkostanalyysi, sentimenttianalyysi ja päätöspuuanalyysit. EDM:ssä puolestaan hyödynnetään esimerkiksi luokittelua, klusterointia, korrelaatioanalyysiä ja visualisointia. Tutkimussuunnat pyrkivät yhtenevästi kehittämään datan hyödyntämistä erityisesti oppimisen arvioinnissa. Tutkimussuuntia yhdistää myös pyrkimys tutkimuksen ja käytännön edistämiseen dataa hyödyntävässä opetuksessa. (Siemens & Baker 2012) EDM:n ja LA:n erojen ja yhtäläisyyksien määrittämisen myötä, voidaan tunnistaa tämän tutkimuksen keskittyvän nimenomaan EDM:ään.

3.3 EDM:n hyödyntäminen oppimisen arvioinnissa

Tehdyssä selvitystyössä ei löydetty tutkimusta EDM:n hyödyntämisestä oppimisen arvioinnissa tätä tutkimusta vastaavassa kontekstissa. Kansainvälisesti analytiikkaa hyödyntävää oppimisen arviointia on tutkittu jonkin verran, mutta tehty tutkimus kohdistuu lähinnä vanhempiin oppijoihin, useimmiten korkeakoulutasolla oleviin. Merkittävimpiin aihepiiriin tutkimuksiin lukeutuu Falchikovin ja Boudin (1989) meta-analyysi, joka osoittaa merkittävää korrelaatiota itsearviointien ja opettaja-arviointien välillä. Tutkimuksen perusteella tunnistettiin useita itsearviointien tarkkuuteen vaikuttavia tekijöitä: Pidemmälle edenneet oppijat olivat tarkempia itsearvioijia kuin opintojen alkuvaiheessa olevat. Toisena tekijänä huomattiin, että tarkasti suunnitellut opinnot paransivat oppijoiden itsearviointien tarkkuutta. (Falchikov & Boud 1989)

Ross et al. (1999) tutkivat itsearviointien opetuksen vaikutusta vuosiluokilla 4–6 olevien oppilaiden itsearviointitarkkuuteen ja kirjoitustaitoon. Tutkijat tarkkailivat oppilasryhmää, jolle tietoisesti opetettiin itsearviointien tekemistä, ja vertasivat tätä kontrolliryhmään, jolle itsearviointia ei opetettu. Tuloksina löydettiin selkeää itsearviointitarkkuuden paranemista, sekä heikoiten suoriutuvilla oppilailla jopa kirjoitustaidon kehittymistä ryhmässä, jossa itsearviointia tietoisesti opetettiin. (Ross et al. 1999) Olettaessa huomioon otettuna opetussuunnitelman määrittämä edellytys itsearviointien opettamiseen (Opetushallitus 2014b), voidaan arvella suomalaisten peruskoululaisten itsearviointitarkkuuden kehittyvän vuosiluokalta toiselle siirryttäessä.

Dochy et al. (1999) käsitelivät aiempaa tutkimusta liittyen itse-, vertais- ja yhteisarviointiin. Tutkimuksen tulokset osoittivat itsearviointien korreloivan hyvin opettaja-arviointien kanssa. Itsearviointiin osallistuminen korreloi tutkimuksen mukaan myös korkeiden arvosanojen kanssa, eli itsearviointia tekevät oppijat suoriutuivat keskimäärin paremmin opinnoissaan. Käytetyn mittarin perusteella itsearviointi paransi siis oppijoiden kykyä oppia. Itsearviointiin huomattiin vaikuttavan positiivisesti oppijan korkea motivaatio ja omasta arvioinnistaan saatu palaute. Lisäksi itsearviointien positiivisten vaikutusten huomattiin korostuvan tehtäessä itsearviointeja pidemmällä aikavälillä. (Dochy et al. 1999) Tuoreemman tutkimuksen osalta Karnilowicz (2012) esittää 64 psykologiaopiskelijan itsearviointien korreloivan hyvin opettaja-arviointien kanssa. Karnilowiczin (2012) tutkimus tukee myös Falchikovin ja Boudin (1989) sekä Boudin et al. (2013) tuloksia, joiden mukaan hyvin suoriutuvat oppijat itsearvioivat suoriutumistaan heikommaksi kuin se todellisuudessa on. Sama ilmiö tapahtuu myös päinvastaisesti, sillä heikommin suoriutuvat arvioivat suoriutumistaan yleensä todellisuutta paremmaksi (Falchikov & Boud 1989; Karnilowicz 2012; Boud et al. 2013).

Fitzgerald (2003) huomasi tutkiessaan lääketieteellisen koulun opiskelijoita, että itsearviointien tarkkuus on riippuvainen vastaavien itsearviointien määrästä. Tarkatkaan itsearviointien eivät osanneet arvioida suoriutumistaan tehtävätyypin vaihtuessa. (Fitzgerald et al. 2003) Itsearviointien tarkkuudella viitataan itsearviointien tekijän kykyyn arvioida mahdollisimman samankaltaisesti kuin vertailukohteena käytettävä arvioija. Samansuuntaisia tuloksia esittävät Boud et al. (2013), jotka huomasivat tutkimuksessaan alemman korkeakoulutason oppijoiden itsearviointikyvyn paranevan, mikäli oppijat harjoittelevat itsearviointia pidemmällä aikavälillä.

Lew et al. (2010) tutkivat lukukauden ajan 3 588 ensimmäisen vuoden yliopisto-opiskelijan itsearviointien ja opettaja-arviointien suhdetta toisiinsa, ja aiemmista tutkimuksista poiketen onnistuivat löytämään ainoastaan kohtuullista korrelaatiota arviointien väliltä. Tutkimuksen löydöksenä oli myös, että riippumatta siitä, pitivätkö oppijat itsearviointia merkityksellisenä oppimiselle, pysyi itsearviointien tarkkuus samanlaisena. Tutkimuksessaan he eivät myöskään löytäneet todisteita itsearviointikyvyn kehittymisestä pidemmälle edenneillä opiskelijoilla. (Lew et al. 2010) Tuloksen perusteluita voidaan etsiä oppijoiden korkeasta kouluasteesta, sillä ala-aste-ikäisten oppilaiden kohdalla esimerkiksi

Butler (1990) esittää itsearviointikyvyn muuttuvan merkittävästi vuosiluokkien välillä. Epävarmoja tuloksia itsearvioinnin roolista oppimisen tukijana tarjoavat myös Maguire et al. (2001), jotka esittivät tutkimuksessaan, että itsearvioinnista tietoiset opiskelijat suhtautuvat koko tehtävään strategisesti ja hyödyntävät tulevaa itsearviointia taktikoidakseen jo suoriutumisen aikana. Tämä puolestaan vie painoarvoa itse oppimiselta.

Vaikka Suomen peruskouluun liittyvää vastaavaa tutkimusta ei löytynyt, voidaan Johanssonin (2013) tutkimusta ruotsalaisten kolmasluokkalaisten itsearviointikyvystä verrata tähän tutkimukseen. Tutkimus osoittaa, että jo kolmasluokkalaiset oppilaat pystyvät itsearvioimaan omaa lukutaitoaan sellaisella tasolla, että itsearviointien ja opettaja-arviointien välillä on selkeä korrelaatio. Kokonaisuudessaan tutkimusta voidaan pitää luotettavana, sillä se sisälsi 5 271 kolmasluokkalaisten oppilaan kokeet, joita oli arvioinut 351 opettajaa. (Johansson 2013) Suomessa tehtyä tutkimusta itsearvioinnin ja opettaja-arviointien korrelaatioista edustaa Lindblom-Ylänne et al. (2006) tekemä tutkimus, jossa verrattiin 15 lukiopiskelijan itse-, vertais- ja opettaja-arviointeja. Tutkimuksen tuloksena eri arviointitavoista saatujen tulosten väliltä löydettiin selkeitä positiivisia korrelaatioita. Itsearvioinnin tarkkuuteen huomattiin vaikuttavan voimakkaasti ohjeistuksen tarkkuus, sillä yksityiskohtaisemmat arviointiohjeet paransivat odotetusti myös itsearviointien tarkkuutta. (Lindblom-Ylänne et al. 2006)

Kokonaisuudessaan löydetyt tulokset EDM:n hyödyntämisestä oppimisen arvioinnissa ovat pääsääntöisesti yhteneviä. Kansainvälisesti tehdyissä tutkimuksissa ei vaikuta olevan merkittäviä eroavaisuuksia, vaan itsearviointikyky vaikuttaa kirjallisuuden perusteella universaalilta. Läpikäydyn tutkimuksen perusteella voidaan koostaa data-aineistoa vasten testattavat hypoteesit:

- H₁: Oppilaiden itsearviot ja opettajien tekemät arviot korreloivat keskenään positiivisesti.
- H₂: Itsearviointien suurempi määrä parantaa oppilaan kykyä arvioida omaa oppimista.
- H₃: Korkeammalla vuosiluokalla olevat oppilaat ovat parempia itsearvioijia.
- H₄: Paremmin suoriutuvat oppilaat aliarvioivat suoriutumistaan ja heikommin suoriutuvat yliarvioivat suoriutumistaan.

Hypoteesien kohdalla on huomioitava, että niissä voidaan käsitellä itsearviointiin liittyen vain asioita, joita on esitellyn data-aineiston perusteella mahdollista testata.

4. KORRELAATIOT DIGITAALISESSA ARVIOINTIAINEISTOSSA

Luvussa tutkitaan suomalaisesta peruskoulusta kerättyä data-aineistoa ja esitetään mitaustulokset. Kerätystä arviointidatasta etsitään korrelaatioita, jotka selittäisivät oppilaan kykyä itsearvioida omaa oppimistaan. Hypoteesien pohjalta tehtävien testien tulokset visualisoidaan ja esitetään matemaattisina arvoina.

4.1 Oppilaan itsearviointikyky

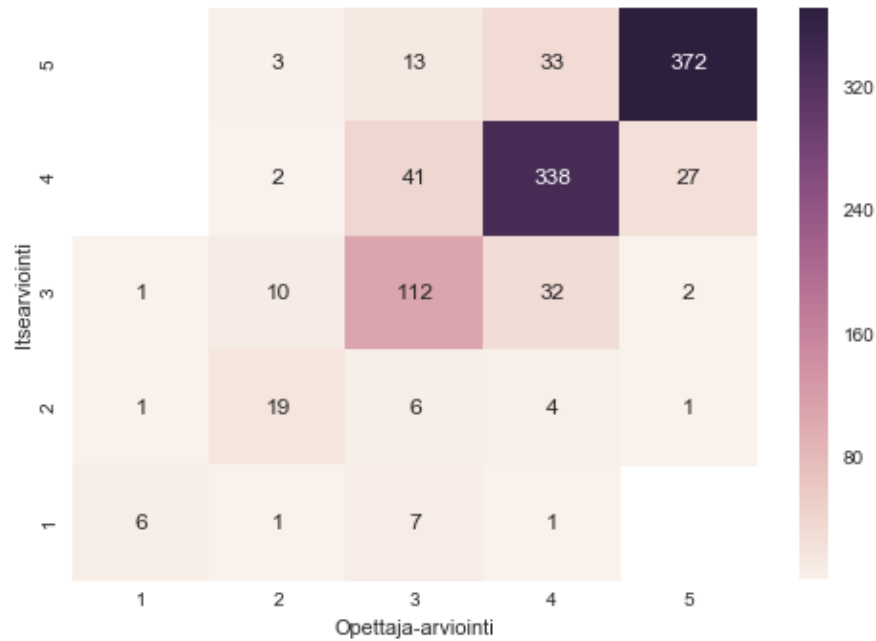
Kappaleessa tutkitaan kirjallisuuden perusteella tunnistettuja hypoteeseja hyödyntäen data-aineistosta saatuja tutkimustuloksia ja Pythonin matplotlib- sekä seaborn-kirjastojen avulla visualisoituja tuloksia.

4.1.1 Itse- ja opettaja-arviointien korrelaatio

Korrelaatiotaulukkoa varten luotiin aluksi tyhjä Pythonin dict-kirjasto, jonka avaimena toimi yhdiste *skill_id* ja *user_id* sarakkeista ja arvona lista, jonka ensimmäinen alkio vastasi opettajan tekemää arviota ja toinen alkio oppilaan tekemää arviota. Kirjaston avaimet ja arvot saatiin suodatetusta aineistosta etsimällä aina kunkin *skill_id* ja *user_id*-kombinaation *is_latest* sarakkeesta arvo 't'. Sama haku toistettiin sekä itse- että opettaja-arvioinneille. Lopputuloksena saadussa kirjastossa oli arvona aina viimeisin annettu arviointi sekä opettajalta että oppilaalta itseltään.

Kirjaston perusteella koostettiin pandas-datataulu, joka indeksoitiin *skill_id*:n ja *user_id*:n yhdisteellä, ja jonka ensimmäinen sarake koostui oppilaiden itsearviointeista ja toinen sarake opettajien antamista arvioinneista. Tätä datataulua voitiin käyttää suoraan korrelaatiokertoimen laskemiseen ja tulosten visualisointiin. Yksityiskohtaisempi datan käsittely ja visualisointi on esitetty liitteessä B.

Laskemalla oppilaan itsearviointi ja opettaja-arviointi sarakkeiden välille Spearmanin järjestyskorrelaatiokerroin, saadaan tulokseksi 0,828. Lasketun korrelaatiokertoimen p-arvoksi saadaan $3,13 * 10^{-235}$, joka on hyvin alhainen. Itse- ja opettaja-arviointien riippuvuus toisistaan on esitetty kuvassa 4.1.



Kuva 4.1: Itse- ja opettaja-arviointien korrelaatio

Kuvasta 4.1 nähdään, että merkittävin osuus tehdyistä arvioinneista sijoittuu suoralle, niin että itse- ja opettaja-arvioinnissa on annettu sama arvosana. Arvioinnit, joissa itse- ja opettaja-arvosanojen välillä on ollut eroa, ovat jakautuneet melko tasaisesti molemmille puolille suoraa.

4.1.2 Arviointien määrän vaikutus itsearviointikykyyn

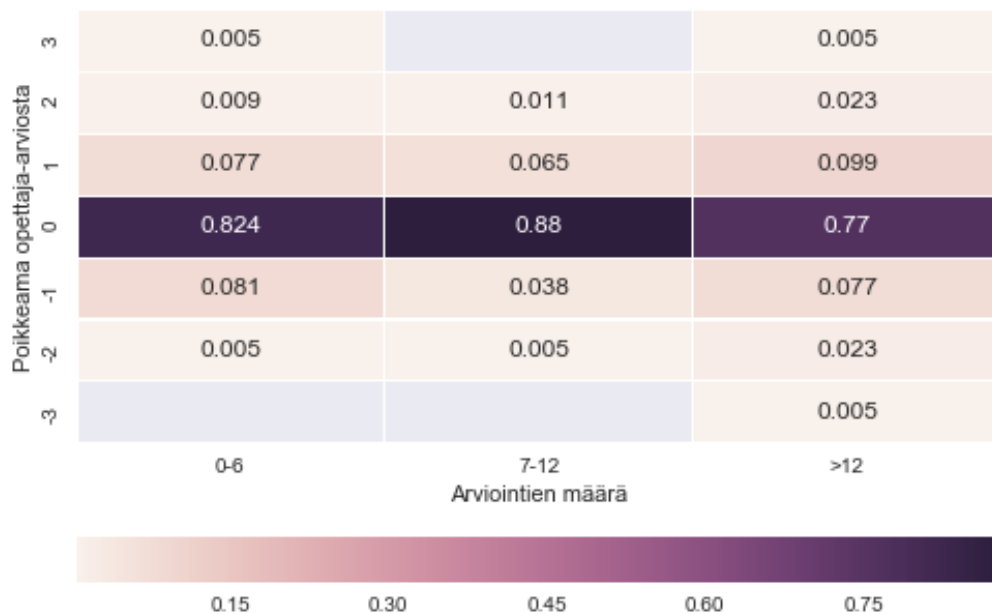
Arviointimäärän, eli kuinka paljon arviointeja on tehty, vaikutusta oppilaan kykyyn itsearvioida omaa oppimistaan tutkitaan jakamalla käytettävän aineiston oppilaat kolmeen eri käyttäjäryhmään aktiivisuuden mukaan: oppilaat, jotka ovat tehneet 6 itsearviointia tai vähemmän (73 oppilasta), oppilaat, joilla on 7–12 itsearviointia (74 oppilasta) ja aktiivisimpana ryhmänä oppilaat, joilla on 13 itsearviointia tai enemmän (78 oppilasta). Tehdyissä itsearviointikohteissa tulee olla myös opettajan arvio, jotta sen tarkkuutta voidaan verrata. Kategoriat jaoteltiin siten, että kaikissa eri käyttäjäryhmissä on mahdollisimman samanlainen määrä arvioita tehneitä henkilöitä. Tästä syystä ryhmien väliset arviointimäärät ovat hyvin erilaisia ja tulokset esitetään prosentiosuuksina oman vertailuryhmän sisällä.

Segmentoinnin jälkeen luodaan kaikille oppilasryhmille omat datataulut, joista löytyy ainoastaan segmenttiin hyväksyttävien rivien itse- ja opettaja-arvioinnit. Näiden datataulujen osalta lasketaan korrelaatio kullekin datataululle samalla tavalla kuin kaikkien hyväksytyjen itsearviointien korrelaatio. Yksityiskohtaisempi aineiston käsittely ja visualisointien toteutus on nähtävillä liitteessä C. Ryhmille saadut korrelaatiot on esitetty taulukossa 4.1.

Taulukko 4.1: Käyttäjäsegmenttien vaikutus korrelaatioon

Arviointien määrä	Korrelaatiokerroin	P-arvo
1–6	0,817	$2,0 * 10^{-54}$
7–12	0,880	$1,4 * 10^{-120}$
>12	0,799	$7,6 * 10^{-100}$

Taulukosta 4.1 huomataan, että korrelaatiokerroimet ovat hyvin lähellä toisiaan, eikä niissä ole merkittäviä eroja. P-arvo on odotetusti pienin vähiten arvioiteja tehneellä käyttäjäsegmentillä otoskoon takia. Kaikki lasketut p-arvot ovat kuitenkin hyvin alhaisia ja täten tukevat korrelaatiokerroimen tilastollista merkitsevyyttä. Kuvassa 4.2 on esitetty eri käyttäjäsegmenttien itsearviointien poikkeavuudet opettaja-arvioista prosenttisuuksina omasta käyttäjäsegmentistään.

**Kuva 4.2: Arviointimäärän vaikutus arvioinnin tarkkuuteen**

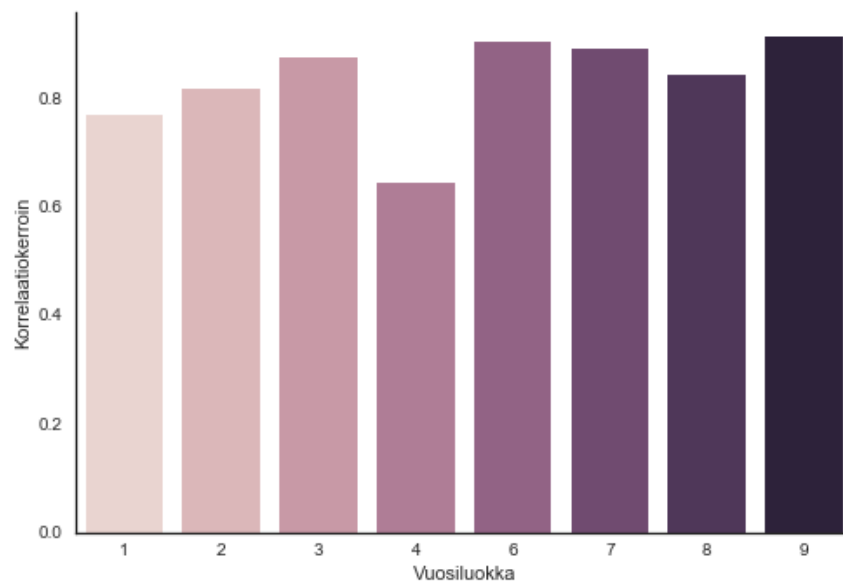
Kuva 4.2 tukee taulukossa 4.1 esitettyjä laskutuloksia. Kaikilla käyttäjäsegmenteillä suurin prosentuaalinen osuus itsearvioinneista on arvioitu samalla tavalla kuin opettaja-arviointi. Keskenään eriävät arvioinnit ovat jakautuneet melko tasaisesti paremmiksi ja huonommiksi kaikissa käyttäjäsegmenteissä, mutta paljon arvioiteja tehneillä käyttäjillä on kuitenkin lievästi enemmän hajontaa tekemissään arvioinneissa.

4.1.3 Vuosiluokan vaikutus itsearviointikykyyn

Vuosiluokan vaikutusta oppilaan itsearviointikykyyn tutkitaan laskemalla korrelaatiot erikseen peruskoulun eri vuosiluokille. Luokitteluun käytetään arviointiin liittyvää *year_name*-saraketta, jonka perusteella oppilaan vuosiluokka voidaan tunnistaa. Korrelaatioiden selvittämiseksi luotiin dict-kirjasto, jonka avaimiksi alustettiin vuosiluokkien

numerot ja arvoiksi tyhjat listat. Tämän jälkeen käytiin kaikki käyttökelpoiset arvostelut läpi ja lisättiin arviointikohteen yksilöivä *skill_user_id*-kenttä oikeaan vuosiluokka liittyvään listaan.

Tarkastelusta joudutaan jättämään pois 5. vuosiluokka, sillä kyseiseltä luokalta on saatavilla vain 2 vertailukelpoista arviointia. Muiden vuosiluokkien osalta otoskoot vaihtelevat välillä 97 (9. vuosiluokka) – 358 (1. vuosiluokka), keskiarvon ollessa 258 arviointia vuosiluokkaa kohden. Otoskoissa on merkittäviä eroja ja tämän takia tuloksiin onkin suhtauduttava varovaisesti. Ryhmittelyn jälkeen luodaan jokaiselle vuosiluokalle oma datataulu, johon kirjataan kriteerit täyttävät itse- ja opettaja-arvioinnit. Näille datatauluille lasketaan puolestaan omat korrelaatiokertoimet sarakkeiden välille. Tarkempi kuvaus datan käsittelystä ja visualisoinneista on nähtävillä liitteessä D. Vuosiluokille lasketut korrelaatiokertoimet on esitetty kuvassa 4.3.



Kuva 4.3: Vuosiluokkakohtaiset korrelaatiokertoimet

Kuvasta 4.3 voidaan huomata lievää positiivista trendiä korrelaatiokertoimessa vuosiluokan kasvaessa. Alakoulun keskimääräiseksi korrelaatiokertoimiseksi saadaan laskennallinen arvo 0.801 ja yläasteen keskimääräiseksi korrelaatiokertoimeksi 0.882. Vuosiluokkakohtaiset p-arvot ovat välillä 10^{-15} – 10^{-63} . P-arvojen perusteella saadut korrelaatiokertoimet ovat kaikki tilastollisesti merkitseviä otoskoon vaihteluista huolimatta.

4.2 Opiskelumenestyksen vaikutus itsearviointiin

Opiskelumenestyksen vaikutuksista itsearviointitarkkuuteen tutkittiin jakamalla oppilaat opettajien arviointien keskiarvon mukaan parhaiten ja heikoiten suoriutuneisiin segmentteihin. Oppilastyypin tunnistamiseksi data-aineisto jaettiin kahdeksi erilliseksi datatauluksi: opettaja-arviointitauluun, jossa aineiston *rating_is_self_evaluation*-solun arvo on 't', ja itse-arviointitauluun, joilla solun arvo on 'f'. Opettaja-arviointitaulusta luodaan

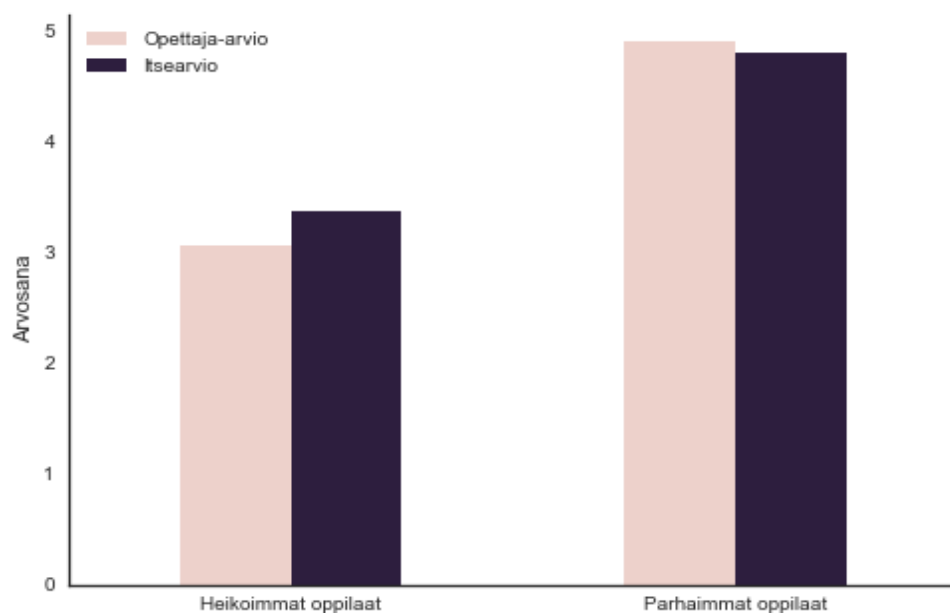
dict-kirjasto, jossa avaimena on oppilaan *user_id* ja arvona lista tälle annetuista arvosanoista. Listasta lasketaan oppilaskohtainen opettaja-arviointien keskiarvo, jonka mukaiseen suuruusjärjestykseen dict-kirjasto järjestetään. Heikoimpia arviointeja saaneiden oppilaiden joukkoon valitaan 20 % menestysjärjestyksen alkupäästä ja parhaita arviointeja saaneiden oppilaiden joukkoon puolestaan 20 % menestysjärjestyksen loppupäästä. Tämän jälkeen käytössä on kaksi 45 oppilaan *user_id* listaa, joissa on parhaiten ja heikoiten arvioidut oppilaat.

Näistä listoista lasketaan kummankin ryhmän osalta opettaja-arviointien keskiarvo vähennettynä itsearviointien keskiarvosta jokaiselle oppilaalle. Tämän jälkeen näistä erotuksista otetaan keskiarvo, joka kuvaa segmentin keskimääräistä itsearviointia suhteessa opettaja-arvioon. Listoista lasketaan myös, kuinka moni 45 oppilaasta arvioi suoriutumisansa paremmaksi kuin opettaja ja toisaalta, kuinka moni heikommaksi kuin opettaja. Tarkempi datan käsittelyn ja visualisoinnin suorittava ohjelmakoodi on esitetty liitteessä E. Laskemalla saadut tulokset on kirjattu taulukkoon 4.2.

Taulukko 4.2: Itse- ja opettaja-arviointien eroavaisuudet opintomenestyksen perusteella

Oppilastyyppi	Parhaiten suoriutuneet oppilaat	Heikoiten suoriutuneet oppilaat
Suorituksen yliarvioineet	2	24
Suorituksen aliarvioineet	7	3
Opettaja-arvioiden keskiarvo	4,907	3,053
Itsearvioiden keskiarvo	4,800	3,372
Arvioiden erotus	0,107	-0,319

Taulukosta 4.2. nähdään, että arviointien keskiarvojen välillä on molemmissa ryhmissä eroja. Myös suoritustaan yli- ja aliarvioineet ovat hyvin riippuvaisia oppilastyypistä. Molempien oppilastyypin arviointien keskiarvot on visualisoitu kuvassa 4.4.



Kuva 4.4: Koulumenestyksen vaikutus arviointien keskiarvoon

5. TULOSTEN ANALYSOINTI

Luvussa vastataan asetettuihin hypoteeseihin neljännen luvun laskennallisten arvojen ja visualisointien perusteella eli analysoidaan mittaustuloksia. Tulosten esittämisen jälkeen arvioidaan tulosten merkittävyyttä ja niihin vaikuttaneita tekijöitä.

5.1 Tutkimuksen tulokset

Ensimmäisen hypoteesin mukaan itse- ja opettaja-arviointien välillä esiintyy positiivista korrelaatiota. Hypoteesia testattiin laskemalla korrelaatiokerroin kaikille arviointikohteille, joista oli saatavilla itse- ja opettaja-arvio. Arviointien välille lasketuksi korrelaatiokertoimeksi saatiin lukuarvo 0,828, joka viittaa voimakkaaseen positiiviseen korrelaatioon muuttujien välillä. Korrelaatiokertoimen laskennalliseksi p-arvoksi saatiin $3,13 * 10^{-235}$. Tämä tarkoittaa tuloksen olevan tilastollisesti erittäin merkitsevä verrattaessa esimerkiksi Holopaisen ja Pulkkinen (2002) sekä Heikkilän (2005) näkemyksiin tilastollisesta merkitsevyydestä, jotka esittävät alle 0,001 p-arvon olevan tilastollisesti erittäin merkitsevä. Saatu p-arvo viittaa korrelaation ilmenemättömyyden olevan lähes mahdotonta, eli muuttujien välillä selvästikin on havaittavissa voimakasta korrelaatiota.

Voimakas positiivinen korrelaatio voidaan myös visuaalisesti huomata kuvasta 4.1, jossa suurin osa arvioinneista sijoittuu suoralle, jolla itse- ja opettaja-arvioinnit ovat samoja eli molemmissa arvioinneissa on annettu sama arvosana. Muut arvioinnit ovat jakautuneet melko tasaisesti molemmille puolille suoraa, eli tämän perusteella ei ole huomattavissa merkittävää suorituksen yli- tai aliarvioimista. Kokonaisuudessaan laskennallinen korrelaatiokerroin, p-arvo ja tulosten visualisointi tukevat asetettua ensimmäistä hypoteesia, joten sen hylkäämiseen ei ole syytä. Aineiston perusteella itse- ja opettaja-arvioinnit korreloivat positiivisesti.

Toisena hypoteesina itsearviointien määrän arveltiin korreloivan positiivisesti itsearvioinnin tarkkuuden kanssa. Hypoteesia testattiin jakamalla aineiston oppilaat kolmeen eri käyttäjäsegmenttiin arviointiohjelmiston käytön aktiivisuuden perusteella ja laskemalla näille segmenteille omat korrelaatiokertoimet ja p-arvot. Käyttäjäsegmenteille lasketut korrelaatiokertoimet sijoittuivat lukuarvojen 0,7999 ja 0,880 välille (taulukko 4.1) niin, että alhaisin korrelaatiokerroin oli eniten arviointeja tehneillä oppilailla ja korkein keskinkertaisen määrän arviointeja tehneillä oppilailla. Kullekin segmentille laskettu p-arvo oli hyvin alhainen ja tuki laskettujen korrelaatiokertoimien tilastollista merkitsevyyttä. Korrelaatiokertoimien perusteella ei voida tunnistaa positiivista tai negatiivista trendiä itsearviointikyvyn ja arviointikertojen määrän välillä, sillä kaikkien segmenttien korrelaatiokertoimet ovat hyvin lähellä toisiaan eikä niiden välillä ole monotonista yhteyttä

arviointikertojen määriin. Tulokset eivät siis ole linjassa Fitzgeraldin (2003) ja Boudin et al. (2013) löydösten kanssa.

Kuvassa 4.2 esitettiin eri käyttäjäsegmentteihin kuuluvien oppilaiden tekemien arviointien poikkeamat opettaja-arvioista prosenttiosuuksina kyseisestä segmentistä. Kuvassa 4.2 visualisoidut tulokset tukevat laskennallisia korrelaatiokertoimia. Keskinkertaisesti arviointeja tehneillä oppilailla on selkeästi eniten itsearviointeja, jotka eivät poikkea opettaja-arvioinnista. Eniten arviointeja tehneillä oppilailla puolestaan on suurin hajonta arviointien poikkeavuudessa. Tätä voidaan selittää rohkeampana suhtautumisena itsensä arvioimiseen, jonka suuremmat arviointimäärät ovat mahdollistaneet. Enemmän arviointeja tehneet oppilaat eivät koe enää yhtä voimakasta tarvetta seurata opettajan arviointia. Toisaalta myös opettajat voivat pyrkiä korjaamaan oppilaan itsearviointeja lähemmäksi omaa käsitystään antamalla itsearvioinneista voimakkaammin poikkeavia arviointeja enemmän itsearviointeja tehneille oppilaille. Visuaalisten tuloksien perusteella tässä tapauksessa ei ole mahdollista tehdä tarkempia havaintoja hypoteesiin liittyen. Kokonaisuudessaan laskennallisista ja visualisoiduista tuloksista ei löydetä hypoteesia tukevia merkkejä, joten hypoteesi tulee hylätä. Aineiston perusteella itsearviointien suurempi määrä ei vaikuta itsearviointien tarkkuuteen.

Kolmannen hypoteesin mukaan oppilaan korkeampi vuosiluokka tarkoittaisi tarkempaa itsearviointikykyä. Hypoteesin testaamiseksi oppilaat erotettiin vuosiluokkakohtaisiin segmentteihin, ja laskettiin kullekin vuosiluokalle oma korrelaatiokerroin. Vuosiluokkakohtaiset p-arvot olivat alhaisia, joten saatuja korrelaatiokertoimia voidaan pitää tilastollisesti merkitsevinä. Kuvasta 4.3 nähdään, että vuosiluokan ja korrelaatiokertoimen välillä on lievä positiivinen trendi, jonka mukaan edistyneemmät oppilaat ovat tarkempia itsearvioijia. Verrattaessa neljän ensimmäisen vuosiluokan korrelaatiokertoimen keskiarvoa 0,776 ja neljän viimeisen vuosiluokan korrelaatiokertoimen keskiarvoa 0,887, voidaan huomata selkeä itsearvioinnin tarkkuuden kehitys myöhemmällä vuosiluokalla. Kuten kuvasta 4.3 huomataan, korrelaatiokerroin ei kuitenkaan kasva monotonisesti, joten hypoteesia ei voida täysin hyväksyä sellaisenaan. Aineiston perusteella edistyneemmät opinnot korreloivat lievästi itsearviointien tarkkuuden kanssa, eli korkeammalla vuosiluokalla oleva oppilas osaa todennäköisesti itsearvioida oppimistaan paremmin kuin alhaisemmalla vuosiluokalla oleva oppilas.

Neljännän hypoteesin mukaan parhaiten koulussa menestyvät oppilaat aliarvioivat itsearvioinnissa suoriutumistaan ja heikoiten suoriutuvat oppilaat yliarvioivat itsearvioinnissa suoriutumistaan. Hypoteesia testattiin rajaamalla aineistosta opettaja-arviointien perusteella parhaiten ja heikoiten arvioidut 45 oppilasta omiksi aineistoikseen ja vertaamalla niitä. Vertailussa käytettiin sekä itse- ja opettaja-arviointien keskiarvoa että sitä, kuinka moni kummastakin ryhmästä arvioi suoriutumistaan paremmaksi tai huonommaksi kuin opettaja. Vertailun tulokset on esitetty taulukossa 4.2 ja kuvassa 4.4. Kuvasta 4.4 voidaan

nähdä, että aineiston heikoiten arvioidut oppilaat ovat selvästi yliarvioineet omia taitojaan. Parhaiten arvioidut oppilaat ovat lievästi, mutta kuitenkin tilastollisesti merkitsevästi, aliarvioineet omaa suoriutumistaan.

Taulukon 4.2 pohjalta voidaan laskea, että heikoiten arvioidut oppilaat ovat arvioineet suoriutumistaan 10,4 prosenttia paremmaksi kuin opettajat. Tilanne näyttäytyy samanlaisena myös määrällisenä vertailuna, sillä 53 % heikoiten arvioiduista oppilaista oli yliarvioinut omaa suoriutumistaan verrattuna opettaja-arvioihin. Parhaiten arvioitujen oppilaiden osalta arviointien eroavaisuudet ovat huomattavasti pienempiä, sillä keskiarvollisesti itsearviointit olivat vain 2,1 % heikompia kuin opettaja-arviointit. Omaa suoritustaan aliarvioineita oli 16 % parhaiten arvioiduista oppilaista. Tulokset vastaavat Falchikovin & Boudin (1989), Karnilowiczin (2012), ja Boudin et al. (2013) löydöksiä. Edellä esitetyt ja lasketut tulokset puoltavat hypoteesia, joten sen hylkäämiseen ei ole perusteita. Aineiston perusteella heikoiten arvioidut oppilaat yliarvioivat suoriutumistaan ja parhaiten arvioidut oppilaat aliarvioivat suoriutumistaan.

5.2 Tutkimuksen arviointi

Tutkimuksessa selvitettiin data-analytiikan hyödyntämistä peruskoulun oppimisen arvioinnissa empiirisen data-aineiston perusteella. Osana tutkimusta luotiin ohjeet, joilla aineistosta on mahdollista tutkia korrelaatioita ja verrata niitä keskenään. Ohjeiden osalta on huomioitava, että tutkimuksessa esitetyt toimenpiteet on yksilöity käytetylle aineistolle, eikä niitä suoranaisesti voida hyödyntää tästä poikkeaviin aineistoihin. Ne tarjoavat kuitenkin ajatusmallin vastaaviin korrelaatiotutkimuksiin.

Keskeinen ongelma itsearviointien tarkkuuden arvioinnissa liittyy vertailukohteeseen. Yleisesti itsearviointikykyyn liittyvissä tutkimuksissa (e.g. Falchikov & Boud 1989; Lew et al. 2010; Karnilowicz 2012; Johansson 2013) verrataan oppilaan arviota opettajan arvioon. Tämä herättää kuitenkin kysymyksen, onko opettajan antama arvio absoluuttisen oikea? Opettajan tekemiin arviointeihin vaikuttaviksi tekijöiksi tunnistetaan kansainvälisesti niin arviointikulttuurin vaikutus, arvioitavan yleinen suoritustaso kuin myös arvioitavan sukupuoli, rotu ja sosiaalinen tausta (Newstead 2002). Suomen koulujärjestelmää käsittelevissä tutkimuksissa esitetään myös kritiikkiä opettaja-arvioiteja kohtaan. Esimerkiksi Hirvonen (2012) huomauttaa arvioitavan sukupuolella olevan vaikutusta opettajan tekemiin arviointeihin. Ouakrim-Soivio (2013) puolestaan esittää tutkimuksessaan, että kansallinen seuranta-arviointi tuotti enemmän vaihtelua arvosanoihin kuin opettaja-arviointi, jossa oli tapana pyöristää arvosanoja kohti yleistä keskiarvoa. Opettaja-arvioiteja ei siis voida pitää totuutena, vaan parhaana saatavilla olevana vertailukohteena.

Tutkimuksessa testattujen hypoteesien osalta on otettava huomioon teoria-aineistoon liittyvät heikkoudet. Vaikka teoriaa löydettiinkin kansainvälisestä tutkimuksesta runsaasti, oli sen kohderyhmä pääosin lukio- tai korkeakoulutason opiskelijoita. Kokonaisuudes-

saan voidaan todeta, että tiedonhakuprosessin aikana löydettiin vain vähän tämän tutkimuksen fokukseen osuvaa tieteellistä tekstiä, jossa olisi tutkittu nimenomaan 6–16-vuotiaita oppilaita. Tämä on tuloksissa huomioitava, sillä esimerkiksi Butler (1990) esittää, että nuorella iällä itsearviointitarkkuus kasvaa merkittävästi lyhyellä aikavälillä. Toisaalta Sung et al. (2005) huomauttavat, että oppilaiden itsearviointikyky riippuu voimakkaasti siitä, opetetaanko itsearviointia tietoisesti. Vaikka Suomen perusopetuksen opetussuunnitelmassa veloitetaan itsearvioinnin opetukseen (Opetushallitus 2014b, ss. 47–49), on syytä nostaa esiin kysymys, opetetaanko sitä todella varhaisilta vuosiluokilta lähtien?

Tutkimuksen tulosten osalta on huomioitava aineistoon liittyvä kritiikki. Vaikka kyseessä onkin todellinen lukuvuoden ajalta kerätty aineisto suomalaisista peruskouluista, ei sitä laajuudeltaan voida yleistää koko Suomea kattavaksi, sillä aineisto piti sisällään arviointeja vain kahdesta peruskoulusta. Suurempi määrä arviointeja olisi myös parantanut tutkimuksen tulosten tarkkuutta, vaikka nykyiselläkin otannalla onnistuttiin saavuttamaan selkeää tilastollista merkitysvyyttä. Kuten jo aineiston esittelyssä todettiin, ei aineisto kuvannut tasapuolisesti peruskoulun vuosiluokkia, sillä esimerkiksi viidenneltä vuosiluokalta ei ollut vertailuun kelpuutettavia arviointeja saatavilla. Aineisto vaikutti lievästi myös teoriaosuudessa luotuihin hypoteeseihin. Teoriaosuudessa ei voitu esittää hypoteeseja esimerkiksi oppimistulosten muutoksiin liittyen, sillä käytettävissä ei ollut vertailukohtaa ajalta ennen digitaalisen arviointiohjelmiston käyttöönottoa. Täten hypoteesit eivät ole suoranaisesti kirjallisuudessa eniten tunnistettuja itsearviointiin liittyviä näkemyksiä, vaan näkemyksistä on suodatettu jäljelle vain tutkimuksellisesti mahdolliset hypoteesit.

Tuloksia luettaessa on myös syytä huomioida, että oppilaat ja opettajat pystyivät molemmat näkemään toistensa antamat arvioinnit ennen oman arviointinsa kirjaamista. Varsinkin alempien vuosiluokkien kohdalla on huomioitava, että osapuolet eivät välttämättä halua esittää kovin eriäviä mielipiteitä. Oppilailla tämä voi liittyä epävarmuuteen omista arviointitaidoistaan ja opettajilla siihen, ettei oppilaan haluta tulevan epävarmaksi tai vähemmän motivoituneeksi liian ankaralta tuntuvan arvioinnin johdosta. Erityisesti toisen hypoteesin, jonka mukaan arviointimäärät parantavat itsearviointikykyä, kohdalla on syytä ottaa huomioon, että vähän itsearviointeja tekevät oppilaat voivat pelätä eriävän arvioinnin esittämistä ja tämän johdosta asettaa oman arvionsa mahdollisimman lähelle opettajan arviota.

6. PÄÄTELMÄT

Kuudennessa luvussa tiivistetään tutkimuksessa esitetyt tulokset ja niiden merkitys. Tämän lisäksi luvussa esitetään jatkotutkimusehdotuksia.

6.1 Tulosten yhteenveto

Teoriaa varten tehdyssä selvitystyössä huomattiin selkeä tutkimusaukko liittyen digitaaliin oppimisen itse- ja opettaja-arviointeihin suomalaisissa peruskouluissa. Vaikka kansainvälistä tutkimusta löydettiin jonkin verran, keskittyi se pääosin korkeakoulutason arviointeihin. Taustateorian perusteella voitiin tunnistaa oppilaiden itsearviointikykyyn vaikuttavia tekijöitä, mutta näiden peilautuvuudesta tutkimuksen kontekstiin ei voitu olla varmoja. Käsitellyn teorian pohjalta luotiin suomalaisista peruskouluista kerättyä aineistoa vasten testattavat hypoteesit.

Tutkimuksen perusteella saadut tulokset noudattivat yhtä lukuun ottamatta melko hyvin kansainvälisten tutkimustulosten perusteella luotuja hypoteeseja. Itse- ja opettaja-arviointien väliltä löydettiin voimakasta positiivista korrelaatiota. Tämä tarkoittaa, että itse- ja opettaja-arviointeja voidaan lähtökohtaisesti käyttää toistensa ennustamiseen, sillä niiden välillä on huomattavaa riippuvuutta. Tulos on yhteneväinen useiden kansainvälisten tutkimusten kanssa (e.g. Falchikov & Boud 1989; Dochy et al. 1999; Karnilowicz 2012; Johansson 2013), vaikka tutkimuksen konteksti eroaakin merkittävästi. Arviointimäärien ja itsearviointitarkkuuden väliltä ei löytynyt teoriaa tukevaa korrelaatiota, vaikka hypoteesin pohjalta korrelaatiota odotettiin. Ennen väitteen kiistämistä on kuitenkin huomioitava, että oppilaat ja opettajat pystyivät näkemään toisen osapuolen tekemät arvioinnit, joten lähellä toisiaan olevat arviointitulokset vähäisillä arviointimäärillä saattoivat johtua muista syistä kuin itsearviointitaidosta. Saavutettu tulos on eriävä Fitzgeraldin (2003) ja Boudin et al. (2013) tulosten kanssa.

Korkeamman vuosiluokan ja itsearviointikykyyn väliltä löydettiin lievää positiivista korrelaatiota. Vaikka tulokset eivät olleetkaan suoraviivaiset, oli ero huomattava verrattaessa peruskoulun neljää ensimmäistä vuosiluokkaa neljään viimeiseen vuosiluokkaan. Tulos tukee esimerkiksi Falchikovin ja Boudin (1989) löydöksiä opintojen etenemisen vaikutuksista arviointikykyyn. Oppilaan taitotasolla todettiin myös olevan vaikutusta tämän tekemiin itsearviointeihin, sillä heikoimmin arvioitujen oppilaiden huomattiin yliarvioiden omaa taitotasoaan ja parhaiten arvioitujen puolestaan aliarvioivan suoritustaan itsearviointeissa verrattuna opettaja-arviointeihin. Aineistosta laskettu tulos oli yhteneväinen Falchikovin ja Boudin (1989), Karnilowiczin (2012) sekä Boudin et al. (2013) tutkimustulosten kanssa.

Kokonaisuudessaan kaikki lasketut korrelaatiokertoimet olivat p-arvoiltaan erittäin alhaisia ja täten tilastollisesti merkitseviä. Korrelaatiokertoimet itsessään osoittivat kaikilla oppilasryhmillä erittäin voimakasta positiivista korrelaatioita, josta voidaan päätellä arviointien välillä olleen selkeä yhteys ja oppilaiden itsearviointitarkkuuden olleen digitaalisessa arviointiohjelmistossa kokonaisuudessaan hyvä. Saavutetut tulokset tarjoavatkin tutkimustietoa digitaalisen arviointiohjelmiston vaikutuksista suomalaisten peruskoulujen oppilaiden itsearviointikykyihin ja avaavat uusia kiinnostavia jatkotutkimuskohteita.

6.2 Jatkotutkimus ja rajoitteet

Tutkimuksessa saatujen tulosten ja niistä tehdyn arvion pohjalta voidaan tunnistaa aihetta syventävien ja tuloksia tarkentavien jatkotutkimusten mahdollisia aihepiirejä. Keskeinen jatkotutkimustarve liittyy tutkimuksen tuloksena tunnistettuihin korrelaatioihin. Korrelaatiot itsessään eivät vielä tarjoa merkittäviä työkaluja arvioiden pohjalta tehtäviin päätelmiin, mutta niiden perusteella voidaan luoda ennustavia malleja, jotka hyödyntävät arvioitavan oppilaan taustatietoja. Ennustavien mallien avulla mahdollistettaisiin poikkeuksien havaitseminen aikaisessa vaiheessa, joka antaisi arvioijille mahdollisuuden reagoida tilanteisiin ennakoivasti. Esimerkkitapauksena hyvin opinnoissaan menestyneen oppilaan äkillinen arvosanojen romahtaminen voi kieliä kiusaamisesta tai ongelmista kotona. Tällaisten tilanteiden tunnistamisella ja niihin reagoimisella voidaan pienentää syrjäytymisen riskiä ja tarjota entistä yksilöllisempää oppimisen ohjausta suurissakin kouluissa.

Jatkotutkimusehdotuksissa on huomioitava myös tutkimuksen tuloksiin liittyneet rajoitukset. Useampien koulujen ja arviointien sisällyttäminen tutkimukseen mahdollistaisi koulu-, paikkakunta- ja oppiainekohtaiset vertailut tunnistettujen korrelaatioiden osalta. Tehtyjä itse- ja opettaja-arviointeja olisi syytä verrata myös sillä tasolla, etteivät osapuolet tietäisi toistensa antamia arviointeja, jotta arvioinnit pohjautuisivat puhtaasti arvioinnin tekijän omaan käsitykseen. Samantapaisen jatkotutkimusaiheen avaa opettaja-arvioinnin korvaaminen vertailukohtena esimerkiksi objektiivisilla arviointikohteilla tai arviointikohteilla, joissa arvioijan vaikutus on minimaalinen.

LÄHTEET

Alshammari, I., Aldhafiri, M. & Al-Shammari, Z. (2013). A Meta-Analysis of Educational Data Mining on Improvements in Learning Outcomes, *College Student Journal*, Vol. 47(2), pp. 326–333.

Anttila, P. (2013). Taitojen ja luovien alojen arvioinnin kysymyksiä, kokoelmassa Räsänen, A. Oppimisen arvioinnin kontekstit ja käytännöt, Opetushallitus, Juvenes Print – Suomen Yliopistopaino Oy, ss. 89–118.

Atjonen, P. (2007). Hyvä, paha arviointi, Kustannusosakeyhtiö Tammi, Gummerus Kirjapaino Oy, Jyväskylä, 256 s.

Baker, R. (2011). Data mining for education, *International Encyclopedia of Education* (3rd edition). Oxford, UK: Elsevier.

Baker, R. & Yacef, K. (2009). The State of Educational Data Mining in 2009: A Re-view and Future Visions, *Journal of Educational Data Mining*, Article 1. Vol 1. No 1.

Boud, D. (1995). Enhancing learning through self assessment, 1. publ. ed. Kogan Page, London, 247 p.

Boud, D., Lawson, R. & Thompson, D.G. (2013). Does student engagement in self-assessment calibrate their judgement over time? *Assessment & Evaluation in Higher Education*, Vol. 38(8), pp. 941–956.

Butler, R. (1990). The Effects of Mastery and Competitive Conditions on Self-Assessment at Different Ages, *Child Development*, Vol. 61(1), pp. 201–210.

Chandler, D. & Rod, M. (2016). Data analytics, *A Dictionary of Social Media*. Oxford University Press.

Chatti, M.A., Dyckhoff, A.L., Schroeder, U. & Thüs, H. (2013). A reference model for learning analytics, *International Journal of Technology Enhanced Learning*, Vol. 4(5–6), pp. 318–331.

Colwell, D. & Gillett, J. (1982). Spearman versus Kendall. *The Mathematical Gazette*. Vol. 66, No. 438 (Dec., 1982), pp. 307-309.

Dochy, F., Segers, M. & Sluijsmans, D. (1999). The use of self-, peer and co-assessment in higher education: A review, *Studies in Higher Education*, Vol. 24(3), pp. 331–350.

Drosg, M. (2007). *Dealing with Uncertainties: A Guide to Error Analysis*, 1. ed. Springer-Verlag, Berlin, Heidelberg, 190 p.

- Elias, T. (2011). Learning Analytics: Definitions, Processes and Potential (Report), Available: <http://learninganalytics.net/LearningAnalyticsDefinitionsProcessesPotential.pdf> (Retrieved 12.7.2017).
- Eva, K.W., Cunnington, J.P.W., Reiter, H.I., Keane, D.R. & Norman, G.R. (2004). How Can I Know What I Don't Know? Poor Self Assessment in a Well-Defined Domain, *Advances in Health Sciences Education*, Vol. 9(3), pp. 211–224.
- Falchikov, N. & Boud, D. (1989). Student Self-Assessment in Higher Education: A Meta-Analysis, *Review of Educational Research*, Vol. 59(4), pp. 395–430.
- Fitzgerald, J.T., White, C.B. & Gruppen, L.D. (2003). A longitudinal study of self-assessment accuracy, *Medical education*, Vol. 37(7), pp. 645–649.
- Hatami, Ali. (2015). The effect of collaborative learning and self-assessment on self-regulation, *Educational Research and Reviews*, Vol. 10(15), pp. 2164–2167.
- Hauke, J. & Kossowski, T. (2011). Comparison of Values of Pearson's and Spearman's Correlation Coefficients on the Same Sets of Data, *Quaestiones Geographicae*, Vol. 30(2), pp. 87–93.
- Heikkilä, T. (2014). Muuttujien väliset riippuvuudet, kokoelmassa: Heikkilä, T. Tilastollinen tutkimus, verkkomateriaali, Edita Publishing Oy, Helsinki, 41 s.
- Heikkilä, T. (2005). Tilastollinen tutkimus, 5.-6. painos. Business Edita, Helsinki, 328 s.
- Hirvonen, K. (2012). Onko laskutaito laskussa? Matematiikan oppimistulokset peruskoulun päättövaiheessa 2011, Opetushallitus, Juvenes Print - Tampereen Yliopistopaino Oy, Tampere, 133 s.
- Holopainen, M. & Pulkkinen, P. (2002). Tilastolliset menetelmät, 1. painos. WSOY, Helsinki, 338 s.
- Hung, C., Huang, I. & Hwang, G. (2014). Effects of digital game-based learning on students' self-efficacy, motivation, anxiety, and achievements in learning mathematics, *Journal of Computers in Education*, Vol. 1(2), pp. 151–166.
- Hunter, J.D. (2007). Matplotlib: A 2D Graphics Environment, *Computing in Science & Engineering*, Vol. 9(3), pp. 90–95.
- Jalonen, H., Helander, N. & Mäkelä, L. (2017). D-pilleriä etsimässä, DEEVA-tutkimus-hanke, Saatavissa: <http://deeva.fi/uncategorized/d-pilleria-etsimassa/> (Haettu 12.7.2017).
- Johansson, S. (2013). The relationship between students' self-assessed reading skills and other measures of achievement, *Large-scale Assessments in Education*, Vol. 1(1), pp. 1–17.

Karnilowicz, W. (2012). A comparison of self-assessment and tutor assessment of undergraduate psychology students, *Social Behavior and Personality: an international journal*, Vol. 40(4), pp. 591–604.

Ketokivi, M. (2009). *Tilastollinen päättely ja tieteellinen argumentointi*, Gaudeamus Helsinki University Press, 251 s.

Laitinen, J. (2015). Koulu 2.0 – Teknologia tulee vauhdilla luokkiin, ovatko koulut siihen valmiita? *Etelä-Suomen Sanomat*, Saatavissa: <http://www.ess.fi/uutiset/kotimaa/2015/08/30/koulu-2.0---teknologia-tulee-vauhdilla-luokkiin-ovatko-koulut-siihen-valmiita> (Haettu 12.7.2017).

Lew, M.D.N., Alwis, W.A.M. & Schmidt, H.G. (2010). Accuracy of students' self-assessment and their beliefs about its utility, *Assessment & Evaluation in Higher Education*, Vol. 35(2), pp. 135–156.

Lindblom-Ylänne, S., Pihlajamäki, H. & Kotkas, T. (2006). Self-, peer- and teacher-assessment of student essays, *Active Learning in Higher Education*, Vol. 7(1), pp. 51–62.

Maguire, S., Evans, S.E. & Dyas, L. (2001). Approaches to Learning: a Study of First-year Geography Undergraduates, *Journal of Geography in Higher Education*, Vol. 25(1), pp. 95–107.

McClarty, K., Orr, A., Frey, P., Dolan, R., Vassileva, V. & McVay, A. (2012). *A literature review of gaming in education*, Research report, NJ: Supper Saddle River: Pearson.

Mellin, I. (2006). Lineaarinen regressioanalyysi, kokoelmassa: Mellin, I. *Tilastolliset menetelmät*, Teknillinen korkeakoulu, Matematiikan laboratorio, ss. 231–433.

Newstead, S. (2002). Examining the Examiners: Why are We So Bad at Assessing Students? *Psychology Learning & Teaching*, Vol. 2(2), pp. 70–75.

Nironen, S. (2016). Digiloikka taaksepäin? Koulujen digitalisoinnin työkalut saavat opettajilta huutia Helsingissä, *Yle*, Saatavissa: <https://yle.fi/uutiset/3-9176889> (Haettu 12.7.2017).

Opetushallitus. (2014a). *Oppimisen arviointi*, kokoelmassa: Opetushallitus. *Perusopetuksen opetussuunnitelman perusteet. Määräykset ja ohjeet 2014:96*, 4. painos., Next Print Oy, Helsinki, ss. 47–60.

Opetushallitus (2014b). *Perusopetuksen opetussuunnitelman perusteet 2014*. Opetushallitus määräykset ja ohjeet 2014:96. Next Print Oy, Helsinki, 473 s.

Ouakrim-Soivio, N. (2013). Toimivatko päättöarvioinnin kriteerit? Oppilaiden saamat arvosanat ja Opetushallituksen oppimistulosten seuranta-arviointi koulujen välisten osaamiserojen mittareina. Opetushallitus Raportit ja selvitykset 2013:9, Opetushallitus, Juvenes Print - Suomen Yliopistopaino Oy, Tampere, 282 s.

Papamitsiou, Z. & Economides, A.A. (2014). Learning Analytics and Educational Data Mining in Practice: A Systematic Literature Review of Empirical Evidence, *Educational Technology & Society*, Vol. 17(4), pp. 49–64.

Paris, S.G., Lawton, T.A., Turner, J.C. & Roth, J.L. (1991). A Developmental Perspective on Standardized Achievement Testing, *Educational Researcher*, Vol. 20(5), pp. 12–20.

Perusopetuslaki, (1998). 22 § Oppilaan arviointi (21.8.1998/628).

Python Software Foundation (2017a) Python is powerful... and fast; plays well with others; runs everywhere; is friendly & easy to learn; is Open. Available: <https://www.python.org/about/> (Retrieved 14.8.2017)

Python Software Foundation. (2017b) pandas 0.20.2 Python documentation. Available: <https://pypi.python.org/pypi/pandas/> (Retrieved 12.7.2017).

Romero, C. & Ventura, S. (2013). Data mining in education, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, Vol. 3(1), pp. 12–27.

Ross, J.A., Rolheiser, C. & Hogaboam-Gray, A. (1999). Effects of self-evaluation training on narrative writing, *Assessing Writing*, Vol. 6(1), pp. 107–132.

Ruohonen, K. (2011). Tilastomatemiikka, Tampereen teknillinen yliopisto, Matematiikan laitos. Saatavissa: [https://tutcris.tut.fi/portal/en/publications/tilastomatemiikka\(e5a82a7a-a34d-4c12-8ce7-49571107a553\).html](https://tutcris.tut.fi/portal/en/publications/tilastomatemiikka(e5a82a7a-a34d-4c12-8ce7-49571107a553).html) (Haettu 12.7.2017),

Rust, C., Price, M. & O'Donovan, B. (2003). Improving Students' Learning by Developing their Understanding of Assessment Criteria and Processes, *Assessment & Evaluation in Higher Education*, Vol. 28(2), pp. 147-164.

Räisänen, A. (2013). Oppimisen arvioinnin kontekstit ja käytännöt, Opetushallitus, Juvenes Print - Suomen Yliopistopaino Oy, ss. 7–10.

Saari, A. & Säntti, J. (2016). Kouluja ei pidä digitalisoida väkisin – opettajien omaa harjontaa pitäisi kunnioittaa, *Helsingin Sanomat*, Saatavissa: <http://www.hs.fi/paakirjoitukset/art-2000002919633.html> (Haettu 12.7.2017).

Salumäki, T. (2016). Koulua digitalisoidaan – moni opettaja tuskailee puuttuvia laitteita, *Yle*, Saatavissa: <https://yle.fi/uutiset/3-9198973> (Haettu 12.7.2017).

Siemens, G. (2012). Learning analytics, Proceedings of the 2nd International Conference on learning analytics and knowledge, ACM, pp. 4–8.

Siemens, G. & Baker, R. (2012). Learning analytics and educational data mining: towards communication and collaboration, Proceedings of the 2nd International Conference on learning analytics and knowledge, ACM, pp. 252–254.

Sung, Y., Chang, K., Chiou, S. & Hou, H. (2005). The design and application of a web-based self- and peer-assessment system, Computers & Education, Vol. 45(2), pp. 187–202.

Taras, M. (2010). Student self-assessment: processes and consequences, Teaching in Higher Education, Vol. 15(2), pp. 199–209.

Valtioneuvoston kanslia (2016). Toimintasuunnitelma strategisen hallitusohjelman kärkihankkeiden ja reformien toimeenpanemiseksi 2015-2019 - päivitys 2016, Hallituksen julkaisusarja 2/2016, 90 s.

Warren, S.J., Dondlinger, M.J. & Barab, S.A. (2008). A MUVE Towards PBL Writing, Journal of Research on Technology in Education, Vol. 41(1), pp. 113.

Waskom, M. (2015). Seaborn: statistical data visualization, Seaborn, Available: <https://seaborn.pydata.org/index.html> (Retrieved 12.7.2017).

Xiao, C., Ye, J., Esteves, R.M. & Rong, C. (2016). Using Spearman's correlation coefficients for exploratory data analysis on big dataset, Concurrency and Computation: Practice and Experience, Vol. 28(14), pp. 3866–3878.

LIITE A: DATAN_PUTSAUS.PY

```
1 # Datan ylimääräisistä ja käyttökelvottomistariveistä puhdistava moduuli.
2 # Return: final_data = Puhdistettu datataulu, jossa ei turhia rivejä
3 # comb_asses_dict = kirjasto, jossa avaimena arviointikohteen yksilöivä
4 # tunniste ja arvona lista annetuista arvioinneista.
5
6 import pandas as pd
7 from collections import defaultdict
8
9 def clean_df():
10     data = pd.read_csv('P:/arvioinnit.csv')
11
12     # Luodaan arvioinnin (oppilaan ja arviointikohteen) yksilöivä sarake
13     data['skill_user_id'] = data['skill_id'].apply(str) + data['user_id'].apply(str)
14
15     comb_asses_dict = defaultdict(list)
16     for index, row in data.iterrows():
17         if row['is_latest'] == 't':
18             if row['skill_user_id'] not in comb_asses_dict:
19                 # Aineistossa ei ole mahdollista antaa 0-arvoa, joten voidaan käyttää
20                 # sitä "puuttuvana arviona"
21                 comb_asses_dict[row['skill_user_id']] = [0, 0]
22             if row['ratings_is_self_evaluation'] == 'f':
23                 comb_asses_dict[row['skill_user_id']][0] = row['ratings_value']
24             else:
25                 comb_asses_dict[row['skill_user_id']][1] = row['ratings_value']
26     del_list = []
27     # Poistetaan arvioinnit, joissa ei ole sekä itse- että opettaja-arviota
28     for key, value in comb_asses_dict.items():
29         if (value[0] == 0) or (value[1] == 0):
30             del_list.append(key)
31     for key in del_list:
32         del comb_asses_dict[key]
33     # Lopulliseen datatauluun vain rivit, joihin on arviointeja
34     final_skill_user_ids = []
35     for key, value in comb_asses_dict.items():
36         final_skill_user_ids.append(key)
37     final_data = data.loc[(data['skill_user_id'].isin(final_skill_user_ids) &
38                          (data['is_latest'] == 't'))]
39     return final_data, comb_asses_dict
```

LIITE B: YLEINEN_KORRELAATIO.PY

```
1 # Kaikkien hyväksyttävien rivien korrelaation ja p-arvon laskeva ja tulokset
2 # visualisoiva moduuli.
3 import pandas as pd
4 from scipy import stats
5 import seaborn as sns
6 import datan_putsaus
7
8 data, asses_dict = datan_putsaus.clean_df()
9
10 final_self_asses = []
11 final_teacher_asses = []
12 final_keys = []
13 for key, value in asses_dict.items():
14     final_self_asses.append(value[1])
15     final_teacher_asses.append(value[0])
16     final_keys.append(key)
17 # Luodaan datataulu
18 corr_df = pd.DataFrame({'self' : final_self_asses,
19                        'teacher' : final_teacher_asses}, index=final_keys)
20 # Korrelaatio ja merkitsevyys
21 results = stats.spearmannr(corr_df['self'], corr_df['teacher'])
22 print("Spearmanin korrelaatiokerroin:", results[0])
23 print("Korrelaatiokertoimen p-arvo:", results[1])
24
25 # Suoritetaan tulosten visualisaatio
26 # Luodaan kirjasto, jossa avaimena itsearviointi ja opettaja-arvio -pari ja arvona
27 # esiintymiskertojen määrä
28 count_comb_dict = {}
29 for value in asses_dict.values():
30     if tuple(value) in count_comb_dict:
31         count_comb_dict[tuple(value)] += 1
32     else:
33         count_comb_dict[tuple(value)] = 1
34 self_ass = []
35 teach_ass = []
36 amount = []
37 # Siirretään arvot listaan ja luodaan pandas taulu
38 for key, value in count_comb_dict.items():
39     self_ass.append(list(key)[1])
40     teach_ass.append(list(key)[0])
41     amount.append(value)
42 visual_df = pd.DataFrame({'Itsearviointi' : self_ass,
43                          'Opettaja-arviointi' : teach_ass,
44                          'Amount' : amount})
45 result = visual_df.pivot_table(index='Itsearviointi', columns='Opettaja-arviointi',
46                                values='Amount')
47 fig = sns.heatmap(result, annot=True, fmt='g')
48 fig.invert_yaxis()
```

LIITE C: ARVIOINTIMAARIEN_VAIKUTUS_KORRELAATIOON.PY

```
1 # Metodi jakaa oppilaat kolmeen samankokoiseen ryhmään, laskee ryhmille omat
2 # korrelaatiokertoimet ja p-arvot ja lopuksi visualisoi tulokset.
3 import pandas as pd
4 import seaborn as sns
5 from scipy import stats
6
7 import datan_putsaus
8
9 data, ass_dict = datan_putsaus.clean_df()
10 # Luodaan kolmelle käyttäjätyypille (ahkeruuden mukaan) oma lista, johon lisätään
11 sopivat opiskelijanumerot
12 lazy_list = []
13 ave_list = []
14 busy_list = []
15 for key, value in data['user_id'].value_counts().items():
16     if value <= 6:
17         lazy_list.append(key)
18     elif value >= 13:
19         busy_list.append(key)
20     else:
21         ave_list.append(key)
22
23 # Lasketaan korrelaatiot kullekin käyttäjäryhmälle
24 user_lists = [lazy_list, ave_list, busy_list]
25 for lis in user_lists:
26     self_asses = []
27     teacher_asses = []
28     keys = []
29     for key, value in ass_dict.items():
30         if int(key[5:9]) in lis:
31             self_asses.append(value[1])
32             teacher_asses.append(value[0])
33             keys.append(key)
34     corr_df = pd.DataFrame({'self' : self_asses,
35                             'teacher' : teacher_asses}, index=keys)
36     results = stats.spearmanr(self_asses, teacher_asses)
37     print('Korrelaatiokerroin', results[0], 'P-arvo', results[1])
38
39 # Visualisointi, jossa lasketaan kullekin ryhmälle itsearviointin ja opettaja-
40 arviointin poikkeaman mukainen prosenttiosuus
41 class_list = []
42 diff_list = []
43 percentage_list = []
44 i = 0
45 # Täytetään listat poikkeamille, arviointimäärille ja poikkeamille
46 for lis in user_lists:
47     count_comb_dict = {}
48     for dict_key, dict_value in ass_dict.items():
49         if int(dict_key[5:9]) in lis:
50             if tuple(dict_value) in count_comb_dict:
51                 count_comb_dict[tuple(dict_value)] += 1
52             else:
53                 count_comb_dict[tuple(dict_value)] = 1
54     diff_count = {}
55     for key, value in count_comb_dict.items():
56         diff = list(key)[1]- list(key)[0]
```

```

57     if diff in diff_count:
58         diff_count[diff] += value
59     else:
60         diff_count[diff] = value
61 for key, value in diff_count.items():
62     diff_list.append(key)
63     if i == 0:
64         class_list.append('0-6')
65         percentage_list.append(float("{0:.3f}".format(value / 221)))
66     elif i == 1:
67         class_list.append('7-12')
68         percentage_list.append(float("{0:.3f}".format(value / 368)))
69     elif i == 2:
70         class_list.append('>12')
71         percentage_list.append(float("{0:.3f}".format(value / 443)))
72     i += 1
73 # Visualisoidaan tulokset luotujen listojen avulla
74 visual_df = pd.DataFrame({'Poikkeama opettaja-arviosta' : diff_list,
75                          'Arviointien määrä' : class_list,
76                          'Määrä' : percentage_list})
77 result = visual_df.pivot_table(index='Poikkeama opettaja-arviosta', columns='Ar-
78 viointien määrä', values='Määrä')
79 fig = sns.heatmap(result, annot=True, fmt='g', linewidth=.1, cbar_kws={"orienta-
80 tion": "horizontal"})
81 fig.invert_yaxis()

```


LIITE D: VUOSILUOKAN_VAIKUTUS_KORRELAATIOON.PY

```
1 # Moduuli laskee luokkakohtaiset korrelaatiokertoimet ja p-arvot sekä visualisoi
2 # tulokset pylväsdiagrammiin
3
4 import pandas as pd
5 from scipy import stats
6 import seaborn as sns
7
8 import datan_putsaus
9
10 data, ass_dict = datan_putsaus.clean_df()
11
12 # Luodaan kirjasto, jossa avaimena on vuosiluokka ja arvona lista luokkaan
13 # kuuluvista skill_user_id:eistä
14 class_dict = {1: [], 2: [], 3: [], 4: [],
15              6: [], 7: [], 8: [], 9: []}
16 for index, row in data.iterrows():
17     if row['year_name'] != 5:
18         class_dict[row['year_name']].append(row['skill_user_id'])
19 class_corrs = []
20 for class_number, class_list in class_dict.items():
21     self_ass = []
22     teacher_ass = []
23     keys = []
24     for ass_key, ass_value in ass_dict.items():
25         if ass_key in class_list:
26             self_ass.append(ass_value[1])
27             teacher_ass.append(ass_value[0])
28             keys.append(ass_key)
29     corr_df = pd.DataFrame({'self' : self_ass,
30                          'teacher' : teacher_ass}, index=keys)
31     print(stats.spearmanr(corr_df['self'], corr_df['teacher']))
32     # Lisätään tulos listaan visualisaatiota varten
33     class_corrs.append(stats.spearmanr(self_ass, teacher_ass)[0])
34
35 # Visualisoidaan tulokset pylväsdiagrammilla
36 sns.set(style='white')
37 class_no = ['1', '2', '3', '4', '6', '7', '8', '9']
38 ax = sns.barplot(class_no, class_corrs, palette=sns.cubehelix_palette(8))
39 sns.despine()
40 ax.set(xlabel='Vuosiluokka', ylabel='Korrelaatiokerroin')
```

LIITE E: SUORIUTUMISEN_VAIKUTUS_ITSEARVIOINTIIN.PY

```
1 # Moduuli jakaa oppilaat parhaaseen ja heikoimpaan 20 % otoksesta, jonka jälkeen
2 # vertaa itse- ja opettaja-arviointien keskiarvoja oppilaskohtaisesti ja visualisoi
3 # tulokset
4 import pandas as pd
5 import seaborn as sns
6
7 import datan_putsaus
8
9 data, ass_dict = datan_putsaus.clean_df()
10 teach_asses = data[data['ratings_is_self_evaluation'] == 'f']
11 self_asses = data[data['ratings_is_self_evaluation'] == 't']
12
13 # Luodaan kirjasto, jossa avaimena user_id ja arvona lista oppilaalle annettujen
14 # arviointien arvosanoista
15 id_ratinglist_dict = {}
16 for index, row in teach_asses.iterrows():
17     if row['user_id'] not in id_ratinglist_dict:
18         id_ratinglist_dict[row['user_id']] = [row['ratings_value']]
19     else:
20         id_ratinglist_dict[row['user_id']].append(row['ratings_value'])
21
22 # Korvataan lista arvioinneista listan keskiarvolla
23 id_teach_ave_dict = {}
24 for key, value in id_ratinglist_dict.items():
25     id_teach_ave_dict[key] = (sum(value) / len(value))
26
27 # 225 oppilasta, valitaan keskiarvoista 20 % parhaita ja huonointa
28 rank_no = 0
29 worst_ids = []
30 best_ids = []
31 for key in sorted(id_teach_ave_dict, key=id_teach_ave_dict.get):
32     if (rank_no < (len(id_teach_ave_dict) * 0.2)):
33         worst_ids.append(key)
34     elif (rank_no >= (len(id_teach_ave_dict) * 0.8)):
35         best_ids.append(key)
36     rank_no += 1
37 # 45 parhaimman ja huonoimman oppilaan id:t listassa
38
39 # Tarvitaan myös itsearviointien keskiarvot
40 id_self_ratinglist_dict = {}
41 for index, row in self_asses.iterrows():
42     if row['user_id'] not in id_self_ratinglist_dict:
43         id_self_ratinglist_dict[row['user_id']] = [row['ratings_value']]
44     else:
45         id_self_ratinglist_dict[row['user_id']].append(row['ratings_value'])
46 id_self_ave_dict = {}
47 for key, value in id_self_ratinglist_dict.items():
48     id_self_ave_dict[key] = (sum(value) / len(value))
49
50 # Lisätään keskiarvot sopivan käyttäjäryhmän listaan
51 best_teach_ass_ave = []
52 worst_teach_ass_ave = []
53 best_self_ass_ave = []
54 worst_self_ass_ave = []
55 for key in best_ids:
56     best_teach_ass_ave.append(id_teach_ave_dict[key])
```

```

57     best_self_ass_ave.append(id_self_ave_dict[key])
58 for key in worst_ids:
59     worst_teach_ass_ave.append(id_teach_ave_dict[key])
60     worst_self_ass_ave.append(id_self_ave_dict[key])
61
62 # Lasketaan keskiarvojen erotukset ja poikkeamien määrät
63 sum_worst = []
64 sum_best = []
65 worst_bigger = 0
66 worst_smaller = 0
67 best_bigger = 0
68 best_smaller = 0
69 for (w_t_a, w_s_a, b_t_a, b_s_a) in zip(worst_teach_ass_ave, worst_self_ass_ave,
70                                         best_teach_ass_ave, best_self_ass_ave):
71     sum_worst.append(w_t_a - w_s_a)
72     if (w_t_a > w_s_a):
73         worst_smaller += 1
74     elif (w_t_a < w_s_a):
75         worst_bigger += 1
76     if (b_t_a > b_s_a):
77         best_smaller += 1
78     elif (b_t_a < b_s_a):
79         best_bigger += 1
80     sum_best.append(b_t_a - b_s_a)
81 print("Parhaat opettajalla parempia", best_smaller, "ja huonompia:", best_bigger)
82 print("Opettajan arvio - oppilaan arvion keskiarvo parhaimmilla", sum(sum_best) /
83       len(sum_best))
84 print("Huonot opettajalla parempia", worst_smaller, "ja huonompia:", worst_bigger)
85 print("Opettajan arvio - oppilaan arvion keskiarvo heikoimmilla", sum(sum_worst)
86       / len(sum_worst))
87
88 # Visualisoidaan tulokset pylväsdiagrammilla
89 sns.set(style='white')
90 sns.set_palette(sns.cubehelix_palette(2))
91 df = pd.DataFrame([[3.053, 3.372], [4.907, 4.800]], index = ['Heikoimmat oppilaat',
92                 'Parhaimmat oppilaat'], columns=['Opettaja-arvio', 'Itsearvio'])
93 printable = df.plot(kind='bar', rot=0)
94 sns.despine()

```