

Moniselitteiset PISA-tulokset ja niiden ongelmalliset koulutuspoliittiset tulkinnat

OSMO KIVINEN JA JUHA HEDMAN

ABSTRACT
Ambiguous PISA results and their problematic interpretations in education policy

PISA has not been designed to assess how well students master the contents of the school curricula, yet opinion leaders in Finland keep reading PISA results as schools' report cards indicating the excellence of education policy in a country. Finland took top positions from the very beginning, which surprised the Finns themselves. The experienced PISA miracle was then kept alive, until the 2012 assessment's declining scores delivered a minor "PISA shock" in Finland. PISA 2015, however, proved that the decline was temporary, but it did not eliminate the worry about Finnish gender gap in literacy, claimed as widest among the OECD countries, entailing alleged boys' severe underachievement. Success in PISA became an integrated part of national identity and Finnish country brand. The paper problematizes the one-dimensional reasoning of top achievements in PISA tests as an indication of excellent results of education policy. PISA methodology has to certain extent been questioned internationally, but the paper examines further the relations between success of countries in the league tables and the test performance of participating students. The paper argues that countries' movements up and down the league tables have necessarily no connection with students' performance in PISA's literary tests. Every three years PISA produces masses of standardized numerical information, which unfortunately does not redeem the strict requirements of policy based on scientific facts.

Johdanto

Kolmen vuoden välein toistuvista PISA-arvioinneista (*Programme for International Student Assessment*) on tullut kansainvälinen mediatapahtuma ja kihelmöivä odotuksen hetki oman maansa PISA-rankin-geissa sijoittumista jännittäville hallituksille ja ope- tusalan virkamiehille. PISAn kaltaisten laajojen kansainvälisten arviointien saattamana kansallis- valtioiden koulutuspolitiikka on enenevässä määrin OECD:n (*The Organisation for Economic Co-operation and Development*) ylikansallisessa ohjauksessa. Yhteistä PISA-tulosten vastaanotolle eri maissa on, että koulutuspoliittisia tulkintoja tai vaikkapa vaa- timuksia koulutus uudistuksista on voitu esittää ym- märtämättä paljoakaan siitä, mitä – jos mitään – PISA numerot kouluasioista tosiasiaassa kertovat.

PISAssa alusta asti mukana ollut Suomi on kokenut enimmäkseen menestystä, mutta myös yhden laske- vien pisteiden jakson ja siitä palautumisen.

Alun perin talouden asiantuntijaorganisaationa tunnetuksi tullut OECD on nyttemmin suuntautunut myös koulutuspolitiikkaan; maailmanlaajuista huomiota se on saanut PISAn kärkeen sijoittuvien maiden ympärille kehitellyillä koulutuspoliittisilla 'menestys- tarinoilla'. OECD:n koulutuspoliittisena strategiana on tarjota PISAn kärkimaista tunnistettuja parhaita käy- täntöjä levitettäväksi kaikkiin maihin (vrt. Salokangas ja Kauko 2015). Aiemmin kansallisvaltioiden omissa käsissä ollut koulutuspolitiikka on sulautumassa osak- si ylikansallista ohjausjärjestelmää, yhtenä tarkoituk- senaan sovittaa yksittäisten maiden tavoitteet globaa- lin markkinatalouden toimintaedellytyksiin. (Vrt. Breakspear 2012, 4–5; Rutkowski 2015, 683; Sellar ja Lingard 2013; Sjøberg 2016, 106.) PISA-tuloksista joh-

detut koulutuspoliittiset suositukset hyväksytään usein näyttöön perustuvan politiikan nimissä.

PISA on ymmärretty näyttöön perustuvan politiikan airueksi. Ankarimmillaan evidenssin tulisi olla tieteellisesti tutkittua tietoa; OECD ei kuitenkaan itse harjoita lainkaan tieteellistä tutkimusta, vaan lähinnä hyödyntää PISA-koneiston tuottamaa massiivista numeroaineistoa omien koulutuspoliittisten strategioidensa oikeuttamisessa. PISA tuottaa runsaasti standardoitua numeerista tietoa, mutta sitä ei pidä samaistaa tieteellisesti tutkittuun tietoon. Poliittikkaa faktojen jälkeen käsittelevässä tuoreessa kirjassaan Paul-Erik Korvela ja Johanna Vuorelma (2017) kirjoittavat siitä, miten valistusajalta periytyvät ihanteet järkeen ja tietoon perustuvasta politiikasta ovat uhattuina, kun vallalla on pelkkiin mielipiteisiin, tunteisiin ja valheisiin perustuva puhe. Kuten artikkelista käy ilmi, myös näennäisesti empiirisiin tosiasiatietoihin perustuvat PISA-puheet ovat omalla tavallaan ongelmallisia.

PISAn tapa hyödyntää tilastollista Raschin-mallia rankinglistojen laadinnassa on eri yhteyksissä syytäkin kyseenalaistettu (esim. Fernandez-Cano 2016; Guglielmi ja Brekke 2017; Kreiner ja Christensen 2014; Prais 2003, Wuttke 2007). PISAn metodologisia ratkaisuja arvioivassa keskustelussa on turhan vähälle käsittelylle jäänyt kysymys siitä, missä määrin rankingeja varten eri osallistujamaille tilastolliset arvioidut PISA-pisteet käyvät yksiin näiden maiden oppilaiden testitehtävistä suoriutumisen kanssa (vrt. Cosgrove 2015; Prais 2003). Artikkelimme tuo oman lisänsä tämän tietoaution täyttämiseen.

Verrattaessa PISAn eri aikoina tekemien arviointien tuloksia keskenään on syytä ottaa huomioon, että osallistuvien maiden ja hallintoalueiden lukumäärä on kaksinkertaistunut vuoden 2000 ensimmäisen arvion luvusta 32 vuoden 2015 kuudennen arvion lukuun 72. Vasta myöhemmässä vaiheessa mukaan tulleet Aasian maat, kuten Singapore, ja hallintoalueet, kuten Kiinan Shanghai, nousivat välittömästi kärkisijoille. Ironista kyllä, laskevien sijoitusten Suomessa synnyttämä ”PISA-shokki” olisi saattanut jäädä kokematta, jos nämä Aasian edustajat olisivat olleet mukana heti ensimmäisestä arvioinnista alkaen. Tässä yhteydessä sopii soveltaa muista yhteyksistä tuttua fraasia ja todeta, että PISA on niin kuin se luetaan. Samaa voi sanoa myös PISA-näyttöön perustuvasta politiikasta.

Toisin kuin usein luullaan, Suomen sen paremmin kuin muidenkaan maiden sijoitusten nousu tai lasku

PISA-rankingeissa ei kerro suoranaisesti mitään siitä, mitä maan koulutuspolitiikassa on tehty tai jätetty tekemättä. On tärkeää huomata, että OECD:n omassa PISA-kielenkäytössä *education system* tarkoittaa maata tai hallintoaluetta, mutta ei varsinaista koulujärjestelmää (OECD 2003, 3; OECD 2016, 25), eivätkä PISAn testitiedot sisällä kouluja tai koulutusjärjestelmiä koskevia faktoja (Fernandez-Cano 2016; OECD 2016, 25).

OECD on alusta asti tähdentänyt, että PISA ei mittaa sitä, kuinka hyvin oppilaat eri maissa hallitsevat koulun opetussuunnitelman mukaiset oppisällöt, vaan PISAn on tarkoitus ottaa selvää siitä, mitä elämässä ja elämää varten on opittu (OECD 2003, 12; OECD 2010, 22; OECD 2014, 3; OECD 2016, 25). Koulu ei ole millään muotoa ainoa konteksti elämässä tarvittavien taitojen oppimiselle, mihin viittaa myös David Berliner (2015) arviollaan siitä, että yhdysvaltalaislasten päivittäisestä elämästä vain vajaa kymmenesosa on koulua ja valtaosa kuluu kotiväen, kavereiden ja harrastusten piirissä. PISA ilmoittaa arvioivansa sitä, missä määrin pakollisten yhteisten opintojen päättövaiheessa olevat nuoret hallitsevat niitä tietoja ja taitoja, joita täysipainoinen osallisuus globalisoituvassa tietoyhteiskunnassa tulee heiltä edellyttämään (Breakspear 2014; Murphy 2014; Schleicher 2007).

Täysipainoinen toiminta tietoyhteiskunnassa edellyttää OECD:n mukaan sellaiset universaalit perustaidot, joita voidaan kuvata lukemisessa, matematiikassa ja luonnontieteessä määriteltyjen PISA-pisteiden avulla (OECD 2015, 21). Christopher Tienkenin (2014) kärkevä huomautuksen mukaan PISA:n tunnistamat tiedot ja taidot edustavat pikemmin 1800-luvun lopun kuin 2000-luvun avaintaitoja. Robert Mislevy (1993) puolestaan luonnehtii PISAn rankingien laadinnassa soveltamaa Raschin-mallia ”karikatyyriksi”, joka tuo lähinnä mieleen 1800-luvun psykologian.

Sijoittuminen PISA-kärkeen tuli suomalaisille yllätyksenä, samoin kuin siitä seurannut kansainvälisen koulutuspoliittisen keskustelun kasvava kiinnostus suomalaista peruskoulua kohtaan (Chung 2015; Waldow 2017). Asiantuntijat, virkamiehet ja mielipidevaikuttajat omaksuivat tavan lukea PISA-tuloksia kuin koulutodistuksia. Suomen PISA-menestystä on selitetty maan etevällä koulutuspolitiikalla, maisteritasoisella opettajankoulutuksella ja osaavilla opettajilla (ks. esim. Sahlberg 2015; Vetteranta ym. 2016; Välijärvi ym. 2002). Vaille huomio-

ta on jäänyt se keskeinen seikka, että PISA on alusta asti tähdentänyt poikkeavansa muista laajamittaisista kansainvälisistä arvioinneista kuten esimerkiksi PIRLS:istä (*Progress in International Reading Literacy Study*) ja TIMSS:istä (*Trends in International Mathematics and Science Study*) siinä, että sen tuottamat tulokset eivät koulutodistusten tapaan osoita, miten hyvin oppilaat hallitsevat pakollisten perusopintojen oppisisällöt (OECD 2010, 22; OECD 2014, 3; OECD 2016, 25).

Suomi sai elää ”PISA-ihmettä” (vrt. Simola 2005), kunnes järjestyksessään viidennen PISA 2012-arvioinnin tulokset tulivat julki ja synnyttivät pienimuotoisen ”PISA-shokin”, sillä suomalaisnuorten oppimistulosten nähtiin kääntyneen laskuun. Koulutuksen mallimaaksi jo identifioitunut Suomi sai nyt kokea Irlannin, Japanin ja Australian tavoin heikenevien maapisteiden tuottaman pettymyksen samalla kun Aasiasta mukaan tulleet uudet PISA-osallistajat, Singapore ja Shanghai, kiilasivat kärkeen ohi muiden. Tuoreimpien PISA 2015 -tulosten myötä tilanne muuttui ja Suomen voitiin taas uutisoida sijoittuvan OECD-maiden kärkeen (HS 6.12.2016). Opetus- ja kulttuuriministeriön tiedotteen mukaan (OKM 2016) kymmenen vuotta jatkunut oppimistulosten lasku oli lukutaidossa tahtunut, matematiikassa hidastunut ja Suomi palannut kärkimaiden joukkoon. Opetusalan Ammattijärjestö (OAJ 6.12.2016) kuitenkin tyrmäsi PISA 2015 -tulokset Suomen kaikkien aikojen huonoimpina; Helsingin Sanomien Marjukka Liiten (HS 6.12.2016) teki OAJ:ta osuvamman tulkinnan todetessaan, että näiden numeroiden valossa Suomella ei ole ”syytä häpeään”.

PISA 2015 -tulokset saivat suomalaisasiantuntijat puhumaan oppilaiden sosioekonomisen taustan kasvaneesta merkityksestä selitettäessä oppimistulosten eroja (Vettenranta ym. 2016, 52–55). Huoli koulutuksen tasa-arvon heikkenemisestä kävi ilmi myös opetus- ja kulttuuriministeriön tiedotteesta (OKM 2016). Helsingin Sanomat (HS 6.12.2016) tulkitsi PISA-tuloksia siten, että kotitaustan vaikutus oppilaiden PISA-menestykseen olisi kasvussa. Suomalaispoikien ja -tyttöjen välisen eron PISA-pisteissä nähtiin revähtäneen OECD-maiden suurimmaksi, ja huoli poikien alisuoriutumisesta vahvistui entisestään. OKM:n julkaisemassa PISA-raportissa pojilla todettiin ongelmia niin lukemisen, matematiikan kuin myös luonnontieteen edellyttämässä kirjallisuudessa osaamisessa (Vettenranta ym. 2016).

Eri maiden rankingsijoitusten sekä oppilaiden testisuoriutumisen monin tavoin ongelmallista keskinäisyhteyttä eritellään tuonnempana; tarkastellessamme sukupuolten välisiä eroja lukemistehtävistä suoriutumisesta arvioimme myös väitettyä suomalaispoikien alisuoriutumista (ks. Vettenranta ym. 2016, 45–52; 87–88; vrt. HS 6.12.2016). PISAn luku-taitotiedoista puhuttaessa on hyvä pitää mielessä se, että PISAn kolmen aihealueen pisteetykset korreloivat keskenään voimakkaasti, ja siten lukemista koskevat havainnot kuvaavat soveltuvin osin myös muita aihealuetta (ks. Soh 2014).

Tässä artikkelissa problematisoimme sen varsin yksiuolotteisen PISA-järkeilyn, jonka mukaan oppilaiden erinomaiset suoritukset PISA-testeissä voitaisiin tulkita osoitukseksi maan koulutuspolitiikan erinomaisista tuloksista (vrt. Feniger ja Lefstein 2014, 845). Artikkelin on jäsennetty seuraavasti: johdannon jälkeen käsittelemme problemaattista tapaa tulkita PISA-tuloksia koulutusjärjestelmiä koskeviksi tiedoiksi. Sen jälkeen avaamme PISAn viitekehysten metodologista monitulkintaisuutta sekä tarkastelemme maiden PISA-rankingmenestyksen ja oppilaiden testisuoriutumisen ongelmallista keskinäisyhteyttä. Lopuksi otamme tarkasteluun valittujen maiden vaihtelevat reaktiot PISA-tuloksiin. Yksi keskeinen juonne koko artikkelissa on tarkastella eri maissa käytyjä keskusteluja siitä näkökulmasta, mitä PISA-tuloksista perustellusti voi päätellä, ja toisaalta mitä kaikkea PISA-tulosten nimissä on voitu väittää.

PISA-mittaustulosten suoraviivaiset tulkinnat koulua koskeviksi tiedoiksi

Suomalaisasiantuntijat ovat antaneet omat peruskoulusta lähtevät selityksensä Suomen mainioille PISA-tuloksille. 2000-luvun alkuvuosien korkeat PISA-pisteet ja kärkisijat on tulkittu osoitukseksi suomalaisen koulutuspolitiikan etevyydestä sekä yliopistojen antaman maisteritasoisen opettajankoulutuksen erinomaisuudesta (Sahlberg 2015). Aiemmissa kirjoituksissaan monin tavoin ansiokkaasti suomalaista koulua käsitellyt Hannu Simola tuoreessa ”Suomalainen koulu on taitekohdassa” -artikkelissaan kuitenkin toistaa aiemman uskomuksensa siitä, että peruskoulun menestys voitaisiin lukea PISA-mittauksista. Simola (2017, 10) kirjoittaa, että suomalainen peruskoulu on yhä maineensa huipul-

la ja tähdentää, että peruskoulu on PISA-tutkimuksessa viidesti todistettu ainutlaatuisiksi yhdistelmäksi korkeatasoista, tasa-arvoista ja pienillä kustannuksilla tuotettua laatua. Mutta kuten tässä artikkelissa osoitamme, PISA ei voi perustellusti tuollaisia todistuksia antaa. Viimeisestä 2015 PISAsta Simola päätelee, että peruskoulu on vaarassa ränsistyä.

Kiinnostavaa on panna merkille sellainen yksityiskohta, että kun Suomen sijoitus vuoden 2012 PISAssa laski, muulloin ylistetystä opettajankoulutuksesta ei sillä kertaa puhuttu halaistua sanaa. Toisin kuin monissa muissa maissa, Suomessa PISA-pisteiden lasku ei herättänyt kriittisiä kysymyksiä PISA:n metodologiasta tai viitekehuksesta (Cosgrove 2015; Jerrim 2013; Kreiner ja Christensen 2014; Prais 2003; Wuttke 2007).

PISA avasi kansainväliset asiantuntijamarkkinat suomalaisille koulutuspolitiikan tuntijoille, joilla PISA-asioiden sijasta on ollutkin kansainvälisessä katsannossa omaperäistä kerrottavaa suomalaisen peruskoulun ja koko koululaitoksen erityispiirteistä. Innokkaita PISA-turismin organisoijia ja koulutusviennin suunnittelijoita on Suomesta niin ikään ollut helppo löytää (ks. Schatz 2016). Kansainvälisessä metodikeskustelussa suomalaisasiantuntemuksen sijaan loistaa poissaolollaan.

PISAn johtohahmolle Andreas Schleicherille osoitetussa avoimessa kirjeessä (The Guardian 6.5.2014) kansainvälinen ryhmä eturivin koulutustutkijoita esitti huolensa siitä, miten PISAn ruokkima kilpailuasetelma saattaa yllyttää hallitukset lyhytnäköisiin koulutuspoliittisiin uudistuksiin, joiden avulla uskotaan voitavan parantaa maan sijoitusta rankinglistauksissa. Kriitikot pitivät PISAA aivan liian yksilotteisena arvioimaan kasvatuksellisilta ja kulttuurisilta traditioiltaan moninaista osallistujamaiden joukkoa ja arvelivat PISAA nojaavan koulutuspolitiikan todennäköisesti johtavan opetussuunnitelmiin köyhtymiseen ja oppimismotivaation ja -ilon näivettymiseen.

Niin PISAA puolustavat kuin vastustavatkin suomalaisasiantuntijat jakavat sen uskomuksen että PISA kertoisi erityisesti kouluasioista. Asiantuntijoita vaijaa pahanlainen PISA-lukutaidottomuus ja sen seurauksena hurskas tietämättömyys siitä, mitä PISA:n julkituomat numerot tosiasiaa tarkoittavat. PISAsta puhuneista suomalaisista kansainvälisesti tunnetuin on Pasi Sahlberg muun muassa kirjansa *Finnish Lessons* kahden eri painoksen ansiosta (tässä Sahlberg 2015). Sahlbergin kansainvälinen asiantuntijamaine

nojaa Harvard-vierailujen lisäksi ennen kaikkea hänen suomalaista koulua käsitteleviin kansainvälisiin esityksiinsä, joihin hän on sisällyttänyt sopivaksi katsomassaan määrin huomioita myös Suomen PISA-menestyksestä. Mutta toisin kuin Suomessa paremmin tunnetut Simola ja Risto Rinne, Sahlberg ei niinkään profiloitu Suomen PISA-ihmeen kriitikoksi, vaan tasapainottelee PISAA puoltavien ja vastustavien kantojen välimaastossa päätyäkseen kuitenkin miltei OECD:n äänenpainoin tähdentämään PISAn korvaamatonta koulutuspoliittista arvoa. Tämä käy ilmi muun muassa siitä, miten hän yhdessä Andy Hargreavesin kanssa *The Washington Postin* (March 24, 2015) julkaisemassa kannanotossa tekee pesäeroa jo mainittuun *The Guardianissa* ilmestyneeseen koulutustutkijoiden voimaperäiseen PISA-kritiikkiin.

Simolan ”esseiden”, kuten hän kirjoituksiaan kutsumu, myyvät otsikot ulottuvat PISAn ”ihmeestä”, paradoksaalisen ”vastatuuleen luovimisen” kautta suomalaisen koulutuksen ”mysteeriin” (Simola 2005; 2013; vrt. 2015). Simolan kansainvälinen julkisuusarvo näyttää paljolti nojaavan Suomen PISA-menestykseen, mutta eivät Simolankaan kriittiset puheenvuorot osoita perehtyneisyyttä siihen, mitä PISAn tuottama massiivinen data-aineisto tosiasiaa pitää sisällään. Päinvastoin, hän pitää edelleen kiinni ongelmallisesta tulkinnastaan, että PISA kertoo suomalaisen peruskoulun tilanteesta (Simola 2017), mitä PISA ei tämän artikkelin osoittamalla tavalla voi lainkaan tehdä.

Finnish Lessons 2.0 on Sahlbergin (2015) kansainvälisesti tunnetun kirjan toinen painos. Mahtipontisen alaotsikkonsa mukaan siinä kerrotaan, mitä opittavaa maailmalla on koulutuksen muutoksista Suomessa. Muiden suomalaisasiantuntijoiden tavoin myös Sahlberg nojaa Suomen korkeiden PISA-sijoitusten ja koulujen erinomaisuuden välille oletettuun vahvaan keskinäisyhteyteen. Sahlbergin mukaan Suomen koulutusjärjestelmä siirtyi parissa vuosikymmenessä keskinkertaisuudesta erinomaisuuteen. Mitään empiirisiä faktoja keskinkertaisuusväitteiden tueksi ei kuitenkaan ole tarjolla.

Sahlberg on kyllä tietoinen siitä, että pohjoismainen hyvinvointivaltio ja suomalaiskotien vankka sivistysmyönteinen perinne ovat olleet ratkaisevat tekijät omaleimaisen suomalaisen koulukäyttäytymisen taustalla. Mutta kaikesta päätellen pohjoismaista hyvinvointivaltiota ei – omassa epämuodikkueudessaan – kannata nykyisessä mielipideilmastos-

sa yrittää pukea kansainvälisen yleisön kaipaamaksi parhaiden käytäntöjen resepteiksi. Sen sijaan Sahlbergin ”paradokseiksi” muotoilemat suomalaisreseptit ovat lennokkaita tyyliin ”opeta vähemmän, opi enemmän” ja ”testaa vähemmän, opi enemmän”. Eri asia on, missä määrin ne pitävät yhtä suomalaisen koulutodellisuuden kanssa ja miten opi enemmän realisoituu suomalaiskouluissa. Koulutusihmeen ytimessä Sahlbergilla ovat maisteritasoisen koulutuksen saaneet suomalaisopettajat.

Itse ajattelemme, että lasten pärjääminen koulussa ja elämässä on Suomessa kuten muuallakin edelleen vahvasti yhteydessä kotien elinolojen ja toimintatapojen vakauteen. Ratkaisevan tärkeä elämänvaihe on kouluikä edeltävä varhaislapsuus, joka antaa, jos antaa, eväät tulevaan elämään. Mitä opettajuuteen tulee, suomalaistilanne toki poikkeaa monista maista siinä, että opettajankoulutukseen on runsaasti hakijoita, joista valita (Malinen ym. 2012, 581). Tiedossa ei kuitenkaan ole näyttöä siitä, että nykyinen koulunpenkiltä suoraan maisteriopettajaksi koulittu nuori olisi jollakin erityisellä tavalla parempi opettaja kuin aikanaan kansakoulupohjalta parivuotisen seminaarin käyneet elämänmakuiset opettajat. Itseltään selvää ei myöskään ole se, että vuosikymmenten takaiset suomalaisnuoret olisivat olleet osaamiseltaan jotenkin nykynuoria huonompia tai koulut oman aikansa kriteereillä leimallisesti keskinkertaisia kuten on annettu ymmärtää.

Rinne (HS 5.12.2016) ottaa kriittisen kannan ”ylikansalliseen indikaattoripolitiikkaan” ja numeroilla hallitsemiseen, mistä PISA käy hänelle paraatiesimerkiksi. Emme kiistä PISAn heikkouksia, mutta evidenssiperustaisen politiikan aikakaudella ylimalkainen puhe indikaattoripolitiikasta kuulostaa epäanalyttiseltä (vrt. Varjo ym. 2016). Olisikin terve tullutta, jos kasvatustieteilijät ottaisivat numerot ja indikaattorit vakavasti ja ryhtyisivät kehittämään niitä nykyistä mielekkäimmiksi, kuten Simon Marginson (2014) rankingeista puhuessaan asian muotoilee. Marginsonin (ks. Kivinen ym. 2017) mukaan olisi hyvä, jos yhteiskuntatieteilijät kiihkeän mutta ylimalkaisen vastustuksen sijaan ottaisivat pohdittavakseen, miten rankingeja ja valtion rajat ylittäviä arviointeja voitaisiin yhteistuumin kehittää siten, että niiden negatiiviset vaikutukset minimoituisivat ja että ne saataisiin tuottamansa tiedon varassa nykyistä paremmin palvelemaan yleistä etua. On tärkeätä, että rankingit ja arvioinnit nojaisivat asianmukaisesti organisoituihin globaaleihin tietokantoi-

hin, koska ne mahdollistaisivat maksimaalisen objektiiviset arvioinnit. Ikävä kyllä PISA-asiantuntijoi- ta, niin kriitikoita kuin puolestapuhujiakin, näyttää vaivaavan vakava PISA-lukutaidottomuus.

PISAn viitekehyksen metodologinen monitulkintaisuus

Yhteistä kolmivuotisvälein (2000, 2003, 2006, 2009, 2012 ja 2015) järjestetyille kuudelle PISA-arviointikerralle on ollut kerätä kustakin osallistujamaasta vähintään 150 koulusta minimissään 4 500 oppilasta käsittävät maanäytteet (OECD 2014, 66; 70–73). PISAssa sovelletaan IRT-periaatetta (*Item Response Theory*) tuottamaan tilastolliset arviot siitä ”taitotasosta”, jolle osallistujien kaltaiset henkilöt todennäköisesti yltaisivät kullakin kolmella pääaihealueella, jotka ovat lukeminen, matematiikka ja luonnontiede. Taitotasoarviot nojaavat Raschin-mallin avulla tuotettuihin viiteen mahdollisten arvojen jakaumaan. (Ks. NCES 2014.)

Seuraavaksi tarkastelemme PISAn metodologias- ta käytyä keskustelua, johon oman panoksensa antoivat William Stewartin julkisesti esittämät (TES 26.7.2013), kysymysmuotoon puettut keskeiset epäilyt, kuten ”mitä jos saisit tietää, että PISAn maaver- tailut eivät perustu kaikille oppilaille yhtäläisesti teetetyille testeille, vaan valikoidut oppilaat saavat vastattavakseen eri tehtävät ja puuttuvat vastaukset korvataan PISA-laskentamenettelyjen tuottamilla mahdollisilla arvoilla?” Yksi Stewartin kysymyksistä on, ”entä mitä jos saisit tietää, että vaihtamalla tehtävien järjestystä sijoitukset maarankingeissa muuttuvat merkittävästi, esimerkiksi Iso-Britannia voi sijoittua jollekin sijalle välillä 14–30 tai Tanska välil- le 5–37”. Artikkelimme problematisoi kautta linjan maiden PISA-rankingeissa menestymisen sekä oppilaiden testisuoriutumisten välille oletetut yhteydet.

Suomessa kiinnostus PISAn tiedonkeruunet- telyihin heräsi suomalaispoikien pudottua vuoden 2012 arvioinnissa sijalle 14. Suomen maanäytteen voitiin tuolloin todeta poikenneen merkittävästi aiempien vuosien arvioinneista. Vuoden 2012 arvi- oinnissa päätettiin painottaa maahanmuuttajataus- taisten nuorten osuutta, minkä seurauksena maan- näytteeseen piti erityisesti valita kouluja, joissa oli vähintään viisi maahanmuuttajataustaista 15-vuo- tiasta. Suomen vuoden 2012 maanäytteeseen tuli lopulta aiempiin arviointeihin verrattuna kaksinker-

tainen määrä osallistujia ja maahanmuuttajataustaisten osuus nousi yli seitsenkertaiseksi aiempaan verrattuna (ks. Harju-Luukkainen ym. 2014, 11). Osmo Soininvaara (Soininvaara 8.12.2013) kiinnitti heti tuoreeltaan huomiota muuttuneesta näytteenvalinnasta koituviin seurauksiin huomauttamalla, että lähtökohtaisesti maahanmuuttajien tyypillisesti asuttamien alueiden kouluissa yleensä myös syntyperäisten suomalaisten sosiaalitausta jää muiden alueiden kouluja matalammaksi. Soininvaaran arvelua Suomen PISA-notkahduksen saattavan selityä muuttuneesta näytteenvalinnasta asiasta käynnistyi vilkas nettikeskustelu, mutta kovin tarkkaa selvyyttä näytteenottotavan ja muuttuneiden PISA-sijoitusten keskinäisyhteyksiin ei saatu.

PISA standardoi logististen regressioanalyysin avulla osallistujamaiden pisteet suhteessa teoreettiseen keskilukuun 500, jonka perusteella maat järjestetään parhaimmasta heikoimpaan. Keskiluku palvelee maiden ranking-sijoitusten määrittämistä ja edustaa mahdollisten arvojen laskennassa aineiston kaikkien maiden keskiarvoa. Mutta, toisin kuin on väitetty, oppilaiden oppimistulosten keskiarvo se ei voi olla (ks. Vettenranta ym. 2016). PISASTA ei voida identifioida yksittäisiä oppilaita, eikä niin muodoin myöskään oppilaskohtaisia testituloksia, joten oppilaiden tulosten keskiarvo ei ole PISASTA mielekkäästi määriteltävissä. Tapa, jolla Suomen PISA-ryhmä (Vettenranta ym. 2016) lukee oppilaiden välisiä eroja PISAN maatasolla tuotetuista normaalijakaumista, tunnetaan myös ekologisena virhepäätelmänä (ks. Fernandez-Cano 2016; Guglielmi ja Brekke 2017).

Aineiston ongelmista huolimatta OECD tekee päätelmiä oppimistulosten eroista sen perusteella, missä määrin pistekeskiarvojen vaihtelua voidaan selittää eroilla oppilaiden sosioekonomisessa taustassa (ks. OECD 2013, 27; 37–39). Suomen PISA-raportin mukaan oppilaiden kotitaustan selitysvoima oppimistulosten eroon on kasvussa ja tasa-arvo murenemassa (Vettenranta ym. 2016, 52–55). Väitteen perustana olevat väestötiedot olisi hyvä tarkistaa.

Havainnollistamme parilla esimerkillä, miten arveluttavaa turvautuminen PISAN oppilaiden taustaja koskevan kyselyaineiston tietoihin tosiasia on. Ensimmäinen esimerkki koskee korkeakoulututkinnon suorittaneiden äitien osuutta, minkä väitetään Suomessa nousseen vuoden 2003 osuudesta 25 prosenttia peräti 51 prosenttiin vuoteen 2015 mennessä. Korkeakoulututkinnon suorittaneiden suomalaisäitien osuus olisi siis PISA-tietojen mukaan kaksin-

kertaistunut viidessätoista vuodessa. Tilastokeskuksen koko ikäluokan kattavien rekisteripohjaisten tietojen mukaan puheena olevien PISA-ikäisten suomalaisnuorten äitien ikäluokissa pyöreästi 22 prosenttia on suorittanut korkeakoulututkinnon, eikä tuossa osuudessa ole viidessätoista vuodessa tapahtunut kuin kahden prosenttiyksikön muutos. Näiltä osin PISAN tiedot eivät siis pidä likimainkaan yhtä tosiasioiden kanssa. Emme lähde tekemään muista maista vastaavia vertailuja tässä, toteamme vain, että esimerkiksi Etelä-Koreassa korkeakoulututkinnon suorittaneiden äitien osuus oli vuoden 2012 PISASSA 35 prosenttia, mutta 2015 PISASSA enää viisi prosenttia. Näillä luvuilla ei tietenkään voi olla mitään yhteyttä siihen, mitä suuruusluokkaa korkeakoulutettujen korealaisäitien osuudet ovat tosiasiaa olleet.

PISAN keräämässä oppilaiden kotitaustaa koskevassa kyselyaineistossa puuttuvien tapausten osuus on 28 prosentin luokkaa (OECD 2007: liite A8). Nämä PISA korvaa logistisen regression tuottamalla *Ad hoc* -ryhmillä, joista johdettujen taustatietojakaumien oletetaan noudattavan Gaussin normaali-jakaumaa (vrt. Fernandez-Cano 2016; Rutkowski 2017; Weirich ym. 2014).

Toinen esimerkki koskee PISA-tulosten kirvoittamaa huolenaihetta matalataitoisuudesta, joka päätellään kahdelle alimmalle taitotasolle jäävien osuuksista. Suomalaisraportti esittää, että 11 prosenttia suomalaisnuorista (pojista 16 prosenttia ja tytöistä 7 prosenttia) ei saavuta sellaista lukutaidon tasoa, jolla voisi jatkaa ongelmitta opintoja ja olla täysipainoisesti osallisena yhteiskunnassa (Vettenranta ym. 2016, 27, 51). Väite, jonka mukaan heikko kirjallinen osaaminen kasautuisi samoille nuorille ja että näistä noin kaksi kolmasosaa olisi poikia (Vettenranta ym. 2016, 94) edellyttää omintakeista PISA-luentaa. Jos PISA ei lainkaan eksplikoi osallistujien henkilökohtaisia testituloksia, niin miten ja mistä kasautuminen samoille nuorille on voitu päätellä?

Maailmanpankki (2017) nojaa niin ikään PISA-aineistoon arvioidessaan heikon kirjallisen osaamisen esiintyvyyttä eri maiden nuorisoväestössä, mutta se ei esitä Suomen PISA-ryhmän tavoin yli kymmenen prosentin matalataitoisten osuuksia, vaan raportoi Suomessa heikosti osaavien osuuden jäävän yhteen prosenttiin lukemisessa, neljään matematiikassa ja nolnaan luonnontieteessä. Myös jäsenmaiden väestörekistereihin nojaavat UNESCO:n (2017) tilastot osoittavat, että suomalaisikäluokat täyttävät

täysimääräisesti oppivelvollisuutensa ja ovat hyvin-kin luku- ja laskutaitoisia. Sitä paitsi olisi koko lailla outoa, jos maassa, jossa koko kansa on jo 1800- ja 1900-lukujen vaihteessa osannut lukea, lukutaito olisi katastrofaalisesti häviämässä (Kivinen 1988; vrt. Kivinen ym. 2016).

Svend Kreiner ja Karl Bang Christensen (2014) vauhdittivat PISA-metodologiasta käytävää keskustelua esittämällä perustellun epäilynsä PISAn tavasta soveltaa Raschin-mallia. Kreiner ja Christensen nostivat esiin ongelmallisen DIF-ilmion (*differential item functioning*) (vrt. Zwitser ym. 2017). Kysymys on siitä, että PISAssa testitehtävien järjestys vaikuttaa huomattavasti maiden rankingjärjestykseen. PISAn johtohahmo Andreas Schleicher myöntää PISA-tulosten jonkinasteisen riippuvuuden valituista tehtävistä ja testin koordinoitintavoista, mutta kiistää esitetyt ongelmat ”hölynpölynä” (Science Guide 29.8.2013). Siitä, että Kreiner ja Christensen löytävät DIF-ongelmia tarkastelemastaan yhdestä tehtävähöystä, ei Schleicherin mukaan voi päätellä vastaavien ongelmien esiintyvän yhtä lailla kaikissa kolmesatoista toisistaan poikkeavassa tehtävähöyössä (vrt. Zwitser ym. 2017). Schleicherilta jäi kuitenkin huomaamatta, että Kreiner ja Christensen (2014) olivat tietoisesti ottaneet tarkasteltavakseen tehtävähöyön, johon sisältyivät kaikki PISA 2006 -arvioinnin lukemistehtävät. OECD:n nimissä puhuva Schleicher tuomitsi Kreinerin ja Christensenin (2014) artikkelin tieteellisesti arvottomaksi; tiedeyhteisöä edustavat tutkijat eivät ole yhtyneet Schleicherin tuomitsevaan kantaan.

PISAn ottaessa 2015 arvioinnissaan käyttöön digitaaliset testivalikot digitalisoinnin tarkoittamattomia seurauksia ei osattu ottaa huomioon. Hyvin menestyneille Aasian maille kuten Etelä-Korea ja Japani sekä sellaisille Kiinan hallintoalueille kuin Hong Kong ja Taipei vuoden 2015 PISA osoitti edelliseen arviointiin verrattuna keskimäärin 20 pisteen romahduksen lukutaidossa. Sikäläiset asiantuntijat ja virkamiehet selittivät pudotuksen seuranneen siitä, että PISA oli vaihtanut kynä-paperitehtävät tietokonepohjaisiin testeihin (Komatsu ja Rappleye 2017). Myös Schleicher pitää Aasian lukutaitopisteiden laskua outona ja arvelee niin ikään sen aiheutuneen tietokonepohjaisesta testien käyttöönotosta (TES 24.3.2017). Myös Saksan pärjäämistä PISAssa vuosina 2000–2015 analysoinut tutkijaryhmä päätyi siihen, että tietokonepohjaisten testien käyttöönotto on tuonut ennakoimattomia vaikutuksia PISA-pis-

teisiin edellä mainitulla tavalla (Robitzsch ym. 2017). Nähtäväksi jää, onnistuuko PISA parantelemaan tilastollisia mallejaan ja oikaisemaan mitta-asteikkojaan yhteensopiviksi käyttöönotettujen tietokonepohjaisten testimenettelyjen kanssa.

PISA:n maarankingien ja oppilaiden testaamisen ongelmalliset yhteydet

PISA edellyttää, että kustakin maasta kerättyjen testivastausten lukumäärän tulee joka tehtävässä ylittää tilastollisesti riittäväksi katsottu 400 vastauksen kynnyksen. Vastaukset kerätään BIB-periaatteen (*balanced incomplete block*) nojalla (ks. Murphy ja Schulz 2006, 7; OECD 2014, 30–31). BIB-periaatetta soveltamalla PISA säästää aikaa ja rahaa aineiston keruussa, koska kukin oppilas saa vastattavakseen vain osan tehtävistä, mutta toisistaan poikkeavin tehtävähöyöimien vuoksi PISA menettää mahdollisuuden saada yksiselitteiset testitulokset kaikilta osallistuneilta oppilailta (vrt. von Davier ym. 2009). Esimerkiksi vuoden 2006 PISA-testissä noin puolet oppilaista ei tehnyt ainoatakaan lukemistehtävää (Kreiner 2011). Vuoden 2015 lukutaitotesti sisälsi 88 lukemistehtävää, joihin vastasi 12 prosenttia OECD-maista osallistuneista 15-vuotiaista nuorista. Tämä tarkoittaa, että PISAssa lukemisen aihealueella arvioidut standardoidut maapisteet nojaavat lähes 90 prosenttisesti puuttuvia testivastauksia korvaaviin laskennallisiin arvoihin, jotka saadaan IRT-periaatetta ja Raschin-mallia soveltaen.

Laatiessaan maarankingeja PISA tuottaa tilastolliset arviot erikseen poikien ja tyttöjen lukutaidon tasosta ja järjestää maat näiden perusteella parhaimmasta heikoimpaan. PISAn kollektiivitasen arviot, kuten maan lukutaidon taso, ilmaistaan sekä tasavälisenä järjestyslukuna (rankingsijointus) että etäisyytenä (rankingpisteet) keskiluvusta 500 (ks. Kreiner ja Christensen 2014; Wuttke 2007). Suomalaisen PISA-asiantuntijaryhmän mukaan sukupuolten välinen ero lukutaidossa on maassamme revennyt OECD-maiden suurimmaksi (Vettenranta ym. 2016, 49–51), mutta PISAn ranking-pisteistä ei tosiasiaa voi päätellä taitoeroja tyttöjen ja poikien välillä. Seuraavassa tarkastelemme tyttöjen ja poikien PISA 2015 -lukemistestissä pärjäämisen yhteyttä kahdeksantoista kärkimaan rankingpisteisiin.

Raportoidessaan tyttöjen ja poikien välisistä eroista lukutaidon tasossa Suomen PISA-ryhmä päätelee

tyttöjen olevan yhden kokonaisen kouluvuoden poikia edellä (Vettenranta ym. 2016, 34; 45; 49). Kun muistetaan, että PISAn on tarkoitus arvioida, miten nuoret tulevat pärjäämään koulun ulkopuolisessa elämässä, ei hevin voi nähdä mielekästä yhteyttä elämässä pärjäämisen ja koulussa vietettyjen vuosien välillä – etenkin kun on puhe Suomesta, missä tytöt ja pojat käyvät peruskoulua yhtä monta vuotta ja muutoinkin etenevät samaa tahtia luokalta toiselle. Sukupuolten välille laskennallisesti arvioitujen pistekeskisarvojen eron muuntaminen kouluvuosi on jonkinmoisen väärinymmärryksen tulos ja sellaisenaan täysin irti asiayhteydestä.

PISAn tuottamat kolmen aihealueen sekä kummankin sukupuolen ranking-pisteet saadaan PISA Data Explorer -hakupalvelusta (OECD 2017b); omaa tarkasteluamme varten olemme poimineet sieltä Suomen ja 17 muun kärkimaan tyttöjen ja poikien PISA 2015 lukutaidon ranking-pisteet. PISA 2015 -tietokannasta (OECD 2017a) olemme poimineet tiedot lukemistestin tehtäviin vastanneiden suomalaistyttöjen ja -poikien hyväksymisasteista, jolla tarkoitetaan hyväksytyksi vastanneiden osuutta kaikista vastanneista. Suomesta PISA 2015 -arviointiin poimittiin 2 863 tyttöä ja 3 019 poikaa ja nämä 5 882 oppilasta saivat vastattavakseen kotitaustoja koskevan kyselyn. Testiin kuuluviin 88 lukemistehtävään saatiin vastaukset keskimäärin 700 oppilaalta (350 tytöltä ja 350 pojalta). Yksittäinen oppilas vastasi tyypillisesti 11/88 lukemistehtävään.

PISA hyödyntää aiemmista arvioinneista kertynyttä tietoa oppilaiden tehtäviin hyväksyttävästi antamista vastauksista. Niiden nojalla kullekin tehtävälle voidaan määrittää odotettavissa olevat hyväksymisasteet. Mitä todennäköisempää tehtävän hyväksytyksi läpäiseminen on, sitä matalammaksi tehtävän vaikeusaste on arvioitu. PISA 2015 lukemistestissä helpoimman lukemistehtävän hyväksytyksi läpäisemisen todennäköisyys on 0.94, ja vaikeimman 0.08.

Suomalaisoppilaiden lukemistestissä pärjäämiselle sopiva vertailukohta on kaikkien osallistujamaiden oppilaille laskettu keskimääräinen hyväksymisaste 56 prosenttia, minkä suomalaisoppilaat ylittävät selvästi 64 prosentin hyväksymisasteellaan (tytöt 68.4 prosenttia ja pojat 58.9 prosenttia). Suomalaisyttöjen hyväksymisaste on kymmenisen prosenttiyksikköä suomalaispoikia korkeampi, mutta aiempaa Marjukka Liitenin huomautusta mukaillen voimme kuitenkin todeta, että suomalaispojilla ei ole ”syytä häpeään”, sillä myös poikien hyväksymis-

aste ylittää koko aineiston keskimääräisen hyväksymisasteen noin kolmella prosenttiyksiköllä.

Maakohtaisissa lukutaitoarvioissaan PISA operoi hyväksymisasteiden sijaan ranking-pisteillä, jotka se määrittää suhteutettuna teoreettiseen keskilukuun 500. Vertailllessamme hyväksymisasteita havaitsemme suomalaispoikien (58.9 prosenttia) ja korealaispoikien (59.2 prosenttia) hyväksymisasteiden ylittävän selvästi kaikkien maiden keskimääräisen hyväksymisasteen, joka on vajaat 56 prosenttia, mutta testissä heikommin menestyneet suomalaispojat saavat PISAn lukutaitorankingissa 500 ylittävät pisteet 504, sen sijaan testeissä paremmin menestyneet korealaispojat jäävät 498 pisteellään sen alle. Myös saksalais-, norjalais- ja tanskalaispojat suoriutuvat lukemistehtävistä keskimääräistä paremmilla hyväksymisasteilla, mutta jäävät lukutaitorankingissa alle 500 pisteen teoreettisen keskiluvun.

Suomalaistyöt jäävät 11/88 tehtävässä aiemmista arvioinneista johdetun odotusarvon mukaisten hyväksymistodennäköisyyksien alle; näissä tehtävissä tytöt siis suoriutuvat odotettua heikommin. Suomalaispojat puolestaan suoriutuvat odotettua heikommin 26/88 tehtävässä, mikä tosin sanoen tarkoittaa, että valtaosassa (62/88) tehtävistä suomalaispojat suoriutuvat vähintään odotetusti. Esimerkiksi ruotsalaispojat suoriutuvat odotettua heikommin 54/88 ja islantilaispojat 68/88 tehtävässä. Myös Yhdysvaltain (42), Britannian (44) ja Australian (43) pojat suoriutuvat lukemistehtävistä suomalaispoikia useammin alle odotusten. Suhteellisen vaatimattomasti suoriutuvat myös Tanskan (33), Norjan (37) ja Korean (33) pojat. Nämä esimerkit osoittavat, että suomalaispoikien väitettyä alisuoriutumista ei voida yksiselitteisesti päätellä PISA 2015 -lukemistestin tuloksista.

Tarkastelemme vielä Suomen poikien suoriutumista lukemistehtävissä kaikissa kuudessa eri aikoina tehdyssä PISA-arvioinnissa. Vuoden 2006 arvioinnissa Suomen poikien keskimääräinen hyväksymisaste lukeutui osallistujamaiden parhaaseen neljännekseen 93 prosentissa annetuista tehtävistä. Kolme vuotta myöhemmin vuoden 2009 arvioinnissa suomalaispoikien vastaava lukema oli edelleen korkea 80 prosenttia. Mutta vuoden 2012 arvioinnissa Suomen poikien hyväksymisaste ylsi parhaaseen neljännekseen vain 40 prosentissa tehtävistä. Suomalaisyttöjen vastaavat lukemat olivat myös laskeutuneet vuoden 2006 huippulukemasta 93 prosenttia vuoden 2012 lukemaan 73. Tuoreimmassa vuoden

2015 arvioinnissa suomalaispoikien lukemat ovat taas nousseet 63 prosenttiin ja suomalaistyöillä 78 prosenttiin.

Mutta kuten aiemmin todettiin, Suomen maanäyte vuoden 2012 arvioinnissa poikkesi merkittävästi edeltäneistä arvioinneista. Vastaavanlaisia havaintoja poikkeavan maanäyteen yhteydestä poikien testisuoriutumiseen on tehty myös Irlannissa (ks. Cosgrove 2015). Irlantilaispojat ovat sittemmin kohentaneet rankingasemiaan vuoden 2009 arvioinnista. Myös suomalaispoikien rankingsijoitus on palautunut vuoden 2012 arvioinnin alhaisesta sijaluvusta 14 vuoden 2015 sijalukuun 7. Sen sijaan, että suomalaispoikien arvioinnista toiseen vaihtelevat lukutaitopisteet tulkittaisiin osoituksiksi suomalaispoikien muuttuneesta lukutaidosta, ne pikemminkin osoittavat vain sen, että oppilaiden lukemistehtävistä suoriutumisen vertailu eri PISA-arviointien kesken ei hevin käy päinsä. Suomalaisoppilaiden hyväksymisasteista pääteltävien tulosten vaihtelu arviointikerrasta toiseen tukee epäilyämme siitä, että vaihtelu ei voi selittyä koulutusjärjestelmässä tosiasiallisesti tapahtuvilla muutoksilla, sillä koulutusjärjestelmässä ja varsinkin koulujen käytännöissä olennaiset asiat eivät kolmessa vuodessa ehdi juurikaan muuttua.

Maittain vaihtelevat reaktiot PISA-tuloksiin

Seuraavassa tarkastelussa kiinnitämme huomiota muun muassa eri maissa koettuihin PISA-shokkeihin seurauksineen. Kuten mm. Bob Lingard (2016) huomauttaa, itse OECD saattaa hyvinkin ymmärtää eri maiden kansalaisia järkyttävät PISA-shokit myös tarkoituksenmukaisiksi keinoiksi herätellä hyvään oloon tuudittautuvia hallituksia ryhtymään OECD:n toiveiden mukaisesti poliittisiin uudistuksiin. Miltei poikkeuksetta maiden PISA-keskusteluissa esiintyy PISAn kärkimaiden vakioedustajana nimenomaan Suomi, joskaan sen parhaat käytännöt eivät hevin näytä taipuneen toteuttamiskelpoisiksi resepteiksi kohentamaan muiden maiden PISA-tuloksia. Yhdysvaltoja käsitellessämme nostamme esiin myös siellä tehdyt havainnot PISA-menestyksen tason ja kansantalouden kehittyneisyyden keskinäisyhteyksistä.

Alun perin kärkimaiden joukkoon Suomen lailla kuuluneissa Irlannissa ja Australiassa on heikkenevien PISA-pisteiden vuoksi saatu sittemmin päivittellä murenevaa myyttiä ”maailmanluokan koulutusjärjestelmästä”. Vaikka australialaisnuoret edelleen

pärjäävät PISAssa verraten hyvin, maan putoaminen rankingissa huolestutti australialaispäättäjiä siinä määrin, että nämä päättivät kirjata maan koulutuspolitiikan viralliseksi tavoitteeksi sijoittua PISAn viiden kärkeen vuoteen 2025 mennessä. Kaikesta päätelleen tehtävää on mahdotonta toteuttaa, sillä tuskin kellään voi olla asiallista tietoa siitä, millä koulutuspoliittisilla toimenpiteillä ja kouluja koskevilla uudistuksilla PISAssa voi kiivetä ylöspäin. Australia hakee koulutusreformilleen mallia sellaisista Aasian PISA-kärkimaista kuin Etelä-Korea, Hong Kong ja Singapore. (Gorur ja Wu 2015.)

1980- ja 1990-luvuilla koulutusmaailman mallimaana tunnetuksi tullut Ruotsi ei ensimmäisessä PISAssa yltänyt kärkisijoille, mutta ei kokenut myöskään PISA-shokkia. Sen sijaan vuoden 2003 PISA, samana vuonna toteutettu TIMSS sekä sikäläinen kansallinen koulututkimus käynnistivät Ruotsissa huolestuneen koulukeskustelun. PISA 2012:ssa Ruotsin pisteet kaikilla aihealueilla olivat pudonneet reilusti alle OECD-maiden keskiarvon ja verrattuna vuoteen 2000 Ruotsin pudotus oli kaikista PISAn osallistuneista maista jyrkin. Merkittävimminä syinä heikkeneville tuloksille ruotsalaiset pitivät opettajankoulutuksen heikkoa laatua sekä työrauhaongelmia koululuokissa; lisäksi esiin nousivat kasvavat osaamiserot Ruotsin kantaväestöön kuuluvien ja maahanmuuttajataustaisten nuorten välillä. Ruotsin koulutuspolitiikan mallimaiksi valikoituivat hyvän PISA-menestyksensä vuoksi Kanada ja Suomi. Heikkojen PISA-tulosten parantamiseksi toteutetut uudistukset kohdistuivat opettajankoulutukseen ja koulujen opetussuunnitelmaan. (Ringarp 2016.) Vuoden 2015 PISAssa Ruotsin pisteet yltävät jo keskitasoon kaikilla aihealueilla.

Saksa, Tanska ja Japani kokivat omat PISA-shokkinsa 2000-luvun alkuvuosina. Saksan PISA-menestys jäi reippaasti alle odotusten, kun 30 osallistuneen OECD-maan joukossa Saksa sai alle keskitason sijoitukset kaikilla kolmella aihealueella. Tanskalaismediassa maan suhteellisen vaatimattomien PISA-tulosten nähtiin olevan ristiriidassa Tanskan OECD-maiden korkeimpiin kuuluvien koulutusmenojen kanssa; ristiriitaa nähtiin myös tanskalaisoppilaiden korkeana pidetyn koulumotivaation ja vaatimattomien PISA-tulosten välillä. Tanskassa ja Saksassa odotettua heikompi PISA-menestys käänsi huomion myös erilaisista perhe- ja kulttuuritaustoista tulevien oppilaiden välisiin taiteroihin. PISA-shokin herättäminä Saksa ja Tanska ryhtyivät uudistamaan

omia opetussuunnitelmiaan. Koululaisten osaamista eri kouluaineissa alettiin niin ikään selvittää kansallisin testein. Myös opetusmenetelmiä koskevia ohjeita opettajille tarkennettiin. (Ks. Egelund 2008; Ertl 2006.)

Florian Waldow (2017) on analysoinut PISAn esiintymistä saksalaisessa laatulehdistössä silmämääränä nimenomaan PISA-kärjen mallimaat (*reference societies*) sekä se, mitä aineksia kouluihin ja koulutuspolitiikkaan Saksan on ajateltu voivan niistä lainata. Waldow pelkistää saksalaiskeskustelun PISAn kärkimaista yhtäältä ”skandinaaviseen” Suomeen ja toisaalta Itä-Aasiaan kuten Etelä-Koreaan ja Japaniin. Aikavälillä 2001–2014 Suomen saamat positiiviset maininnat lyövät Itä-Aasian vuosi toisensa jälkeen, kunnes vuodesta 2011 mainintojen määrät kääntyivät Itä-Aasian eduksi. Koko jaksolla 2001–2014 negatiivisten ja positiivisten mainintojen määrät suosivat Suomea (74 prosenttia positiivisia ja 22 prosenttia negatiivisia), Itä-Aasian saadessa voittopuolisesti negatiivisia mainintoja (65.8 prosenttia) ja vain niukasti positiivisia (16.2 prosenttia). Tässä emme voi syventyä siihen, miten keskustelua vaikkapa Saksassa, Yhdysvalloissa ja muualla hämmensi voimakkaalla PISA-esiinmarssillaan Shanghai, jonka OECD on sittemmin häivyttänyt PISA-aineistosta edustavuusongelmiin vedoten.

Mitä Saksaan tulee, Waldowin mukaan siellä Suomen koulut ymmärretään progressiivisiksi ja lapsikeskeisiksi sekä opettajat tiedetään hyvin kouluteuiksi, arvostetuiksi ja luotettaviksi. Siinä missä Suomi esiintyy positiivisessa valossa, saksalaisten suhtautuminen Itä-Aasiaan on negatiivista. Aasiailiskouluja luonnehtii kuri, ”pänttäys” (*rote learning*), ankara koepaine ja ilottomuus. Waldowin maalailemana Suomi näyttäytyy PISA-pyhiinvaeltajille ”Pohjolan ihmemaana”, kun taas Aasian maat kouluineen nähdään osana kurin ja kokeiden epäinhimillistä koneistoa.

Japanin ensimmäiset PISA-tulokset olivat odotuksia paremmat maan vähän yllättäen sijoituessa kärkimaiden joukkoon. Mutta kun vuoden 2003 PISAssa sijoitukset lukemisessa laskivat sijalta kahdeksan sijalle 14 ja matematiikassa ensimmäiseltä sijalta kuudenneksi, Japanin koulutusjärjestelmä sai mediassa osakseen kiivasta arvostelua. Kärkijoukosta putoamisen katsottiin osoittavan koulutusjärjestelmän kriisiä, josta selviämiseksi päättäjiltä alettiin vaatia välittömiä uudistuksia maan koulutuskäytäntöihin. (Takayama 2008.)

Standardoidut osaamistestit olivat Yhdysvalloissa arkipäivää jo 2000-luvulle tultaessa, eikä PISA heittänyt siellä koulutusreformivaateita, vaikka eivät Yhdysvaltain PISA-tulokset globaalia talousmahtia mairitelleet. Selitykseksi tarjottiin historiallisesti vahvaa paikallistoimijoiden roolia Yhdysvaltojen kouluasioista päätettäessä (Henderson ym. 2015), minkä vuoksi osaamisen kansainvälisestä tasosta huolehtiminen ei kuulunut oikein kenenkään vastuualueisiin. Ekonomisti Eric Hanushek on ollut yksi utterimmista Yhdysvaltain heikkojen PISA-tulosten ja vahvan taloudellisen aseman välisen ristiriidan pohdiskelijoista. Eric Hanushek, Paul Peterson ja Ludger Woessmann (2013) halusivat teoksellaan *Endangering prosperity: A global view of the American school* vauhdittaa koulutusreformia esittelemällä laskelmiaan siitä vaihtoehdosta koituvista kustannuksista, jos Yhdysvalloissa reformien tielle ei lähdetä, ja maa jäisi siten jälkeen Suomen kaltaisille PISAn kärkimaille ennakoitavan talouskasvun tahdistusta. Vastaavasti kustannuksiltaan mittavatkin reformit maksaisivat kyseisten laskelmien mukaan helposti itsensä takaisin.

Christopher Tienken (2017) kritisoi otsikon ”*Economic Hocus-Pocus*” alla Hanushekin ja Woessmannin kaltaisten ekonomistien laskelmia ja tulkintoja PISA-testinumeroiden yhteydestä kansakuntien taloudelliseen menestykseen ja tulevaisuuteen. Martin Carnoy ja Richard Rothstein (2013) eivät usko PISA-tuloksilla perusteltuihin väitteisiin Yhdysvalloissa tarvittavista koulutusuudistuksista, koska myös heidän mukaansa PISA-testit ovat sillä tavoin laadittu, että niiden tuottamat tiedot eivät salli eivätkä mahdollista mielekkäiden politiikkajohtopäätösten vetämistä Yhdysvaltain tai minkään muunkaan maan koulutusjärjestelmän tuloksellisuudesta. Carnoy ja Rothstein (2013) eivät vastusta kansainvälisiä osaa-arviointoja sinänsä, mutta eivät pidä OECD:n suosittelua tapaa jäljitellä PISA kärkimaiden ”parhaita koulukäytäntöjä” lainkaan uskottavana. Yhdysvalloissa on kuitenkin käynnistetty ”Race to the Top”-ohjelmahanke edistämään maan PISA-menestystä.

Myös Tienken (2014) kritisoi käsitystä siitä, että PISA-testin pisteet ja sijoitukset olisivat arvioita maan koulutuksen tasosta, ennakoisivat kansantalouden tulevaa taloudellista menestystä tai mittaisivat sitä, kuinka valmiita 15-vuotiaat ovat kilpailemaan globaalissa taloudessa. Tienkenin (2014) mukaan PISAn tuottamat testipisteet ja sijoitukset erottelevat maita ensisijaisesti varallisuuden (köy-

hyden) sekä vanhempien koulutustason perusteella. Tienken muistuttaa myös siitä, että yhdysvaltalaisoppilaat eivät ole kertaakaan, vuodesta 1964 asti laadituissa kansainvälisissä testeissä sijoittuneet maailman kärkeen. Yhdysvaltain kansainvälinen kärkiasema varallisuudessa, kilpailukyvyssä, tieteesä ja tutkimuksessa on kaikesta huolimatta ollut koko ajan kiistaton. Teoksessaan *Defying Standardization* Tienken (2017, 34) tähdentää koulun ulkopuolisten tekijöiden tärkeyttä, siinä miten maat menestyvät PISAn kaltaisissa laajoissa kansainvälisissä arvioinneissa. Tienkenin mukaan lapsikäyhyyden aste selittää 46 prosenttia PISAn pistearvoista 30:ssä maailman huomattavimmista teollisuusmaista. Todettakoon, että UNICEFin (2013) mukaan Suomen ja Tanskan lapsikäyhysaste on alle 5 prosenttia, Norjan ja Saksan alle 10 prosenttia, kun taas Yhdysvalloissa luvut ovat 20–23 prosentin luokkaa.

Brittien PISA-suoriutuminen ei ole ollut mitenkään erinomaista. Aluksi PISA ei kiinnostanut brittimediaa (Grek 2009), mutta tilanne muuttui, kun vuoden 2006 PISAssa erot Englannin, Skotlannin, Walesin ja Pohjois-Irlannin välillä kasvoivat. Englannin PISA-pisteiden laskun tulkittiin osoittavan maan silloisen päähallituspuolueen epäonnistunutta koulutuspolitiikkaa (ks. Jerrim 2013). Samalla kun herättiin *best practice* -ajatteluun, uutta suuntaa alettiin etsiä PISAn kärkimaista. Yhdeksi mallimaaksi kelpuutettiin Suomi sen peruskoulua luonnehtivien erityispiirteiden vuoksi. Näitä olivat yliopistossa maisteriksi koulutetut opettajat, suhteellisen salliva opetussuunnitelma sekä se, ettei oppilaiden osaamista ollut tapana säännöllisesti alistaa standardoituihin testeihin. Suomen vuonna 2012 laskeneet PISA-pisteet ja sijoituksen putoaminen käänsivät brittikatseet Aasiaan; Singaporen ja Shanghain parhaista koulukäytännöistä lähdettiin etsimään Britannialle sopivia ratkaisuja. (Chung 2015.)

Irlannin huolenaiheet nousivat vuoden 2009 PISAssa, jossa maan pisteet lukutaidossa ja matematiikassa jäivät merkittävästi aiempia alemmiksi. Sen sijaan, että Irlannissa olisi ryhdytty valmistelemaan PISA-menestyksen suunnan kääntäviä koulutus uudistuksia, maan koulutusministeriö teetti yksityiskohtaisen selvityksen siitä, miten nimenomaan vuoden 2009 PISA toteutukseltaan erosi edellisistä. Irlantilaiselvitys paikansi ongelmia testattavien oppilaiden valinnassa, testisisällöissä, osaamisen arvioinnissa ja vastauskäyttötymisessä. Eri PISA-kierroksilla kerättyjen aineistojen vertailukelpoisuus todet-

tiin ongelmalliseksi, koska maata kulloinkin edustava testattavien oppilaiden joukko on arvioinnista toiseen vaihdellut kontrolloimattomasti. Toiseksi vertailtavuutta hankaloittavaksi tekijäksi todettiin testisisältöjen vaihtelu. Tarkastellessaan lukutaitotehtävien hyväksymisasteen ja standardoitujen pisteiden keskinäissuhteita irlantilaiset havaitsivat, että yhtä suuri hyväksymisasteen lasku ei muissa maissa ole johtanut Irlannin kaltaiseen laskuun standardoituissa pisteissä. (Cosgrove 2015; Cosgrove ja Cartwright 2014.)

Lopuksi

Suomalaiskansallisen koulutuspolitiikan keskeiset iskusanat ovat olleet koti, uskonto ja isänmaa (Kivinen 1988). Nykyisessä OECD:n tahdittamassa ylikansallisessa koulutuspolitiikassa lähdetään siitä, että koulutuksen ensisijainen tehtävä on valmistaa yksilöt ja kansakunnat kilpailemaan globaalissa tietoloudessa. PISAn kaltaisten ylikansallisten mittareiden ajatellaan kertovan siitä tasosta, missä eri maissa kulloinkin mennään.

Suomessa kuuluminen PISA kärkimaiden joukkoon on asiantuntijapiireissä luettu itsestään selvästi maisteritasoisen ja tiedeperustaisen opettajankoulutuksen ansioksi, mutta kun vuonna 2012 tulokset notkahtivat, opettajankoulutuksesta ei puhuttu mitään. Sen sijaan selityksiä notkahdukseen löydettiin poikien (ja maahanmuuttajataustaisten nuorten) heikosta opiskelumotivaatiosta ja puutteellisesta osaamisesta. Toisin kuin muualla, laskevat pisteetkään eivät Suomessa herättäneet juurikaan epäilyksiä PISAn luotettavuudesta, mutta poikien huonoudesta nouseva ”moraalipaniikki” tuli jäädäkseen. Vähänlaisesti on jaksettu pohtia sitä, olisiko itse koulun menettelytavoissa – mitä poikiin tulee – jotakin korjattavaa.

PISA-menestyksellä on hämmästyttävän voimakas vaikutus, ei vain koulutusasiantuntijoihin, vaan maassa käytävään julkiseen keskusteluun. Ani harva maan tilaa käsittelevä poliittinen puheenvuoro on voitu esittää julkisuudessa, eduskunnasta puhumattakaan mainitsematta PISAA, eikä kai siinä sinänsä mitään pahaa ole, vaikka kansakunnan identiteettiä rakennetaan myyttisillä puheilla kansallisen koulutuksen erinomaisuudesta. PISA-menestyksestä on tullut Suomen maabrändin keskeinen elementti.

PISA on ymmärretty näyttöön perustuvan poli-

tiikan airueksi ja PISA-tuloksista johdetut koulutuspoliittiset suositukset on voitu hyväksyä näyttöön perustuvan politiikan nimissä. Mutta kuten edellä olemme todenneet, PISA kyllä tuottaa runsaasti standardoitua numeerista tietoaineistoa, mutta sitä ei pidä samaistaa tieteellisesti tutkittuun tietoon. Toisin kuin usein luullaan, maiden sijoitusten nousu tai lasku PISA-rankingeissa ei kerro suoraan mitään siitä, mitä maan koulutuspolitiikassa on tehty tai jätetty tekemättä.

Yksi keskeinen huomio artikkelissamme on, että PISA-asiantuntijoita, yhtä hyvin kriitikoita kuin puolestapuhujiakin, näyttää vaivaavan vakava PISA-lukutaidottomuus. Todettakoon vielä, että PISAn rankingpisteistä ei tosiasiaa voi päätellä taitoeroja tyttöjen ja poikien välillä. Sen sijaan että suomalaispoikien arvioinnista toiseen vaihtelevat lukutaitopisteet tulkittaisiin osoituksiksi suomalaispoikien muuttuneesta lukutaidosta, ne pikemminkin osoittavat vain sen, että oppilaiden lukemistehtävistä suoriutumisen vertailu eri PISA-arviointien kesken ei hevin käy päinsä. Suomalaisoppilaiden testitulosten vaihtelu arviointikerrasta toiseen ei voi selittyä koulutusjärjestelmässä tosiasiaa tapahtuvilla muutoksilla, sillä koulutusjärjestelmässä ja varsinkin koulujen käytännöissä olennaiset asiat eivät kolmessa vuodessa ehdi juurikaan muuttua.

LÄHTEET

- Berliner, David. 2015. The many facets of PISA. *Teachers College Record* 117:1, 1–20.
- Breakspear, Simon. 2012. The policy impact of PISA: An exploration of the normative effects of international benchmarking in school system performance. *OECD Education Working Papers*, 71. Paris, France: OECD Publications.
- Breakspear, Simon. 2014. *How does PISA shape education policy making? Why how we measure learning determines what counts in education*. Centre for Strategic Education Seminar Paper, 240. <http://simonbreakspear.com/wp-content/uploads/2015/09/Breakspear-PISA-Paper.pdf> Viitattu 24.11.2017.
- Carnoy, Marten ja Rothstein, Richard. 2013. *What do international tests really show about U.S. student performance?* Washington, DC: Economic Policy Institute.
- Chung, Jennifer. 2015. International comparison and educational policy learning: Looking north to Finland. *Compare: A Journal of Comparative and International Education* 45:3, 475–479.
- Cosgrove, Jude. 2015. Changes in achievement in PISA from 2000 to 2009 in Ireland: Beyond the test scores. *The Irish Journal of Education / Iris Eireannach an Oideachais* 40, 29–44.
- Cosgrove, Jude ja Cartwright, Fernando. 2014. Changes in achievement on PISA: The case of Ireland and implications for international assessment practice. *Large-Scale Assessments in Education* 2:1, 1–17.
- von Davier, Matthias, Gonzalez, Eugenio ja Mislevy, Robert. 2009. What are plausible values and why are they useful? Teoksessa Matthias von Davier ja Dirk Hastedt (eds.), *IERI monograph series: Issues and methodologies in large-scale assessments*, Volume 2, 9–36. www.ierinstitute.org/fileadmin/.../IERI_Monograph_Volume_02_Chapter_01.pdf Viitattu 6.6.2017.
- Egelund, Niels. 2008. The value of international comparative studies of achievement: A Danish perspective. *Assessment in Education: Principles, Policy & Practice* 15:3, 245–251.
- Ertl, Hubert. 2006. Educational standards and the changing discourse on education: The reception and consequences of the PISA study in Germany. *Oxford Review of Education* 32:5, 619–634.
- Feniger, Yariv ja Lefstein, Adam. 2014. How not to reason with PISA data: An ironic investigation. *Journal of Education Policy* 29:6, 845–855.
- Fernandez-Cano, Antonio. 2016. A methodological critique of the PISA evaluations. *Revista Electrónica de Investigación y Evaluación Educativa* 22:1, art. M15. DOI: 10.7203/relieve.22.1.8806.
- Gorur, Radhika ja Wu, Margaret. 2015. Leaning too far? PISA, policy and Australia's 'top five' ambitions. *Discourse: Studies in the Cultural Politics of Education* 36:5, 647–664.
- Grek, Sotiria. 2009. Governing by numbers: The PISA 'effect' in Europe. *Journal of education policy* 24:1, 23–37.
- The Guardian. 6.5.2014. OECD and Pisa tests are damaging education worldwide – academics. <https://www.theguardian.com/education/2014/may/06/oecd-pisa-tests-damaging-education-academics> Viitattu 6.6.2017.
- Guglielmi, Sergio ja Brekke, Nancy. 2017. A framework for understanding cross-national and cross-ethnic gaps in math and science achievement: The case of the United States. *Comparative Education Review*. DOI: 10.1086/689656.
- Hanushek, Eric, Peterson, Paul ja Woessmann, Ludger. 2013. *Endangering prosperity: A global view of the American school*. Washington, DC: Brookings Institution Press.
- Harju-Luukkainen, Heidi, Nissinen, Kari, Sulkunen, Sari, Suni, Minna ja Vettenranta, Jouni. 2014. *Avaimet osaamiseen ja tulo- ja elämäntilanteeseen. Selvitys maahanmuuttajataustaisten nuorten osaamisesta ja siihen liittyvistä taustatekijöistä PISA 2012 -tutkimuksessa*. Koulutuksen tutkimuslaitos, Jyväskylän yliopisto.
- Helsingin Sanomat. 5.12.2016. Suomen Pisa-ihme hiipuu hiljalleen – professori kritisoi vertailun ”ylikansallista indikaattoripolitiikkaa” ja numeroilla hallitsemista.
- Helsingin Sanomat. 6.12.2016. Pisa-tulokset tulivat: Suomella ei syytä häpeään mutta alueelliset erot huolestuttavat – pääkaupunkiseudun koululaiset rynnivät muiden ohi.
- Henderson, Michael, Lergetporer, Philipp, Peterson, Paul, Werner, Katharina, West, Martin ja Woessmann, Ludger. 2015. Is seeing believing? How Americans and Germans think about their schools. *Ifo Working Paper*, 202.
- Jerrim, John. 2013. The reliability of trends over time in international education test scores: Is the performance of England's secondary school pupils really in relative decline? *Journal of Social Policy* 42:2, 259–279.
- Kivinen, Osmo. 1988. *Koulutuksen järjestelmäkehitys. Peruskoulutus ja valtiollinen kouludoktriini Suomessa 1800- ja 1900-luvuilla*. Turku: Annales Universitatis Turkuensis. Series C:67. University of Turku.
- Kivinen, Osmo, Hedman, Juha ja Nurmi, Jouni. 2016. Suomalaisaikuisten osaaminen väitettyä parempaa. *Janus* 24:2, 172–184.
- Kivinen, Osmo, Hedman, Juha ja Artukka, Kalle. 2017. Scientific publishing and global university rankings: How well are top publishing universities recognized? *Scientometrics* 112:1, 679–695.

- Komatsu, Hikaru ja Rappleye, Jeremy. 2017. Did the shift to computer based testing in PISA 2015 affect reading scores? A view from East Asia. *Compare: A Journal of Comparative and International Education*. DOI: 10.1080/03057925.2017.1309864.
- Korvela, Paul-Erik ja Vuorelma, Johanna (toim.). 2017. *Puhun niin totta kuin osaan. Poliittikka faktojen jälkeen*. Jyväskylä: Docendo.
- Kreiner, Svend. 2011. *Is the foundation under PISA solid? A critical look at the scaling model underlying international comparisons of student attainment*. Research report 11/1, Department of Biostatistics, University of Copenhagen.
- Kreiner, Svend ja Christensen, Karl Bang. 2014. Analyses of model fit and robustness: A new look at the PISA scaling model underlying rankings of countries according to reading literacy. *Psychometrika* 79:2, 210–231.
- Lingard, Bob. 2016. Rationales for and reception of the OECD's PISA. *Educação & Sociedade*, 37 (136), 609–627. <https://dx.doi.org/10.1590/es0101-73302016166670> Viitattu 23.10.2017.
- Malinen, Olli-Pekka; Väisänen, Pertti ja Savolainen, Hannu. 2012. Teacher education in Finland: A review of a national effort for preparing teachers for the future. *The Curriculum Journal* 23:4, 567–584.
- Marginson, Simon. 2014. University rankings and social science. *European Journal of Education* 49:1, 45–59.
- Mislevy, Robert. 1993. Foundations of a new test theory. Teoksessa Norman Frederiksen, Robert J. Mislevy ja Isaac I. Bejar (toim.), *Test theory for a new generation of tests*. New York-London: Routledge, 19–39.
- Murphy, Dan. 2014. Issues with PISA's use of its data in the context of international education policy convergence. *Policy Futures in Education* 12:7, 893–916.
- Murphy, Martin ja Schulz, Wolfram. 2006. *Sampling for national surveys in education*. Australian Council for Educational Research.
- NCES 2014. *NCES handbook of survey methods: Program for International Student Assessment (PISA)*. <https://nces.ed.gov/statprog/handbook/pdf/pisa.pdf> Viitattu 6.6.2017.
- OAJ 2016. Tasa-arvoinen peruskoulu on pelastettava nyt! <http://www.oaj.fi/cs/Satellite?c=Page&pagename=OAJWrapper&childpagename=OAJ%2FPPa%2Fsisalto&cid=1363787850492&browseContent=true&contentID=1408918197845> Viitattu 6.6.2017.
- OECD 2003. *Literacy skills for the world of tomorrow: Further results from PISA 2000*. Paris: OECD Publishing. DOI: 10.1787/9789264102873-en.
- OECD 2007. *PISA 2006: Science competencies for tomorrow's world, Vol. 1. Annex A8*. <https://www.oecd.org/pisa/39730305.pdf> Viitattu 6.10.2017.
- OECD 2009. *Learning mathematics for life: A perspective from PISA*. Paris: OECD Publishing. DOI: 10.1787/9789264075009-en.
- OECD 2010. *PISA 2009 Results: What students know and can do: Volume I*. Paris: OECD Publishing. DOI: 10.1787/9789264091450-en.
- OECD 2013. *PISA 2012 Results: Excellence through equity: Giving every student the chance to succeed: Volume II*. Paris: OECD Publishing. DOI: 10.1787/9789264201132-en.
- OECD 2014. *PISA 2012 Technical report*. Paris: OECD Publishing. <http://www.oecd.org/pisa/pisaproducts/PISA-2012-technical-report-final.pdf> Viitattu 6.6.2017.
- OECD 2015. *Universal basic skills: What countries stand to gain*. OECD Publishing. <http://dx.doi.org/10.1787/9789264234833-en> Viitattu 20.10.2017.
- OECD 2016. *PISA 2015 Results: Excellence and equity in education: Volume I*. Paris: OECD Publishing. DOI: 10.1787/9789264266490-en.
- OKM 2016. PISA 2015: Suomalaisnuoret edelleen huipulla, pudotuksesta huolimatta. http://minedu.fi/artikkeli/-/asset_publisher/pisa-2015-suomalaisnuoret-edelleen-huipulla-pudotuksesta-huolimatta Viitattu 6.6.2017
- Pearson 2014. Pearson to develop PISA 2018 Student Assessment 21st Century Frameworks for OECD. News announcement 10.12.2014. <https://www.pearson.com/corporate/news/media/news-announcements/2014/12/pearson-to-develop-pisa-2018-student-assessment-21st-century-fra.html> Viitattu 6.6.2017
- Prais, Sig J. 2003. Cautions on OECD's recent educational survey (PISA). *Oxford Review of Education* 29:2, 139–163.
- Ringarp, Johanna. 2016. PISA lends legitimacy: A study of education policy changes in Germany and Sweden after 2000. *European Educational Research Journal* 15:4, 447–461.
- Robitzsch, Alexander, Luedtke, Oliver, Koeller, Olaf, Kroehne, Ulf, Goldhammer, Frank ja Heine Joerg-Henrik. 2017. Challenges in estimations of trends in large-scale assessments: A calibration of the German PISA data. *Diagnostica* 63:2, 148–165.
- Rutkowski, Leslie. 2017. Design considerations for planned missing auxiliary data in a latent regression context. *Psychological Test and Assessment Modeling*, 59: 1, 55–70.
- Rutkowski, David. 2015. The OECD and the local: PISA-based test for schools in the USA, *Discourse: Studies in the Cultural Politics of Education*, 36:5, 683–699.
- Sahlberg, Pasi. 2015. *Finnish lessons 2.0: What can the world learn from educational change in Finland?* New York: Teachers College Press.
- Sahlberg, Pasi ja Hargreaves, Andy. 2015. The tower of PISA is badly leaning: An argument for why it should be saved. *The Washington Post* 24.3.2015. https://www.washingtonpost.com/news/answer-sheet/wp/2015/03/24/the-tower-of-pisa-is-badly-leaning-an-argument-for-why-it-should-be-saved/?utm_term=.d6de6eb9c628 Viitattu 6.6.2017.
- Salokangas, Maija ja Kauko, Jaakko. 2015. Borrowing Finnish PISA success? Critical reflections from the perspective of the lender. *Educação e Pesquisa* 41: n. especial, 1353–1364.
- Schatz, Monika. 2016. Engines without fuel? – Empirical findings on Finnish higher education institutions as education exporters. *Policy Futures in Education* 14:3, 392–408.
- Schleicher, Andreas. 2007. Can competencies assessed by PISA be considered the fundamental school knowledge 15-year-olds should possess? *Journal of Educational Change* 8:4, 349–357.
- Science Guide 29.8.2013. Schleicher's quantitative counterattack. <http://www.scienceguide.nl/201308/schleicher%E2%80%99s-quantitative-counterattack.aspx> Viitattu 6.6.2017.
- Schmidt, William ja Burroughs, Nathan. 2015. Puzzling out PISA: What can international comparisons tell us about American education? *American Educator* 39:1, 24–31.
- Sellar, Sam ja Lingard, Bob. 2013. The OECD and global governance in education. *Journal of Education Policy* 28:5, 710–725.
- Simola, Hannu. 2017. Suomalainen koulu on taitekohdassa. *Kanava* 1(45): 6–11.
- Simola, Hannu. 2015. *The Finnish Education Mystery: Historical and sociological essays on schooling in Finland*. London and New York: Routledge.
- Simola, Hannu. 2005. The Finnish miracle of PISA: Historical and sociological remarks on teaching and teacher education. *Comparative Education* 41:4, 445–470.

- Sjøberg, Svein 2016. OECD, PISA, and globalization: The influence of the international assessment regime. Teoksessa Christopher H. Tienken ja Carol A. Mullen (toim.), *Education policy perils: Tackling the tough issues*. New York and London: Routledge; 102–133.
- Soh, Kaycheng 2014. Finland and Singapore in PISA 2009: Similarities and differences in achievements and school management. *Compare* 44:3, 455–471.
- Soininvaara, Osmo. 8.12.2013. Vaikuttiko näytteen valinta Pisa-katastrofiin? <http://www.soininvaara.fi/2013/12/08/vaikuttiko-naytteen-valinta-pisa-katastrofiin/> Viitattu 6.6.2017.
- Takayama, Keita. 2008. The politics of international league tables: PISA in Japan's achievement crisis debate. *Comparative Education* 44:4, 387–407.
- Tienken, Christopher. 2017. *Defying standardization: Creating curriculum for an uncertain future*. Lanham, MD: Rowman & Littlefield.
- Tienken, Christopher. 2014. PISA problems. *AASA Journal of Scholarship and Practice* 10:4, 4–14.
- Times Educational Supplement. 24.3.2017. Exclusive: Pisa data may be incomparable, Schleicher admits. <https://www.tes.com/news/school-news/breaking-news/exclusive-pisa-data-may-be-incomparable-schleicher-admits> Viitattu 6.6.2017.
- Times Educational Supplement. 26.7.2013. Is Pisa fundamentally flawed? <https://www.tes.com/news/tes-archive/tes-publication/pisa-fundamentally-flawed> Viitattu 6.6.2017.
- UNICEF 2013. *Child well-being in rich countries: A comparative overview*. Innocenti Report Card 11, UNICEF Office of Research, Florence https://www.unicef-irc.org/publications/pdf/rc11_eng.pdf Luettu 6.6.2017
- Varjo, Janne, Simola Hannu ja Rinne, Risto 2016. *Arvioida ja hallita. Perään katsomisesta informaatio-ohjaukseen suomalaisessa koulupolitiikassa*. Kasvatusalan tutkimuksia, Nro 70, Suomen kasvatustieteellinen seura, Helsinki.
- Vettenranta, Jouni, Välijärvi, Jouni, Ahonen, Arto, Hautamäki, Jarkko, Hiltunen, Jenna, Leino, Kaisa, Lähteinen, Suvi, Nissinen, Kari, Nissinen, Virva, Puhakka, Eija, Rautapuro, Juhani ja Vainikainen, Mari-Pauliina. 2016. *PISA 2015 ensituloksia. Huipulla pudotuksesta huolimatta*. Opetus- ja kulttuuriministeriön julkaisuja 2016: 41.
- Välijärvi, Jouni, Linnakylä, Pirjo, Kupari, Pekka, Reinikainen, Pasi ja Arffman, Inga. 2002. *The Finnish success in PISA – and some reasons behind it. PISA 2000*. Koulutuksen tutkimuslaitos. Jyväskylän yliopisto.
- Weirich, Sebastian, Haag, Nicole, Hecht, Martin, Böhme, Katrin, Siegle, Thilo ja Lüdtke, Oliver. 2014. Nested multiple imputation in large-scale assessments. *Large-scale Assessments in Education* 2:1, 1–18.
- Waldow, Florian. 2017. Projecting images of the 'good' and the 'bad school': Top scorers in educational large-scale assessments as reference societies. *Compare: A Journal of Comparative and International Education*. DOI: 10.1080/03057925.2016.1262245.
- Wuttke, Joachim. 2007. Uncertainties and bias in PISA. Teoksessa Stefan Hopman, Gertrude Brinek ja Martin Retzl (toim.), *PISA according to PISA: Does PISA keep what it promises?* Wien: LIT, 265–294.
- Zwitser, Robert, Glaser, Sjoerd ja Maris, Gunter. 2017. Monitoring countries in a changing world: A new look at DIF in international surveys. *Psychometrika* 82:1, 210–232.

TILASTOLÄHTEET

- Maailmanpankki 2017. Maailman Pankin indikaattorit: nuorten lukutaito. <http://data.worldbank.org/indicator/SE.ADT.1524.LT.ZS> Viitattu 6.6.2017.
- OECD 2017a. PISA 2015 Database. Compendia (compressed). Cognitive items. <http://www.oecd.org/pisa/data/2015database/> Viitattu 6.6.2017.
- OECD 2017b. PISA Data Explorer. <http://pisadataexplorer.oecd.org/ide/idepisa/> Viitattu 6.6.2017.
- UNESCO 2017. UNESCO:n koulutustilastot: lukutaidoton väestö. <http://data.uis.unesco.org/Index.aspx?queryid=166> Viitattu 6.6.2017.