

Tilastollinen järkeily ja oppilasarviointi suomalaisen kasvatustieteen historiassa Ian Hackingin tieteenfilosofian näkökulmasta

Antti Saari

Artikkelissa tarkastellaan tilastollisten tutkimusmenetelmien merkitystä suomalaiselle kansanopetuksen oppilasarvioinnille. Analyysi rajataan erityisesti toisen maailmansodan jälkeisille kolmelle vuosikymmenelle 1940-luvun lopusta 1970-luvun loppuun. Tänä aikana modernit tilastolliset menetelmät muodostuivat paradigmaattisiksi kasvatustieteellisen tutkimuksen metodeiksi. Samalla ne toimivat koulujärjestelmän rationaalisen suunnittelun välineinä, joilla pystyttiin tekemään mitattaviksi ja arvioitaviksi koulutuksen populaatiotason ilmiöitä.

Aluksi

Arviointia, jossa yhteiskunnallisten rakenteiden toimintaa ja kehittämistä tarkastellaan tieteellisen tarkkojen kriteerien valossa, voidaan pitää rationaaliseen suunnitteluun uskovan modernin hallinnan perustana. Ulf Lundgren (2002, 103–104) pitääkin Chicagon vuoden 1933 maailmannäyttelyn mottoa ”Tieteet tutkivat, teknologia toteuttaa, ihmiskunta mukautuu” sopivana myös tiedeperustaisen kouluarvioinnin eetokselle. Arviointi on perinteisesti ollut modernien yhteiskuntien rituaali, jossa hallinta julistaa omaa rationaalisuuttaan.

Tänäkin päivänä arviointi toimii keskeisenä tiedon ja hallinnan polttopisteenä koulutuksen alueella. Arvioinnista on tullut maailmanlaajuista: se kykenee ylittämään suuria maantieteellisiä etäisyyksiä ja standardoimaan koulutusta sen eri tasoilla. ”Neljännen tasavallan” Suomessa, jossa valtiollinen, lakeihin ja säädöksiin perustuva ohjaus on menettämässä merkitystään globaalin kapitalistisen talouden rinnalla, on arvioinnista tulossa keskeinen yhtenäistävä tekijä sekä kansallisen kilpailukyvyn indikaattori. (Rinne ym. 2011, 13.) Koulujen tuloksellisuuden ja laadun arviointia voidaan pitää myös osana koulutuspolitiikan niin sanottua *evidence based policies* -trendiä. Siinä poliittiset toimenpiteet pyritään perustelevaan mahdollisimman ”kovan” tilastollisen evidenssin perusteella. (Oakley 2002.) Toisinaan esitetään, että tiedeperustainen, eksakti arviointi on myös yhteiskunnallisen tasa-arvon toteutumisen keskeinen ehto (ks. esim. Wiliam 2010).

Toisaalta koulutuksen arviointi on herättänyt myös paljon kritiikkiä. Esimerkiksi Aini Oravakangas ja Tomi Kiilakoski näkevät, että vallitseva arvioinnin eetos käsittää koulun tuotantokoneistona, jonka toimintaa se pyrkii mahdollisimman tarkasti mittaamaan ja tehostamaan. Koska koulutusta pidetään nykypäivän yhteiskunnassa kansallisen kilpailuky-

vyn avaintekijänä, korostuu sen rooli välineenä, jota tulee kontrolloida mahdollisimman yksityiskohtaisesti ja tehokkaasti. (Kiilakoski & Oravakangas 2010.) Kasvatustieteen filosofi Gert Biesta (2007) puolestaan väittää, että standardoidut arvioinnin käytännöt esiintyvät tiedepuolesta ja objektiivisina, mutta samalla ne tekevät arvioinnin taustalla olevat poliittiset kamppailut näkymättömiksi. Näin ollen arvioinnista, jonka tulisi olla avoimen yhteiskunnan keskeinen itsesäätelymekanismi, tuleekin pohjimmiltaan epädemokraattista.

Erityisesti arviointiin liittyvää tilastollisen testauksen kulttuuria on viime vuosina kritisoitu voimakkaasti. Kenties huomiota herättävin puheenvuoro on ollut yhdysvaltalaisen professorin Diane Ravitchin kirja *The Death and Life of the Great American School System* (2010), jossa George W. Bushin hallinnon *No Child Left Behind* -ohjelman kehittämiseen osallistunut kasvatustieteilijä itse ottaa vahvasti kantaa testaukseen perustuvaa järjestelmää vastaan. Ohjelman tuli kehittää tasa-arvoisempaa koulutusjärjestelmää, mutta koulun tuloksellisuusmittauksiin ja testauksiin perustuva systeemi toimikin keinona rangaista opettajia oppilaidensa huonosta menestyksestä ja luoda tilaa *charter school* -tyyppisille yksityisille kouluille – jotka eivät nekään ole tuottaneet julkisia kouluja parempia oppimistuloksia.

Arvioinnissa on historiallisesti ollut keskeistä erityisesti tilastotieteellinen informaatio, joka on synnyttänyt aivan erityisen tavan tuottaa yhteiskuntaa ja ihmistä koskevia tieteellisiä faktoja. Ilman sitä ei olisi valtavaa byrokraattista koneistoa, joka mahdollistaa koulutusjärjestelmänkin hallinnan. (Ks. Hacking 1991, 181.) Se on luonut arvioinnille uskottavuutta nostamalla sen tieteellistä arvovaltaa ja samalla tehnyt arvioinnista objektiivista asiantuntijatyötä, joka voi toimia koulutusjärjestelmän hallinnan legitiiminä perustana.

Kuitenkin tiedonsosiologinen ja tieteenfilosofinen kritiikki, jossa analysoidaan arviointia erityisesti tilastollisena tiedontuotantona, on ollut harvoin esillä koulutuksen arviointia koskevassa keskustelussa. Tässä artikkelissa tilastollisiin menetelmiin perustuvaa kansanopetuksen arviointia tarkastellaan tieteenfilosofisesti ja -historiallisesti erityisenä järkeilyn tyylinä. Tällä Ian Hackingin käyttämällä termillä viitataan käsitteiden, ontologisten ja epistemologisten olettamusten, tutkimuskäytäntöjen ja -objektien eheään kokonaisuuteen, jossa yleispeivänä pidetty tieteellinen tieto tulee mahdolliseksi.

Tarkoitukseni ei ole kuvata laajasti oppilasarvioinnin historiaa, vaan tuoda tieteenfilosofia ja -historiallisia näkökulmia koulutuksen arvioinnista käytävään kriittiseen keskusteluun. Kiinnitän huomioni erityisesti siihen, miten tilastolliseen järkeilyyn perustuva tieteellinen oppilasarviointi rajaa objektiivisen tiedon ehtoja ja kuinka se samalla pyrkii olemaan koulutuksen hallinnan kannalta merkityksellistä. Lisäksi tarkastelen, miten arviointi kykenee muuttamaan koulun tiedon ja hallinnan ympäristöjä.

Laajojen aineistojen sijasta rajaan tarkasteluni toisen maailmansodan jälkeisille kolmelle vuosikymmenelle. Tällä aikakaudella kehittyi paitsi moderni empiirinen kasvatustiede myös ne tilastolliset menetelmät, joiden pohjalta koulutuksen arviointia tänäkin päivänä harjoitetaan. Siihen saakka epäsystemaattista, opettajien omaan harkintaan jätettyä oppilasarviointia pyrittiin muuttamaan valtakunnallisesti yhtenäiseksi ja objektiiviseksi. Raimo Kontinen (1995, 6–7) pitää tätä merkittävänä uudistuksena, jonka yhteydessä kasvatustiede pyrki yhtäältä nostamaan statustaan ”tieteellisemmäksi” ja toisaalta tekemään itsestään koulutuksen hallinnan kannalta merkityksellistä.

En pyri tarkastelemaan suomalaista oppilasarviointia tyhjentävästi kyseisellä aikakaudella. Sen sijaan käytän harkittuja näytteitä – arviointia käsitteleviä tiedeartikkeleita, menetelmäoppaita ja testiohjeita – joilla havainnollistan tilastollisen järkeilyn tyyliä koskevia ilmiöitä suomalaisessa oppilasarvioinnissa. Aineisto kontekstoidaan erityisesti tilastotieteellisten menetelmien historiaan ja epistemologiseen perustaan.

Mahdollisena vastaväitteenä tämäntapaiselle tarkastelulle voidaan toki esittää, että tällöin analyysi ainoastaan löytää sen, mitä se on olettanutkin, eli tilastollisen järkeilyn tyylin oppilasarvioinnin perustana. Kyseessä ei kuitenkaan ole tämänkaltainen kehäpäätelmä, vaan pikemminkin hermeneuttinen kehä: lähtökohtana on Hackingin ajattelun pohjalta rakennettu analyysikehikko, jonka pohjalta ei pyritä pelkästään toteamaan järkeilyn tyylin olemassaoloa, vaan problematisoidaan sen pohjalta suomalaista koulutuksen arviointia koskevia epistemologisia ja ontologisia lähtökohtia. Toisin sanoen en kysy, onko tilastollista järkeilyn tyyliä arvioinnissa olemassa, vaan tarkastelen, miten se toimii ja minkälaisia arviointia koskevan keskustelun kannalta hedelmällisiä kysymyksiä sen tarkastelu voi herättää.

Etenen esittelemällä aluksi Ian Hackingin järkeilyn tyylin käsitettä ja hänen tapaansa hahmottaa sen avulla tilastollisen tutkimuksen historiaa ihmistieteissä. Tämän jälkeen tarkastelen suomalaisen oppilasarvioinnin piirteitä erityisesti Hackingin hahmottelemien järkeilyn tyylin piirteiden valossa. Kuvaan, miten tilastolliset objektiivisuuden kriteerit arvioinnissa pyrkivät muokkaamaan koulujen oppilasarvioinnin käytäntöjä. Tilastotieteellisiin käytäntöihin pohjaavan arvioinnin osoitetaan muodostuvan elimelliseksi osaksi toisen maailmansodan jälkeistä rationaalisen suunnittelun ideologiaa kansanopetuksen hallinnassa. Tämän jälkeen tarkastelen, millä tavoin tilastollisten tekniikoiden pohjalta luodaan uusia arvioitavia ryhmiä, jotka puolestaan taas mahdollistavat oppilaiden uudenlaisen erotelun ja ryhmittelyn koulun arjessa. Erityisenä ryhmien luomisen perustana tarkastellaan normaalin kategorialta, jonka pohjalta esitetään suhteellisen arvioinnin toteuttamista ja koko koulujärjestelmän laajuista arvioinnin yhtenäistämistä.

Hackingin näkökulman valossa voidaan näin osoittaa, miten tieteellisen tiedon ehdot ovat yhteydessä koulutuksen hallinnan käytäntöihin. Tilastolliset tekniikat eivät siis ole ainoastaan koulutodellisuudesta vieraantuneita menetelmiä, vaan osa tapoja ajatella ja toimia koulumaailmassa. Ne määrittelevät ehtoja sille, miten oppilaista voidaan tuottaa objektiivista tietoa sekä samalla toteuttaa yksilöllisen ja oikeudenmukaisen arvioinnin periaatteita.

Tilastollinen järkeilyn tyyli

Ian Hacking (s. 1936) on kanadalainen tieteenfilosofi, joka on saanut vaikutteita paitsi angloamerikkalaisista tieteenfilosofian traditioista myös Michel Foucault'n ihmistieteiden historian tutkimuksesta. Tämä analyttisen ja mannermaisen tradition yhdistäminen tekee Hackingin ajattelusta erityisen mielenkiintoisen. Hackingin lähestymistapaa ihmistieteiden filosofiaan voidaan nimittää historialliseksi ontologiaksi. Tunnetusti ontologia tutkii sitä, mikä on olemassa ja minkälaisia olemassaolon lajeja voidaan määritellä. Hacking väittää kuitenkin historiallistavansa ontologian: hän siis tutkii, kuinka inhimillisen olemassaolon lajit syntyvät ja katoavat erityisissä historiallisissa yhteyksissään ja miten ne liittyvät ihmistieteelliseen tutkimukseen. (Hacking 2002, 1–9.)

Järkeilyn tyylin (*style of reasoning*) idean Hacking on lainannut tieteenhistorioitsija A.C. Crombielta, joka kirjoitti ”tieteellisen ajattelun tyyleistä”. Tyyllillä ei niinkään haluta viitata tieteellisen tutkimuksen tuloksiin ja löydöksiin kuin erityisesti tutkimuksen käytäntöihin, siihen millä keinoilla tiede katsoo saavuttavansa totuuden. Kyseessä ei ole yksittäinen tieteenalan ”paradigma” tai tutkimusohjelma, vaan jopa useita vuosisatoja kattava argumentaation, luokittelun, tieteellisten teorioiden ja mallien yhdistelmä. Hacking kuitenkin puhuu ajattelun sijasta mieluummin järkeilystä, sillä ajattelu herättää näet helposti

lähinnä kognitiivisiin, tutkijoiden mielensisäisiin prosesseihin liittyviä miellelyhtymiä. (Hacking 2002, 179–183.)

Hackingin ajattelu yhdistyy tällä tavoin 1960-luvulla angloamerikkalaisessa keskustelussa esille nousseeseen tieteen perinpohjaisen historiallisuuden korostamiseen. Siksi Hackingin tuotantoa on toisinaan pidetty myös kasvatustieteissä erityisen mielekkäänä tieteenkriitikin välineenä. Se historiallistaa ja suhteellistaa ikuisina ja paikattomina pidetyt kasvatustieteelliset totuudet ja järjen muodot (Peters 2007).

Hackingin mukaan järkeilyn tyylin voi tunnistaa siitä, että se tuottaa tieteiden kentälle uusia objekteja, väitelauseita sekä niitä koskevan evidenssin ja verifikaation muotoja, ilmiöitä koskevia lakeja tai säännönmukaisuuksia sekä niitä koskevia luokittelujärjestelmiä. Puhuessaan tieteellisten väitelauseiden verifikaatiosta ja niitä koskevasta evidenssistä Hacking korostaa, että jokainen järkeilyn tyyli osoittaa ne keinot, joilla tieteellisten väittämien oikeellisuutta tai erheellisuutta voidaan koetella. Järkeilyn tyylit tuottavat myös aivan uudenlaisia tutkimuksen kohteita, joiden olemassaolosta käydään jatkuvaa kamppailua: voidaanko sanoa, että laboratoriokeissa tarkasteltavat asiat, joita ei voida suoraan havaita, ovat todella olemassa? Koostuuko ihmisen älykkyys yhdestä vai useammasta tekijästä? Niin ikään tavoista tuottaa tieteellistä evidenssiä käydään kamppailua: onko esimerkiksi mahdollista tuottaa varmaa tietoa kokeellisin menetelmin, vai onko deduktio metafysisistä ensimmäisistä periaatteista ainoa oikea tieteellinen menetelmä? (Hacking 2002, 189–191.)

Tilastollisen järkeilyn tyylin historiaa

Järkeilyn tyyliä koskevat kriteerit kuitenkin jäävät hämäräksi, ellei niitä tarkastella konkreettisten historiallisten esimerkkien kautta. Hacking itse on kirjoittanut paljon tilastollisesta järkeilyn tyylin historiasta (Hacking 1975; 1990; 1991), ja osaa hänen tuotannostaan on syytä esitellä lyhyesti myös tämän artikkelin kohteen kannalta. Tilastolliselle järkeilyn tyylin syntymiselle keskeistä on ollut ajatus todennäköisyyteen perustuvasta evidenssistä, mikä on vaikuttanut olennaisesti myös ihmistä tutkivien tieteiden kehitykseen. Vuosisatojen takaiset epistemologiset kysymykset saattavat vaikuttaa käsillä olevan aiheen kannalta etäisiltä, mutta jotta myös koulutuksen arviointia koskevia, erityisesti evidenssin ilmeisyyden ja teoriavapauden ongelmia voidaan ymmärtää, on syytä hieman tarkastella myös niiden historiallista muotoutumista.

Ennen modernien luonnontieteiden syntyä todennäköisyys (lat. *probabilitas*) ei ollut vakavasti otettavan tiedon muoto. Oli näet olemassa varmaa, aristoteelisen scientian tietoa, jota voitiin loogisesti johtaa kaikkea olemassaoloa koskevista metafysisistä olettamuksista. Todennäköisyys koski ennen kaikkea merkkejä: esimerkiksi lääketiede (siinä missä samalla tiedollisella tasolla ollut alkemiakin) tutki sairauden merkkejä kasvoilta, silmistä ja eritteistä. Nämä havaitut ilmiöt eivät itsessään olleet sairautta, vaan ainoastaan enemmän tai vähemmän todennäköisiä merkkejä sairaudesta. Siksi ne tuottivat varsin epävarmana pidettyä tietoa. Vasta kun tämä merkkien tulkinta muodostuu evidenssiksi ja irtoaa aristoteelisista ensimmäisistä periaatteista, saattaa tilastollinen ja sen mukana induktiivinen tieto muodostua tieteiden perustaksi. Uudessa kokeellisessa luonnontieteessä tieto ei enää ollut suuriin auktoriteetteihin, Aristoteleen ja Tuomas Akvinolaisen kaltaisiin menneisyyden mestareihin viittaavaa. Nyt tieteestä tuli ensinnäkin nykyhetkessä välittömästi havaittuun liittyvää tietoa. Tiedon muodostuksessa ei voinut enää viitata historiaan, vaan nykyhetkellä kaikkien havaittavissa oleviin ilmiöihin. Siinä missä vanhat arvovaltaisimmat tieteet sisälsivät kaiken tarvittavan tiedon (josta erityistapauksia koskevat johtopäätökset voitiin pitä-

västi dedusoida), nyt tiedosta tuli täydentyvää. Se ei enää sisältänyt aristoteelisten tieteiden loogista varmuutta, vaan se oli ainoastaan enemmän tai vähemmän todennäköistä. Lisäksi todennäköisyyteen perustuva tieto vaikutti pohjaavan teoreettisista ja filosofisista oletuksista vapaisiin havaintoihin. Ne vaikuttivat siis olevan poliittisista ideologioista riippumattomia. (Hacking 1975; ks. myös Shapin & Schaffer 1985.) Tiedonmuotojen status siis kääntyi ympäri: evidenssistä ja induktiosta tuli nyt todellista tietoa, ja metafyyminen tieto ajautui pelkästään mielipiteiden alueelle (Hacking 1975).

1800-luvulla tapahtui Hackingin mukaan kaksi tilastollisen järjelyn tyylin kannalta merkittävää muutosta. Ensinnäkin newtonilainen mekanistinen tapa ymmärtää maailmaa ja samalla yhteiskuntia murtui. Todennäköisyyden hallitsemassa tieteellisessä maailmankuvassa ihmisen elämää eivät enää määränneet mekaaniset syy- ja seuraussuhteet, vaan juuri todennäköisyydet. Tämän mukana tapahtui toinen muutos erityisesti yhteiskunnallisessa vallankäytössä. Biovallaksi on kutsuttu yhteiskunnan hallintaa, joka kohdistuu populaatio-tason ilmiöihin, kuten varallisuuteen, työllisyyteen, kansanterveyteen ja koulutukseen. Eurooppalaisten valtioiden keskeiseksi voiman tunnukseksi muodostui ”väestön” hyvinvointi terveyden, työllisyyden ja syntyvyyden alueilla. Näitä ilmiöitä puolestaan kuvattiin tilastollisten tekniikoiden avulla, jotka kykenivät kuvaamaan laajoja ja monisyisiä sosiaalisia ilmiöitä yksinkertaisin suurein. (Hacking 1990, 115–124.)

Nämä molemmat muutokset ilmenevät statistiikassa eli valtiota koskeva opissa, joka oli vielä 1800-luvun alkuun saakka lähinnä sanallisesti kuvattua tietoa valtion tilasta. Kuitenkin 1800-luvulla Ranskassa alettiin statistiikka ymmärtää erityisesti numeraalisena kuvauksena. Kuten Hacking toteaa, tämäkin oli varsin kiistanalainen, ihmiskuvaan liittyvä siirtyminen, sillä Saksassa akateemiset statistiikan harjoittajat pilkkasivat ”tabellinikkareita”, jotka kavensivat ihmisen ja kansakunnan vapaan sekä ainutlaatuisen toiminnan ainoastaan numeroin kuvattaviksi säännönmukaisuusiksi. (Alastalo 2005, 30–31; Hacking 1990; Töttö 2000.)

Kyseisen aikakauden suomalaisen tilastotieteen merkkipaaluna voidaan pitää Suomen tilastollisen keskusviraston perustamista vuonna 1864. [1] Vuonna 1848 Johan Vilhelm Snellmankin puolusti noihin aikoihin vielä nuorta tieteenalaa, jota pidettiin usein tylsänä ja pitkävetisenä:

Mutta itse asiassa harva tutkimus on yhtä monipuolista ja tarjoaa enemmän ajattelemisen aihetta, puhumattakaan siitä, että tarkempien tietojen saaminen isänmaan nykyisestä niin fyysisestä kuin henkisestäkin tilanteesta on isänmaataan rakastavalle aina erittäin tärkeä asia. Maan maantieteelliset olot, kansan luonnonolosuhteet, valtiojärjestys ja valtiohallinto, lainkäyttö, kirkolliset olot, lopulta kansakunnan koko fyysinen ja henkinen kulttuuri – kaikki se on tilastotieteen kohteena. (Snellman 2004 [1848], 100.)

Snellman siis näkee, että tilastojen kautta voidaan tuoda näkyväksi koko kansakunnan tila. Tilastojen käyttö muodostuu myös ehdoksi kansakunnan vaurastumiselle. Sen kautta voidaan saada tietoa kaikista väestön varallisuuden, sivistyksen, terveyden ja poliittisen järjestyksen kannalta olennaisista seikoista, ja sen kautta niihin voidaan myös suunnitelmallisesti vaikuttaa. Lisäksi numeroihin perustuvan objektiivisen statistiikan etuna on sen suhteellinen riippumattomuus poliittisista intresseistä sekä teoriasta:

Vaikka teoria aina vaikuttaakin yhteiskuntajärjestykseen, niin jatkuvat ja tarkat havainnot antavat kuitenkin kokemukselle sellaista lujutusta ja varmuutta, joka terveellisesti rajoittaa teorian usein vaarallista ylivaltaa. Varma ja ylei-

sesti levinnyt tieto yhteiskunnan tilasta kunakin hetkenä vaikuttaa varmasti myös yhteiskunnan sisäisen rauhan säilymiseen. Sen paremmin hallittujen toiveet kuin hallitsevien toimenpiteekään eivät nimittäin voi vastustaa selkeiden tosiasioiden todistusvoimaa. Myös kaikkien uudistusten on perustuttava kiistattomiin tosiasioihin, jos uudistuksia vaativat ja toteuttavat ylipäänsä pyrkivät toimimaan järjestelmällisesti. (Snellman 2004 [1848], 101.)

Snellmanin ajattelussa on siis nähtävissä nykypäivänäkin laajasti vallitseva käsitys, jonka mukaan juuri tilastot kykenevät ylittämään vallitsevat erimielisyydet yhteiskunnan tilasta ja suunnasta. Tosiseikat ovat itsessään kiistattomia ja arvovapaita, ja juuri siksi ne voivat toimia politiikan ei-poliittisena perustana (Poovey 1998).

Numeraalisen, suuria populaatioita koskevan tieteellisen kuvauksen mahdollisti 1800-luvulla kehittynyt aivan erityinen tapa tuottaa induktiivista tietoa. John Stuart Millia pidetään toki yhtenä keskeisenä induktiivisen menetelmän teoreetikkona, mutta tilastollisten käytäntöjen kannalta Milliäkin merkittävämpi lienee Siméon-Denis Poisson, joka esitti vuonna 1835, että kaikenlaiset asiat ovat alisteisia niin sanotulle ”suurten lukujen” universaalille laille (joka tunnetaan myös Bernoullin lakina). Tämä tarkoittaa, että kunhan tutkija havainnoi tarpeeksi suurta määrää samankaltaisia tapahtumia, hän tulee huomaamaan, että tapahtumia osoittavien lukujen suhde on kutakuinkin vakio. (Hacking 1990, 95.) Tämä antoi uskoa siihen, että yhteiskuntatiede pystyy paitsi luotettavasti tutkimaan yhteiskunnallisia säännönmukaisuuksia myös vaikuttamaan niihin (Hacking 1990, 109). Näin muodoin suurten lukujen lakia voidaan pitää yhtenä rationaalisen suunnittelun keskeisenä mahdollisuusehtona.

Olennaista induktiivisessa järkeilyssä on myös se, että havainto tapahtuu näennäisesti ilman havainnoivaa subjektia: havaitseminen suoritetaan useamman henkilön voimin ja tarkalleen samoilla menetelmillä. Näin voidaan saavuttaa puolueeton ”kolmannen näkökulma” eli kaikki väitteet saattoivat saada selkeän intersubjektiiivisen merkityksen. (Swijtink 1987; Gigerenzer 1987.) Kollektiivinen todistus oli uudenlainen tiedon muoto myös kasvatustieteessä. Pedagogista tietoa ei enää muodostettu kirjoituspöydän ääressä, vaan siitä muodostui kollektiivisesti organisoitu, kumuloituva projekti, jossa jokainen tutkija saattoi tuoda osan yhteiseen tiedon varantoon. Esimerkiksi Ernst Meumannin jättimäinen, kolmiosainen *Vorlesungen* (1907) ja G. Stanley Hallin lähes yhtä mittava *Adolescence* (1904) muodostivat uudenlaisen kasvua ja oppimista koskevan kansainvälisten tutkimustulosten atlaksen, johon tulevaisuuden tutkijat tulisivat lisäämään oman osuutensa. (Saari 2011.) Tunnetusti esimerkiksi Hallin *Child Study* -ohjelmassa nojaututtiin juuri laajoihin havainnointiohjelmiin, joissa opettajia ympäri Yhdysvaltoja valjastettiin havainnoimaan lasten käyttäytymistä ja mielen toimintaa. Näiden perusteella ajateltiin voitavan rakentaa induktiivista tietoa lapsista.

Tilastollisen järkeilyn tyylin nousu suomalaisissa yhteiskunta- ja käyttäytymistieteissä

Arvioidessaan suomalaisen psykologian historiaa Kullervo Rainio (2006, 240) kiinnittää erityistä huomiota psykologisen tutkimuksen metodifetisismiin: pelkästään tilastollisia menetelmiä käyttämällä uskottiin voitavan tuottaa paitsi sosiaalisesti merkityksellistä, myös luonnontieteiden kaltaista varmaa tietoa. Suomesta muodostuikin 1940-luvun lopulta alkaen faktorianalyysin luvattu maa. Menetelmän esittelijöinä voidaan pitää Toivo Vahervuota ja Yrjö Ahmavaaraa, jotka kirjoittivat aiheesta paljon käytettyjä oppikirjoja (ks.

Ahmavaara 1957; Vahervuo & Ahmavaara 1958; Vahervuo 1952). Usein juuri faktorianaalyttiset menetelmät yhdistetään suomalaisen yhteiskunta- ja käyttäytymistieteellisen tutkimuksen metodologian kapenemiseen (Alastalo 2005, 80–85). Tilastolliset tekniikat, erityisesti faktorianaalyysi, muodostivat ikään kuin yhteiskuntatutkimuksen yleispätevän menetelmän ja tiedeyhteisön kommunikaation yhteisen kielen.

Sama tendenssi näkyi myös suomalaisessa kasvatustieteessä. Joel Kivirauma on havainnut väitöskirjojen menetelmien muuttuneen 1960-luvulla varsin kaavamaisiksi. (Kivirauma 1998, 16–17, taulukko 3.) Sirkka Ahosen (1998) mukaan lähinnä tilastotieteellisten metodien hallintaa osoittavien tutkimusten välitön relevanssi kasvatukseen ja koulutukseen maailmalle jäi hyvin etäiseksi. Tämä vaikutelma varmasti saadaankin, mikäli tarkastellaan ainoastaan raportoituja tutkimustuloksia.

Kuitenkin Hackingin järkeilyn tyylin analyysin valossa tilastotieteellinen tutkimus kietoutuu hyvin vahvasti osaksi suomalaista kansanopetuksen hallintaa. Tilastotieteellinen teknologia kasvatustieteellisen tutkimuksen yhtenäisenä käytäntönä tarjoaa välineet arvioida sen piirissä esitettyjen väitteiden ja tutkimustulosten paikkansapitävyttä. Näin voidaan kyseenalaistaa oletus ”positivistisen” kasvatustieteen vieraudesta suhteessa käytännön kouluun maailmaan.

Koska tilastollisten menetelmien yhtenä keskeisenä airuena pidetään Yrjö Ahmavaaraa, on syytä katsoa lähemmin, miten hän esittelee niiden (ja erityisesti faktorianaalyysin) merkitystä ihmisistä tutkiville tieteille 1950-luvun kirjoituksissaan. Hän valittaa, että psykologian ja sosiologian kaltaisilla tieteenaloilla vallitsee vielä 1950-luvulla täysi sekaannus ja niiden tuottama tieto on epävakaa ja epävarmaa. (Ahmavaara 1957a, 5, 27; Ahmavaara 1957b, 6, 16, 18.)

Ahmavaara julistaa, että faktorianaalyttisin menetelmin voidaan kuitenkin saada järjestyttä epävakaisuuteen ja epävarmuuteen psykologian ja sosiologian kaltaisilla tieteenaloilla. Faktorianaalyysissä ihmistieteellisiä käsitteitä tuotetaan operationaalisesti. Percy Bridgmanilta tutun määritelmän mukaan operationalismi tarkoittaa käsitteen määrittämistä sen mittauksen operaationa. (Ahmavaara 1957a, 3, 22). Faktorianaalyysissä käsitteitä voidaan jalostaa mittaamalla havaittujen ilmiöiden korrelaatioita. Tällä tavoin voidaan sivuuttaa kielen ja maailman vastaavuutta koskeva iänikuinen epistemologinen ongelma: ”Vasta kokeellisesti todettuihin yhteyksiin voidaan sitten perustaa myös sanallisia määritelmiä: Ne testit, jotka tuottavat yhdenmukaisia tuloksia, mittaavat ilmeisesti samaa asiaa, ja tälle voidaan etsiä sopivaa nimitystä tarkastelemalla, mitä yhteisiä piirteitä ao. testien suoritukseen liittyy.” (Ahmavaara 1957a, 22.) Tämänkaltaisen, empiiristä mittauksista lähtevä käsitteenmuodostus voi lopulta luoda myös oikeaa ihmistieteellistä teoriaa (Ahmavaara 1957a, 47, Ahmavaara 1957b, 6, 17). Esimerkiksi Thurstonen mallia seuraava psykologinen tutkimus on Ahmavaaran mielestä täysin ennako-olettamuksista vapaata – se ei ole sitoutunut minikäänlaisiin ennakkokäsityksiin mielen rakenteesta, vaan faktorianaalyysi saa itse paljastaa, miten mielen toimintaa tulee hahmottaa (Ahmavaara 1957a, 73, 114, 118; Ahmavaara 1957b, 11–13, 19, 21).

Tällä epistemologisella lähtökohdalla on vallankumouksellisia seurauksia arviointia koskevalle teorialle ja ennen kaikkea arvioinnin käytännöille. Siinä määrin kun Ahmavaaran kuvaamasta tieteen tekemisen tavasta tuli nopeasti paradigmaattinen, myös kasvatustieteilijät alkoivat entistä enemmän hahmottaa tieteenalansa edistystä juuri mittausteknisen kehityksen valossa. (Ks. esim. Pitkänen 1960, 95; Takala 1963.)

Nämä tekniikat levisivät nopeasti myös mitä erilaisimpiin käytäntöihin, joissa tuotettiin yhä hienojakoisempia luokituksia oppilaiden suoritusten ja luonteenpiirteiden arvioimiseksi (Rinne 2012; ks. myös Ojakangas 1997). Tarkastelen seuraavaksi erityisesti sitä, miten

oppilasarviointia on pyritty uudistamaan tilastollisten tekniikoiden välityksellä, ja miten nämä uudistukset kytkeytyvät laajempaan koulutuksen rationaalisen hallinnan ideologiaan.

Oppilaiden arviointi ja tilastollinen järkeily

Ennen toista maailmansotaa suomalaisen kansanopetuksen oppilasarviointi – esimerkiksi koearvosanojen ja todistusten muodossa – oli varsin epäyhdenäistä. Se perustui lähinnä kunkin opettajan omiin periaatteisiin ja ratkaisuihin. (Simola 2002b, 211.) 1940-luvulta alkaen tätä lähestymistapaa alettiin kuitenkin pitää mielivaltaisena (ks. esim. Salmela 1948; Vahervuo 1947). Vuonna 1943 Kouluhallitus antoikin ohjeet oppilasarvioinnin valtakunnallisesta yhtenäistämisestä (Simola 2002b, 210). Tämä sopi hyvin 1940-luvulla koko maahan ulottuvaan oppivelvollisuuteen, mutta samalla myös opetussuunnitelmissa yhä enenevässä määrin korostuvaan yksilöllisyyteen: juuri oppilasyksilö alkaa sotien jälkeisinä vuosikymmeninä muodostua opetuksen perusyksiköksi, johon paitsi opetuksen, myös arvioinnin täytyy mukautua. Samalla yksilöllisyys kuitenkin kytketään osaksi tieteelliseen tietoon perustuvaa hallinnan järjestelmää. (Rinne 2012; Simola 2002b, 212–213.) Rationaalinen koulujen hallinta alkaa luokitella oppilaiden yksilöllisyyden muotoja yhä yksityiskohtaisemmin, mikä puolestaan tuottaa yhä hienojakoisempia hallinnan mekanismeja (Kivirauma & Kivinen 1988; Rinne 2012). Osoitan, miten juuri tilastollinen tieto toimii tässä yksilöllisyyden muotojen ja koulutusjärjestelmän hallinnan välisen yhteyden välittäjänä: miten se tekee oppilaiden toiminnan ja yksilölliset luonteenpiirteet näkyviksi ja hallittaviksi suhteessa opetussuunnitelman toteuttamiseen.

Varsin usein rationaalisen opetuksen suunnittelun ja arvioinnin perusmallina on pidetty niin sanottua Tylerin rationaalia (Autio 2010; Kiilakoski & Oravakangas 2010, 9). Tämä yhdysvaltalaisen opetussuunnitelmateoreetikko Ralph Tylerin mukaan nimetty periaate yhdistää toisiinsa kaikki opetuksen elementit tieteellisen tiedon nimissä. Ensinnäkin opetuksen tavoitteet tulee määritellä erilaisina yhteiskunnan tarpeina, jotka puolestaan muotoillaan mitattaviksi käyttäytymistavoitteiksi. Tämän jälkeen oppimispsykologian tehtävänä on määritellä ne menetelmät, joilla kyseiset tavoitteet voi kaikkein tehokkaimmin saavuttaa. Opetuksen jälkeen arvioidaan mittaamalla oppimisen tuloksia, joita sitten verrataan tavoitekäyttäytymiseen. Näin paljastuu mitattavissa oleva vaje (*gap*) tavoitteiden ja tulosten välillä, joiden pohjalta voidaan tehdä muutoksia opetusmenetelmiin. (Tyler 1947.) Juuri tämänkaltaista mallia opetuksen arviointiin ja kehittämiseen esitettiin hyvin laajasti 1960- ja 1970-lukujen suomalaisessa kasvatustieteessä (Saari 2011, luku 4.3.). Hannu Simola näkee, että valtiolliseen kouludiskurssiin ilmaantui tänä aikana tavoiterationaalisuus, jossa juuri eksaktisti määritellyt tavoitteet ohjaavat koko koulujärjestelmän toimintaa (Simola 2002a). Arviointi puolestaan toimii ikään kuin koneistoa diagnostisoivana tietona. Tieteellinen arviointi kehittyikin kiinteästi osana 60-luvun ”suuria uudistuksia”, joissa koulutusjärjestelmää pyrittiin kehittämään tiedeperustaisen kokeilutoiminnan ja arvioinnin pohjalta (Konttinen 1995). Esimerkiksi peruskoulukokeilussa tieteellinen tieto kiersi systemaattisesti tutkijoiden ja kehittäjien välillä:

Kouluvaavutusmittausten tulokset ja muu kentältä koottu tieto välitettiin suoraan tietokoneliuskoilta kouluhallituksen pääjohtajalle päätöksentekoa varten, ja päätökset ja ohjeet puolestaan lähtivät kentälle milteipä paluupostissa. Ideaalina pidettiin kentän, opettajien ja koulujen kokemuksen mahdollisimman kattavaa kuulemistä, kokemuksen tehokasta kokoamista arviointitutki-

muksen avulla sekä nopeaa toimenpiteisiin johtavaa päätöksentekoa [– –].
(Konttinen 1995, 7–8.)

Kenties jotakin tämän rationaalien läpäisevyydestä kertoo se, että Heinosen ja Viljasen arvioinnin oppikirjassa (1980) mittaamalla tapahtuvan arvioinnin menetelmä on muodostunut täysin luonnolliseksi ja itsestään selväksi osaksi kouluopetusta. Tylerilaisittain koulutus koostuu Heinosen ja Viljasen mukaan aina neljästä vaiheesta: suunnittelusta, opettamisesta, tuloksista ja evaluaatiosta (so. arvioinnista). Näissä kaikissa vaiheissa tieteellisellä tiedolla on keskeinen rooli. Tilastollinen mittaaminen on puolestaan arviointivaiheessa varsin luonnollinen menetelmä (Heinonen & Viljanen 1980, 12–13; Leimu 1974).

Tämä malli esitetään yleisenä, mitä tahansa opetustapahtumaa, missä tahansa ajassa tai paikassa koskevana kaavana. Sen toteuttaminen vaati merkittäviä muutoksia koulujen ohjaamiseen ja niiden toimintaan: opetussuunnitelmia ei ole perusteltua rakentaa, ellei niiden toteutumista seurata tarkasti. Juuri Tylerin rationaalien puitteissa voidaan ylipäättään odottaa, että määrätyistä tavoitteista tulee seurata jotakin tiettyä oppimistulosten kannalta ja että tämä yhteys on mitattavissa. (Ks. esim. Heinonen 1961, 19.) Rationaali antaa pohjan myös arvioida arviointia eli määrittellä parannuksia vallitseviin mittauskäytäntöihin. Sen pohjalta esitetäänkin oppilasarvioinnin muuttamista matemaattis-tilastollisiin menetelmiin perustuvaksi. Muiden muassa Veikko Heinonen paheksuu vallitsevaa tapaa, jolla koulusäätöjä mitataan varsin epäsystemaattisesti:

Koulutyössä on kylläkin yllättävän pitkään menetellyt alkeellinen mittaaminen, jossa ainutkertainen koe, jonka yksityinen opettaja on suunnitellut intui-tionsa varassa, on hyvin riittänyt. Tulosten tarkkuutta ei ole kysyty, vaan opettaja on yleensä katsonut oppiaineen tuntemuksensa perusteella pystyvän-sä kokeita, mittausvälineitä, suunnittelemaan. Asiantila on ollut verrattavissa pituudenmittaukseen, jota eri mittaajat suorittaisivat käyttäen kukin omaa laatimaansa mittakeppiä. (Heinonen 1961, 16; ks. myös Vahervuo 1947, 9–10; Salmela 1948, 31–34.)

Sen sijaan nykyaikaisen kasvatustieteen ja siihen perustuvan opetuksen tunnuspiirteenä on systemaattinen mittaaminen. Heinonen viittaakin E.L. Thorndiken kuuluisaan toteamukseen: ”Kaikki, mikä on olemassa, on olemassa jossakin määrässä, ja kaikkea mitä on olemassa, voidaan mitata.” (Heinonen 1961, 23). Jotta objektiivinen arviointi voisi ylipäättään olla mahdollista, on vallitsevia koekäytäntöjä muutettava radikaalisti. Heinonen (1961, 50–51) kritisoi kouluissa käytettävää vapaata, esseemuotoista vastaustyyppiä. Esseevastaukset näet tuottavat huonosti yhtenäistettävissä olevia tuloksia.

Kirjoitelmatyyppisessä tehtävässä annetaan oppilaalle yleisluonteinen ärsy-ke, joka edellyttää kuvailua, kertomista, selostamista tai vertailua. Reaktiot ovat koehenkilön vapaasti valittavissa, ellei niitä ole suoritusohjeissa mitenkään rajoitettu. Juuri tämä ilmaisun vapaus merkitsee tämän tehtävätyypin suurinta heikkoutta. (Heinonen 1961, 88.)

Epäjohdonmukainen arviointi aiheuttaa myös ristiriitoja opettajan ja oppilaan välillä, toisi-naan jopa epäoikeudenmukaisuuden tunnetta, sekä motivaation laskua. Lisäksi esseillä on heikko valideetti, ne eivät muodosta edustavaa näytettä, ja sattumalla on paljon vaikutusta niiden sisältöön. (Heinonen 1961, 88–89.) Edelleen, esseiden massamuotoinen tarkastami-nen on varsin vaivalloista. (Heinonen 1961, 55; Mikkonen & Mikkonen 1971, 18.)

Heinosen mielestä täydennys- ja monivalintatestejä tulisikin suosia kouluissa nykyistä enemmän. Ne ovat ensinnäkin objektiivisempia, sillä jokainen testin koehenkilö ymmärtää tehtävän samalla tavalla. Niin ikään kyseessä on varsin tehokas, rationaalinen työtapa. On näet pyrittävä siihen, että informaation määrä ja laatu kokeisiin uhrattuun työpanokseen verrattuna on hyvä. (Heinonen 1961, 50–57.)

Tilastotieteellinen tutkimus alkaa sotien jälkeen tarjota yksityiskohtaisia kriteereitä koulukokeille ja niiden arvioinnille. Ensinnäkin kokeiden on oltava ”sisällöllisesti edustavia”, jotta sisällöllinen validiteetti tulisi taattua. Toisaalta kokeiden tulee erotella johdonmukaisesti yksilöitä. Tämä tarkoittaa sitä, että pelkkää satunnaisvaihtelua tuottavat kysymykset tulee karsia pois. Niin ikään kysymykset, jotka jokainen oppilas osaa ratkaista tai toisaalta kukaan oppilas ei osaa ratkaista, on poistettava kysymysten joukosta. Erityisesti valtakunnallisten kokeiden kohdalla on tärkeää, että kokeet ovat standardoituja eli täysin vertailukelpoisia. (Heinonen 1961; Nurmi 1967; Mikkonen & Mikkonen 1971.) Näiden kriteerien valossa voitaisiinkin sanoa, että tilastotieteeseen pohjaavassa arvioinnissa kyse on objektiivisuudesta ”kollektiivisena empirisminä”: tiedon tulee olla siksi mekaanisesti ja standardoitusti tuotettua, ettei se ole varsinaisesti kenenkään tietoa. Tieto siis edellyttää montaa toimijaa, jotka toimivat samalla tavalla. (Daston & Galison 2010, 29–30.) Validiteetista ja reliabiliteetista, keskiarvosta ja korrelaatiosta tulee näin arviointia ja koulusaavutustestejä koskevissa oppaissa osa opettajien ammattisanastoa, joka antaa heille mahdollisuuden osallistua objektiiviseen arviointiin (Heinonen 1961; Nurmi 1967; Mikkonen & Mikkonen 1971; Heinonen & Viljanen 1980).

Samalla kun koulusaavutusten ja oppilasarvioinnin oppaissa vaaditaan arvioinnin yhtenäistämistä ja yksilöllistä erottelevuutta, väitetään että sen tulisi aina viitata tilastollisesti määriteltyyn populaatioon (ks. Salmela 1948, 69–77; Vahervuo 1958; Heinonen 1961, Nurmi 1967; Mikkonen & Mikkonen 1971; Heinonen & Viljanen 1980). Kuten Ahmavaara, joka painotti tilastotieteellisen operationalisoinnin merkitystä, Heinonenkin toteaa:

Kun olemme laskeneet koehenkilön pisteluvun testissä, emme tämän luvun perusteella vielä tiedä suorituksen tasosta mitään, emme tiedä, onko suoritus hyvä, keskiarvoinen vai huono, emmekä voi suorittaa mitään toimenpiteitä oppilaan suoritusason muuttamiseksi. Yksityinen pisteluku saa merkityksen vasta, kun vertaamme suoritusta muiden oppilaiden suorituksiin, kun asetamme yksilön tuloksen ryhmän tulosluettelon taustaa vasten. (Heinonen 1961, 154; ks. myös Salmela 1948; Mikkonen & Mikkonen 1971, 15.)

Samalla tämä tarkoittaa sitä, ettei oppilaan tai opettajan oma käsitys jonkin tehtävän tai oppiaineen vaikeudesta sellaisenaan ole huomioon otettava. Se, onko testi tai oppiaine vaikea, selviää ainoastaan populaatiotasoon viittaamalla. Arviointi ei siis ole yhden opettajan ja yhden oppilaan välinen asia, vaan osa populaatiotasoon mekaanista objektiivisuutta. Näin ollen mikään arvosana ei puhu vain yksittäisen subjektin kieltä, vaan tuomion julistaa itse objektiivisen totuuden persoonaton ääni. Kuten seuraavassa luvussa tulee tarkemmin ilmi, tapahtuu eräänlainen kehäpäätelmä: oletetaan, että yleinen, populaatiota koskeva totuus on olemassa. Tämän saavuttamisen ehtona on mittaamisen yhtenäistäminen siten, että jokainen mittaaja käyttäytyy tismalleen samalla tavalla. Vasta tällä tavoin tulee mahdolliseksi puhua yksilöistä ja heidän suorituksistaan. Näin saadaan toteutettua se, mikä oletettiin ennalta olemassa olevaksi: ylinnä on populaatiotaso, ja oppilaan yksilöllisyys tulee tiedettäväksi ainoastaan tälle alisteisena.

Edellä kuvatut tilastolliset tekniikat ilmenevät myös siinä, minkälaisia ohjeita kokeiden laa-
timiselle esitetään. Kuten tilastollisessakin tutkimuksessa, on aluksi syytä tehdä erinäisiä
esitestauksia, joissa selvitetään kokeen optimaalinen kesto sekä karsitaan sellaiset tehtävät,
jotka eivät kykene erottelamaan oppilaita toisistaan. Esikokeen perusteella on myös mah-
dollista asettaa kokeen kysymykset vaikeusjärjestykseen ratkaisufrekvenssien mukaisesti.
Edelleen, koe täytyy myös standardisoida tilastollisen tutkimuksen periaatteiden mukaises-
ti. Lisäksi laaditaan koetta koskevat suoritusohjeet, jotka edellyttävät myös opettajilta
kurinalaista käyttäytymistä ja kommunikaatiota. (Heinonen 1961, 64–80, 137–138; Nurmi
1967, 104; Heinonen & Viljanen 1980, 213–224.) Näin voidaan toteuttaa valtakunnallisia-
kin kokeita, joiden tulokset ovat Nurmen (1967, 105) mukaan objektiivisia riippumatta sii-
tä, missä päin maata ne toteutetaan.

Valtakunnalliset kokeet antavat kokonaiskuvan koko Suomen suoritustasosta, jota kou-
luhallinnon viranomaiset voivat käyttää kehittämistyössään ja johon toisaalta myös oppilas
voi objektiivisesti verrata suoritustasoaan (Leimu 1974; ks. myös Saari 1983, 104). Lisäksi
tämä antaa Nurmen mukaan myös opettajalle tietoa hänen omasta kyvykkyydestään opetta-
jana, sillä hän saa tietää luokkansa ”yleistason” yksittäisessä aineessa. ”Jos luokan tulos
osoittautuu heikonlaiseksi, on opettajalla mahdollisuus vaieta asiasta ja tehostaa opetus-
taan.” Luokan hyvää suoritusta voi puolestaan hyvin perustein kehua muillekin opettajille.
(Nurmi 1967, 105.) Edelleen oppilaan siirtyessä kouluasteelta toiselle tai työelämään, voi-
daan hänen kehityksestään ja valmiuksistaan saada objektiivista tietoa. (Vahervuo 1958,
27–31.)

Väliyhteenvetona voidaan todeta, että tilastollinen järkeilyn tyyli pyrkii muodostumaan
toisen maailmansodan jälkeen osaksi koulujen rationaalista hallintaa. Se pyrkii korvaamaan
vanhan arvioinnin mielivallan systemaattiseen havainnointiin ja mittaamiseen perustuvalla
evidenssillä, joka on subjektiivisista, teoreettisista sekä poliittisista ennako-olettamuksista
vapaata. Tällä tavoin oppilasarviointi liittyy osaksi tilastotieteellisen tiedon ja yhteiskunnal-
lisen vallankäytön historiallista jatkumoa, jossa arvovapaa tieto legitimoit hallintaa. Tilas-
tollinen järkeilyn tyyli muodostuu hyvin yksityiskohtaisista säännöistä, jotka kurinalaista-
vat opettajien, oppilaiden sekä tutkijoiden ja virkamiesten toimintaa. Se mahdollistaa
objektiivisen kommunikaation oppilasarviointiin eri tavalla osallisten kesken: se tarjoaa
kaikille yhteiset arvioinnin kriteerit ja käsitteistön. Samalla luodaan valtakunnallinen
koneisto, joka on sekä yhteismitallinen että yksilöivä: se kytkee yksittäisen oppilaan popu-
laatioon sekä valtakunnalliseen koulujen hallinnan järjestelmään ja osoittaa kullekin oman
yksilöllisen paikkansa koulusuoritusten kokonaisuudessa.

Normaali arvioinnin yhteisenä mittapuuna

Kuten Ian Hacking toteaa, järkeilyn tyylit voivat luoda myös täysin uudenlaisia inhimilli-
sen olemassaolon luokkia. Ensinnäkin voidaan sanoa, että tällainen uusi olemassaolon
luokka tilastollisessa järkeilyssä on populaation ilmaantuminen tilastollisen tiedon viitta-
uskohteeksi. Tieto ei ole totta minkään yksittäisen havainnon tasolla, vaan juuri suhteessa
populaatioon. Kuten edellä on todettu, tämä oletus populaatiota koskevasta tiedosta alkaa
toisen maailmansodan jälkeen hallita myös oppilasarviointia.

Faktorianalyysi on ehkä kärkevin esimerkki tavasta, jolla uusien luokkien luonti tapah-
tuu. Kuten yllä todettiin, faktorianalyysi välttää Ahmavaaran mukaan kaikenlaiset ontologi-
set ennako-oletukset ja antaa faktoroinnin ja rotaation kertoa asioiden laidan. Näin esi-
merkiksi älykkyystesteissä tai asennetutkimuksissa voidaan tuottaa uudenlaisia ryhmäenti-
teettejä.

Tilastollinen järkeilyn tyyli voi kuitenkin luoda uusia luokkia arviointiin myös paljon yksinkertaisimmilla menetelmillä. Kenties vaikutusvaltaisin epistemologinen murros, joka on mahdollistanut uusien luokkien luomisen, on normaalin käsitteen ilmaantuminen arvioinnin kentälle.

Tilastollisella normaalilla ja normaalikäyrällä (tunnetaan myös Gaussin käyränä ja kellokäyränä) on alkuperänsä tähtitieteellisissä mittauksissa, joissa havaittiin, että tähden sijaintia koskevissa havainnoissa tapahtui säännönmukaisia mittavirheitä, jotka muodostivat tilastoituna muodostaen keskeltä pullistuneen käyrän. Tähtien sijainti voitiin siis päätellä sen mukaan, mihin enimmäkseen havainnot osuivat. Tämä oli ensimmäinen versio normaalikäyrästä. Astronomi ja matemaatikko Adolphe Quetelet (1796–1874) siirsi normaalikäyrän las-kennallisena periaatteena kokonaan uudelle alueelle, yhteiskunnallisiin ilmiöihin. Normaalikäyrän avulla ei näin ollen enää päätelty jonkin objektin sijaintia, vaan se itsessään kuvasi asioiden tilaa – luonnon- ja yhteiskunnallisten ilmiöiden normaalia jakaantumista populaatiossa. Quetelet jopa ajatteli, että näiden havaintojen pohjalta voitaisiin kuvata yleisiä inhimillisiä ominaisuuksia kuvaava ”keskivertoihminen” (*l’homme type*). (Hacking 1990, 107–113.)

Quetelet’n innoittamana yhteiskuntatieteissäkin alettiin kerätä suuria havaintomääriä, joissa huomattiin normaalia jakautumista. Jo 1910-luvulla suomalainen kasvatustieteilijä Henning Söderhjelm esitti Quetelet’n hengessä, että kasvatustieteellisen tutkimuksen perimmäinen kohde on ”normaalisielu” (*normalsjäl*), eräänlainen ideaalimalli, jossa sielun osat ovat harmonisessa suhteessa toisiinsa, ja josta yksittäisen ihmisten yksittäiset ominaisuudet eroavat jossakin mitattavissa olevassa määrässä. (Söderhjelm 1915.)

Juuri normaalin periaatteesta on tullut 1900-luvun aikana keskeinen pedagogisen vallan väline. Kaikille yhteisen moraalivaruuden rapistuessa siitä, mikä on normaalia, on tullut tavoiteltavaa, ja patologinen poikkeavuus on korvannut moraalisen turmeltuneisuuden muodot sopeuttavan toiminnan kohteena. Jo vuosisadan alussa Suomessa käytettiin esimerkiksi älykkyystestejä erottelemaan normaalit oppilaat erityiskouluihin sijoitettavista heikkolahjaisista (Rinne 2012, 41–42; Ojakangas 1997.)

Normaalilla on monia funktioita arvioinnissa, mutta kaikkein selkeimmin sen yhteys koulun käytäntöihin ilmenee tilastollisia menetelmiä hyödyntävässä oppilasarvostelussa, jossa sillä on toisen maailmansodan jälkeen keskeinen, vaikkakin pysyvästi kiistanalainen asema. Oiva Ketonen ja Toivo Vahervuo kannattivat 1940-luvun lopulta alkaen tahoillaan absoluuttisen ja suhteellisen arvioinnin käyttöä. Näiden oppineiden välisessä erimielisyydessä oli panoksena sekä tieteellinen eksaktisuus – se, millä tavoin tieteellinen mittaaminen voi tavoittaa kohteensa – että oppilaita koskeva oikeudenmukaisuus ja tasapuolinen kohtelu. Ketonen piti eksaktisuutena absoluuttisen arvostelun kytkeytymistä tiettyyn suorituskriteeriin. Hän näki, että kyseinen tapa arvioida antaa tarkan kuvan oppilaan osaamisesta esimerkiksi ammatinvalinnan ohjauksessa. Suhteellisen arvostelutavan ilmaisemat ”ensimmäinen” ja ”viimeinen” eivät tässä sano mitään, vaan ohjauksessa on tiedettävä ”todellinen taito tai kyky”. (Ketonen 1951, 134.) Ketoselle vaa’assa painoi myös ihmiskuva, sillä suhteellisen arvostelun ”(p)erusteluksi ei riitä esim. toteamus että ihmisten pituuden vaihteluissa näkyy Gaussin jakautuminen, sillä oppilaan käyttäytyminen koulussa on toisella tavalla mielekäästä kuin ihmisen pituuden kasvaminen.” Toisin kuin muussa luonnossa, joka saattaa noudattaa normaalijakauman periaatteita, ihmisten osaamisen arvostelussa kohteena on ”elävä ja ajatteleva materiaali”. (Ketonen 1951, 135.)

Sen sijaan Toivo Vahervuolle Ketosen lähtökohta oli karkea tieto- ja mittausopillinen erhe. Hän argumentoi sekä kasvatustieteellisissä artikkeleissa että oppikirjassaan vahvasti

suhteellisen, normaalijakauman ohjaavaan asemaan perustuvan arvioinnin puolesta. Ensinnäkin absoluuttinen arviointi on epätarkkaa. Siinä toimitaan samoin kuin piirrettäessä mittakeppiin mielivaltaisain välein asteiviivoja nolasta kymmeneen. Näin oppilaat kyllä saadaan järjestykseen, mutta suoritusten väleissä ei yhtä suuria etäisyyksiä. (Vahervuo 1947, 10–11.) Kyseessä on siis Ketosen lähtökohdan ympärikäntäminen: pelkästään järjestykseen laittava arviointi on absoluuttisen arvioinnin helmasynti.

Tästä seuraa myös absoluuttisen arvostelun epäoikeudenmukaisuus. Vahervuon mukaan ”(k)oetta, jossa neljäs luokasta kykenee maksimisuoritukseen, ei oikeastaan saa arvosella. Siinä tehdään vääryyttä etevimmille oppilaille, jotka eivät saa tilaisuutta näyttää paremmuuttaan.” (Vahervuo 1951, 240.) Lisäksi Vahervuo väittää, että opettajat antavat absoluuttisessa järjestelmässä helposti mielivaltaisia arvosanoja. Sen sijaan suhteellisessa arvioinnissa oppilaiden lahjakkuustaso heijastuu objektiivisesti. (Vahervuo 1951, 241; Vahervuo 1958, 15–16.) Vaikka huonoihin arvosanoihin saattaa vaikuttaa myös huono opettaja, suhteellisessa arvioinnissa arvostelu kohdistuu aina myös oppilaiden ”luontaisiin ominaisuuksiin” (Vahervuo 1947, 18).

Vahervuo toteaa, että ”arvosana sellaisenaan ei ole hyvä tai huono” (Vahervuo 1951, 240). Sen merkitys on aina suhteessa vertailun kohteena olevaan populaatioon, toisin sanoen sillä ei ole absoluuttista merkitystä. (Vahervuo 1947, 11.) Miten sitten Vahervuo voi väittää, että suhteellinen arviointi voi ylittää absoluuttisen arvioinnin ongelmat? Hän esittää, että suhteelliset arvosanat viittaavat kuitenkin johonkin pelkän konvention ulkopuoliseen todellisuuteen:

Ensinnäkin on kysyttävä, millainen arvosanojen jakaantumiskuvion oikeastaan tulisi olla. Ja siihen kysymykseen näyttää luonto itse antavan vastauksen. Esim. samanikäisten ihmisyksilöiden pituuden jakaantumiskuvio saa suurissa joukoissa tarkoin määrätyn muodon. Tämä muoto [– –] on tunnettu sattuman varaisen vaihtelun jakaantumiskäyrä [– –]. Lukemattomissa muissa tapauksissa, joissa suureen vaihtelu riippuu monen monista eri tekijöistä, on tuloksena Gauss'in käyrän mukainen suureen arvojen jakautuminen. Luonto näyttää pyrkivän juuri tällaiseen jakaantumiseen. Mikä olisikaan silloin lähempänä kuin otaksua, että myöskin koulun eri oppiaineitten edellyttämän ”lahjakkuuden” jakaautuminen on ”todellisuudessa” Gauss'in käyrän mukainen eli kuten sanotaan normaalin jakaautuminen. (Vahervuo 1947, 12.)

Arvosanojen ”todellinen merkitys” löytyy siis viittaussuhteessa populaatioon, jonka Vahervuo olettaa aina noudattavan normaalia jakautumista. (Vahervuo 1947, 12; Salmela 1948, 70–71). Normaali, niin luonnon kuin yhteiskunnallisenkin järjestyksen mallina, toimii kriteerinä kouluarvosanojen objektiivisuudelle sekä oppilaiden ryhmittelylle esimerkiksi luokkanormien mukaisesti (Heinonen 1961, 53–55, 147, 149, 156; Vahervuo 1958; Salmela 1948). Veli Nurmi antaa opettajille kätevän peukalosäännön kokeen onnistuneisuuden tarkistamiseksi: kun koevastaukset asetetaan arvosanan mukaisiin pinoihin, tulisi niiden muodostaa normaalikäyrä, jossa ääripäiden pinot ovat ohuimpia, kun taas keskitason suorituksia on eniten (Nurmi 1967, 97).

Arvosanojen tulee esimerkiksi asteikolla neljästä kymmeneen hajautua seuraavien prosenttilukujen mukaisesti: 3–10–22–30–22–10–3 % (Vahervuo 1947, 14; vrt. Vahervuo 1958, 45; Salmela 1948, 74–76). Tällä tavoin jokaisella arvosanalla on ”tarkoin määrätty merkitys”, ja mikäli kaikissa oppilaitoksissa otettaisiin käyttöön tämä yhteisen mittapuun käytäntö, mahdollistaisi tämä viimein oppilaiden suorituksia koskevan objektiivisen kielen,

joka ei jättäisi mahdollisuutta väärinymmärryksille (Vahervuo 1947, 15). On mielenkiintoista, että Vahervuon mukaan voitaisiinkin lopulta unohtaa kysymys siitä, onko luonto ja yhteiskunta normaalisti jakautunutta – riittää, kun sovitaan siitä, että kaikessa arvostelussa sovelletaan normaalijakaumaa ja sen mukaista arvosanojen hajontaa. Tämä takaa riittävän objektiivisuuden ja yhteisymmärryksen. (Vahervuo 1947, 12–13; Vahervuo 1958, 19.) Näin normaalista tulee itseään toteuttava ennuste: kun sitä käytetään osana kaikkea arviointia ja sitä koskevaa puhetta, siitä todellakin tulee osa yhteiskunnallista todellisuutta. Vahervuo katsookin, että kyseinen periaate on jo toteutunut Kouluhallituksen linjauksessa, jossa opettajan yhdessä aineessa antamien keskiarvojen ei tulisi olla alle 7 eikä yli 8 (Vahervuo 1947, 14; Salmela 1948, 41–42). Hän toteaa vielä, että mikäli koulujen oppilaat olisi ryhmitelty normaalijakauman periaatteiden mukaisesti, tarjoaisi tämä kvantitatiiviselle kasvatustutkimukselle sen menetelmien kannalta otollisen ympäristön (Vahervuo 1958, 39–40; ks. myös Koskenniemi & Grenquist 1934).

Keskusteluun myös osallistunut Arvo Lehtovaara puolestaan vaatii oppikoulun pääsy-tutkintoja muotoiltavaksi normaalikäyrän periaatteiden mukaisiksi. Hän viittaa lisäksi ruotsalaisen Torsten Husénin tutkimukseen, jonka mukaan vain 19 %:lla ruotsalaisista on edellytykset suoriutua ylioppilastutkinnosta. Lehtovaara väittää, että ”(k)oska ei ole syytä olettaa, että Suomen lapsiaineksen jakautuminen eri lahjakkuusasteisiin olisi huomattavasti naapurimaassamme todetusta jakaantumisesta poikkeava ja koska myöskin oppikoulumme vaatimustaso suunnilleen vastaa sikäläistä, näyttää olevan oikeutettua tehdä vastaavat johtopäätökset myöskin Suomen olojen osalta.” (Lehtovaara 1951, 43.) Oppikoulun pääsy-tutkinto olisikin muokattava sellaiseksi, että ainoastaan tuo ”lahjakkuusedellytyksiltään” parhain lapsiaines pääsisi läpi. Näin oppikoulun tulokset paranisivat, oppilasryhmistä tulisi homogeenisempiä, ja myös erinäisistä mentaalihygieenisistä ongelmista päästäisiin eroon. (Lehtovaara 1951, 43–44.) [2]

Vaikka normaalin kategorian esitetään osana arvovapaata tieteellistä arviointia, sitä käytetään toisinaan myös normatiivisesti, mikä tuo kaikkein selkeimmin esille tilastollisen arvioinnin yhteyden vallankäyttöön. Vaikka normaalin ajatus on tässä esitetty lähinnä tilastotieteellisenä kategoriana, on sillä juurensa myös fysiologiassa, jossa organismin tilaa arvioitiin patologisen ja normaalin ulottuvuuksilla (Canguilhem 1978). Erityisesti tässä merkityksessä normaali voi antaa suuntaa myös koulun hallinnan päämäärille. Normaalin molemmat merkitykset esiintyvät tavassa, jolla Heinonen esittää koulukokeiden tuloksien olevan oiva väline patologisesti poikkeavien yksilöiden alustavaan diagnoosiin. (Heinonen 1961, 163.) Tämän oivalluksen pohjalta tulisi myös uudistaa opettajien kielenkäyttöä, sillä ”(a)rkkipuheessa usein lausutut diagnoosit, jolloin oppilaan poikkeavuuden syyksi mainitaan tyhmyys, ilkeys, älyttömyys, laiskuus, saamattomuus tms. ovat liian ylimalkaisia hoidon määräämiseen ja usein sitä paitsi virheellisiä” (Heinonen 1961, 167–168).

Tällä tavoin mukaan arviointiin astuu eriskummallinen, tiedeperustainen normatiivisuus. Heinonen kehottaa kyllä luopumaan epämääräisistä moraalisisista arvioista, mutta korvaa nämä diagnostisilla määritelmillä, jotka itsessään osoittavat suuntaa erityiselle hoitotavalle. Kyseessä on siis eräänlainen immanentti teleologia, jossa normaali osoittaa tieteellisesti terveyden ja patologian muodot ja elämän pyrkimisen kohti normaaliutta. Näin ollen, kuten lääkäritkin, opettajat ja tutkijat voivat antaa ohjata itseään sellaisen tiedon pohjalta, joka on yhtä aikaa sekä normatiivista että arvovapaata. (Hacking 1990, 160–163; ks. myös Ojakangas 1997.) [3]

Tällä tavoin normaali alkaa hallita tapaa, jolla arvosanoja annetaan. Tässäkin se onnistuu välttämään ongelmallisen sanojen ja todellisuuden välisen suhteen, kunhan kaikki toi-

mivat samalla tavalla ja arvosanojen hajonta on ennalta määrätty. Normaalin ohjaama arviointi mahdollistaa oikeuden- ja totuudenmukaisen kommunikaation opettajan ja oppilaan välillä, sillä kaikkia kohdellaan samojen kriteerien mukaisesti. Normaalin periaatteen ajatellaan myös paljastavan elämässä yleensä vallitsevan sisäisen normin, jota koulutuksen hallinnan tulisi noudatella. Näin ollen Vahervuo näkeekin, että normaaliin perustuva arviointi edustaa lapsikeskeistä kasvatusta; siinä missä absoluuttisessa arvioinnissa lapsi pakotetaan sopeutumaan koulun normeihin, suhteellisessa arvioinnissa koulu puolestaan sopeutuu lapsen sisäisiin, luonnollisiin ominaisuuksiin (Vahervuo 1958, 48).

Lopuksi: tilastolliset tekniikat arvioinnin uudistajana

Sotien jälkeen Suomessa kehittynyt oppilasarvioinnin teknologia perustuu pitkälti edeltävänä vuosikymmeninä muodostuneisiin tilastotieteellisiin käytäntöihin. Ne puolestaan nojasivat jo vuosisatoja vanhoihin tapoihin ymmärtää empiirisen ja kumuloituvan tiedon luonne. Vaikka tilastolliseen järkeilyyn tyyliin perustuva arviointi edustikin murrosta koulujen evaaluaatiossa, on sillä siis takanaan jo kauan sitten muotoiltuja tiedontuotannon traditioita.

Olen kuvannut, kuinka arvioinnin oppikirjoissa, testioppaissa ja artikkeleissa luodaan näiden traditioiden mukaisesti jako epäsystemaattisen, subjektiivisen sekä epäoikeudenmukaisen, sekä toisaalta kaikille yhdenmukaisen, objektiivisen ja oikeudenmukaisen arvioinnin välille. Arviointia koskeva kieli näyttää olevan sellaisenaan, opettajien arkipäivässä, helposti subjektiivista ja epäjohdonmukaista. Se ei tavoita oppilaiden suoritusten populaatiotasolla ilmenevää todellisuutta. Sen sijaan erityisesti suhteellisen arvioinnin nähdään korjaavan monia paitsi arvioinnin subjektiivisuuteen, myös epäoikeudenmukaisuuteen ja vaivalloisuuteen liittyviä ongelmia.

Nämä jaottelut lepäivät populaatiota koskevan ontologisen ja epistemologisen perustan varassa. Tunnetusti normaalin kaltaiset tilastolliset periaatteet eivät ole koskaan totta sellaisenaan yksilön tasolla, vaan vasta normaalia koskevan abstraktion kautta voidaan tavoittaa se todellisuus, joka ilmenee merkkinä oppilaiden koulusaavutuksissa. Koska arvioinnin kohdetta todellisuudessa ei voi suoraan havaita saati koskettaa, on tuotettava toissijaisia havainnoinnin kohteita. Nämä ovat oppilaiden tuottamia, laskettavissa olevia objekteja, joita voidaan manipuloida tilastomatematisin menetelmin. Organisoimalla näiden ”kvasiobjektien” tuottaminen ja niitä koskeva analyysi, voidaan luoda vakaa ja yhtenäinen perusta arviointitiedolle.

Osana tilastollista järkeilyä tyyliä arviointi liittyy myös biopoliittiseen hallinnan traditioon. Siinä keskeistä on hallinnan pyrkimys myötäillä populaation tasolla ilmeneviä voimia ja elämän itsensä normia. Tämänkaltaisen hallinnan legitimaatioperustana on puolestaan neutraaliksi ja epäpoliittiseksi oletettu tilastotieteellinen tieto. Näitä periaatteita seuraten arvioinnista pyrittiin sotien jälkeisessä Suomessa tekemään oman aikansa yhtenäistieteen ihanteiden mukaista, arvovapaata ja objektiivista toimintaa, joka ei jättäisi tilaa yksittäisen opettajan mielivallalle. Arviointi esitti olevansa paitsi teoreettisista, myös poliittisista olettamuksista vapaata, ja juuri siksi erinomainen väline koulutuspolitiikan perustaksi. Erityisesti suhteellinen arviointi esitettiin malliksi, joka heijastaisi paitsi luonnon ja yhteiskunnan järjestystä, myös mahdollistaisi eksaktin kommunikaation arviointiin eri tavalla osallisten ihmisten kesken. Tällä tavoin arviointi ei ainoastaan pyri ilmapuntarin tavoin kertomaan passiivisesti koulun todellisuudesta, vaan myös muuttamaan sitä; luomaan uudenlaisia kieltä ja uusia arvioinnin ja oppilaiden ryhmittelyyn käytäntöjä. Lisäksi se toimii symbolisena vallankäyttönä rajaamalla ulos erinäisiä tietämisen ja toimimisen muotoja epätieteellisinä tai epätasa-arvoisina.

Vaikka koulutuksen arviointi onkin tänä päivänä vähemmässä määrin keskusjohtoista ja standardoitua kuin vielä 1970-luvulla, on huomattava, että monet yllä mainitut piirteet ovat paitsi läsnä myös kiistojen kohteena tämän päivän arviointikeskustelussa. Neutraalilta vaihtuvalla tilastollisella evidenssillä on edelleen tilausta demokraattisen yhteiskunnan koulutusjärjestelmässä. Se tiivistää monimutkaisen koulutodellisuuden helposti ymmärrettävissä oleviksi indikaattoreiksi, jotka voivat puolestaan toimia sen kehittämisen kriteereinä. Olisi esimerkiksi mahdotonta kuvitella puhetta kansainvälisesti merkittävästä koulujärjestelmästä ilman PISA-tuloksia, jotka toimivat pohjana niin kansainvälisille asiantuntijavierailuille kuin koulutusviennillekin.

Se, ettei koulutuksen arviointi tänä päivänä juurikaan puhu omasta menneisyydestään – se näet puhuu harvemmin ”klassikoistaan” – kertoo kenties jotain syvällä olevasta ymmärryksestä evidenssin roolista nykyisyyteen pohjaavana ja tulevaisuuteen suuntaavana. Juuri siksi arvioinnin kriittinen ”muistaminen”, siis sitä koskevan historiallisen tiedon tuottaminen, on arvokasta. Se pystyy ainakin muistuttamaan meitä siitä, ettei evidenssiin perustuva kasvatustutkimus ja koulutuksen arviointi sekä sitä koskeva keskustelu ole ainoastaan kapeaan nykyhetkeen perustuvaa, vaan nojaa vuosikymmeniä, jopa vuosisatoja pitkään metodologiseen traditioon.

Viitteet

- [1] Tosin vuonna 1748 perustettiin Ruotsin taulustolaitos statistisen tiedon keräämiseksi ja hyödyntämiseksi.
- [2] Samaa ajatusta normaalin keskeisestä roolista koulujen ja luokkien ryhmittelyssä olivat esittäneet jo Matti Koskenniemi ja Pekka Grenquist 1930-luvulla (ks. Koskenniemi & Grenquist 1934).
- [3] Veli Nurmi kehottaa myös hallitsemaan luokkaa normaalikäyrän avulla siten, että esimerkiksi kyselymenetelmää käytettäessä suunnattaisiin suurin osa kysymyksistä suoritustasoltaan keskitason oppilaille ja joitakin yksittäisiä kysymyksiä myös normaalikäyrän ääripäiden edustajille (Nurmi 1967, 29).

Lähteet

- Ahmavaara, Yrjö 1957a. Henkisten kykyjemme rakenne. Porvoo: WSOY.
- Ahmavaara, Yrjö 1957b. On the Unified Factor Theory of Mind. Helsinki: Suomalainen tiedeakatemia.
- Ahonen, Sirkka 1998. Mitä tapahtui tutkimukselle 1960- ja 1970-luvuilla? Empiristisen paradigman nousu ja pulmallisuus. *Kasvatus* 29 (1), 23–24.
- Alatalo, Marja 2005. Metodisuhdanteiden mahti: lomaketutkimus suomalaisessa sosiologiassa 1947–2000. Tampere: Vastapaino.
- Autio, Tero 2010. Curriculum, opetussuunnitelma ja opettajankoulutus. Teoksessa Eila Lindfors & Jouko Pullinen (toim.), *Cygnaeuksen viitoittamalla tiellä*. Hämeenlinna: Tampereen yliopiston opettajankoulutuslaitos, 115–125.
- Biesta, Gert 2007. Why “What Works” Won’t Work: Evidence-Based Practice and the Democratic Deficit in Educational Research. *Educational Theory* 57 (1), 1–22.
- Canguilhem, Georges 1978. On the Normal and the Pathological. Dordrecht: D. Reidel.
- Gigerenzer, Gerd 1987. Probabilistic Thinking and the Fight Against Subjectivity. Teoksessa Krüger, Lorenz, Gigerenzer, Gerd & Morgan, Mary (toim.), *The Probabilistic Revolution vol. 2*. Cambridge: MIT Press, 11–33.

- Hacking, Ian 1975. *The Emergence of Probability*. Cambridge: Cambridge University Press.
- Hacking, Ian 1990. *The Taming of Chance*. Cambridge: Cambridge University Press.
- Hacking, Ian 1991. How Should We Do the History of Statistics? Teoksessa Burchell, Graham, Gordon, Colin & Miller, Peter (toim.), *Foucault Effect: Studies in Governmentality*. Chicago: University of Chicago Press, 181–196.
- Hacking, Ian 2002. *Historical Ontology*. Cambridge: Harvard University Press.
- Heinonen, Veikko 1959. Koulusaavutustesteistä. *Kasvatus ja koulu* 46, 114–121 .
- Heinonen, Veikko 1961. *Koulusaavutustestit*. Jyväskylä: Gummerus.
- Heinonen, Veikko & Viljanen, Erkki 1980. *Evaluaatio koulussa*. Helsinki: Otava.
- Ketonen, Oiva 1951. Arvosteluista ja arvosanoista. *Kasvatusopillinen Aikakauskirja*, 226–238.
- Kiilakoski, T. & Oravakangas, A. 2010. Koulutus tuotantokoneistona? Tulostavoitteinen koulutuspolitiikka kriittisen teorian valossa. *Kasvatus & Aika* 4 (1), 7–25.
- Kivirauma, Joel 1998. Sivistyshistoriaa, faktoreita ja tulkintaa. *Suomalaiset kasvatustieteen väitöskirjat 1900-luvulla*. *Kasvatus* 29 (1), 13–22.
- Kivirauma, Joel & Kivinen, Osmo 1988. *The School System and Special Education: Causes and Effects in the Twentieth Century*. *Disability, Handicap & Society* 3 (1), 153–164.
- Konttinen, Raimo 1995. Arvostelusta näyttöön. Koulutuksen arvioinnin kehityspiirteitä Suomessa. *Kasvatus* 26 (1), 6–14.
- Koskenniemi, Matti & Grenquist, Pekka 1934. *Kokeita suomalaisilla joukkotesteillä I*. *Kasvatus ja koulu*, 189–203.
- Lehtovaara, Arvo 1952. Oppikoulujemme pääsyutkunnon ongelmia. *Kasvatusopillinen Aikakauskirja* 43–51.
- Lehtovaara, Arvo 1963. Empiirisen pedagogiikan tämänhetkinen vaihe Suomessa. *Kasvatusopillinen aikakauskirja* 99, 105–109.
- Leimu, Kimmo 1974. Opetustoimen evaluaatiotyön hahmotusta ja muuan sovellutus. *Jyväskylän yliopisto: Kasvatustieteiden tutkimuslaitoksen julkaisuja* 221.
- Lundgren, Ulf 2002. Voidaanko koulujärjestelmää arvioida kansallisella tasolla? Teoksessa Honkonen, Risto (toim.), *Koulutuksen lumo: Retoriikka, politiikka ja arviointi*. Tampere: Tampere University Press, 193–207.
- Mikkonen, Valde & Mikkonen, Juhani 1971. OPSAM. Opintosaavutusten mittaus. Helsinki: Tammi.
- Mäkelä, Jukka 1991. Sunnuntaina sataa aina. Tutkimus tilastollisen ajattelun siirtymisestä osaksi empiiristä sosiaalitutkimusta. *Lapin yliopiston yhteiskuntatieteellisiä julkaisuja*. B Tutkimusraportteja ja selvityksiä 13.
- Nurmi, Veli 1967. *Opetustyö tehokkaaksi*. Porvoo: WSOY.
- Oakley, Ann 2002. *Social Science and Evidence-Based Everything: The Case of Education*. *Educational Review* 54 (3), 277–286.
- Ojakangas, Mika 1997. *Lapsuus ja auktoriteetti. Pedagogisen vallan historia Snellmanista Koskenniemeen*. Helsinki: Tutkijaliitto.
- Peters, Michael 2007. *Styles of Thinking, Styles of Reasoning*. *Educational Philosophy and Theory*, 350–363.
- Pitkänen, Pentti 1960. Käytännön ja teorian vuorovaikutus kokeellisessa kasvatustieteessä. *Kasvatus ja koulu* 47, 93–100.
- Poovey, Mary 1998. *A History of the Modern Fact. Problems of Knowledge in the Sciences of Wealth and Society*. Chicago: University of Chicago Press.

- Rainio, Kullervo 2006. Pakinaa psykologian ja psyyken piiloleikistä. *Psykologia* 40 (3), 240–242.
- Ravitch, Diane 2010. *The Death and Life of the Great American School System. How Testing and Choice Are Undermining Education*. New York: Basic Books.
- Rinne, Risto 2012. Koulutus normaaliuden ja poikkeavuuden historiallisena tuottajana. Teoksessa Silvennoinen, Heikki & Pihlaja, Päivi (toim.), *Rajankäyntejä – tutkimuksia normaaliuden, erilaisuuden ja poikkeavuuden tulkinnoista ja määrittelystä*. Turku: Turun yliopiston kasvatustieteiden tiedekunta, 27–57.
- Rinne, Risto, Simola, Hannu, Mäkinen-Streng, Mirka, Silmäri-Salo, Sari & Varjo, Janne 2011. Arvioinnin arvo: suomalaisen perusopetuksen laadunarviointi rehtoreiden ja opettajien kokemana. Jyväskylä: Suomen kasvatustieteellinen seura.
- Saari, Antti 2008. Kasvatuspsykologian Arkhimedeeseen piste. *Soveltava tutkimus ja hallinta. Kasvatus ja Aika* 2 (1), 41–55.
- Saari, Antti 2011. Kasvatustieteen tiedontahto. Kriittisen historian näkökulmia suomalaiseen kasvatukseen tutkimukseen. Jyväskylä: Suomen kasvatustieteellinen seura.
- Saari, Hannu 1983. Koulukoetoiminnan periaatteista peruskoulussa. Teoksessa Hirvi, Vilho (toim.), *Peruskoulun kehittäminen tutkimustulosten perusteella*. Jyväskylä: Jyväskylän yliopisto.
- Salmela, Alfred 1948. *Kansakoulun oppilasarvostelu*. Helsinki: Otava
- Shapin, Steven & Schaffer, Simon 1985. *Leviathan and the Air-Pump. Hobbes, Boyle and the Experimental Life*. Princeton: Princeton University Press.
- Simola, Hannu 2002a. Ilmaan propattu – Toiveiden rationalismi koulureformien diskursiivisena perustana. Teoksessa Honkonen, Reijo (toim.), *Koulutuksen lumo: retoriikka, politiikka ja arviointi*. Tampere: Tampere University Press, 55–74.
- Simola, Hannu 2002b. From exclusion to self-selection: examination of behaviour in Finnish primary and comprehensive schooling from the 1960s to the 1990s. *History of Education* 31 (3), 207–226.
- Snellman, Johan Vilhelm 2004 [1854]. *Tilastoista*. Kootut teokset 8. Helsinki: Edita, 100–104.
- Swijtink, Zeno 1987. *The Objectification of Observation: Measurement and Statistical Methods in the Nineteenth Century*. Teoksessa Krüger, Lorenz, Gigerenzer, Gerd & Morgan, Mary (toim.), *The Probabilistic Revolution, vol 1*. Cambridge: MIT Press, 261–285.
- Takala, Martti 1963. Ihmistä tutkivien tieteiden kehittymisedellytyksistä. *Kasvatus ja koulu* 50, 153–163.
- Tyler, Ralph 1949. *Basic Principles of Curriculum and Instruction*. Chicago: University of Chicago Press.
- Töttö, Pertti 2000. *Pirullisen positivismin paluu*. Tampere: Vastapaino.
- Vahervuo, Toivo 1951. Suhteellinen ja absoluuttinen arvostelusysteemi. *Kasvatusopillinen Aikakauskirja*, 238–243.
- Vahervuo, Toivo 1952. *Psykometriikan metodeja I. Tilastolliset peruskäsitteet*. Porvoo: WSOY.
- Vahervuo, Toivo 1947. *Oppilaitosten oppilasarvostelu*. *Kasvatusopillinen Aikakauskirja*, 5–27.
- Vahervuo, Toivo & Ahmavaara, Yrjö 1958. *Johdatus faktorianalyysiin*. Porvoo: WSOY.

Wiliam, Dylan 2010. What Counts as Evidence of Educational Achievement? The Role of Constructs in the Pursuit of Equity in Assessment. *Review of Research in Education* 34, 254–283.

KT Antti Saari toimii yliopistonlehtorina Tampereen yliopiston Kasvatustieteiden yksikössä.