

VATT VALMISTELURAPORTIT

16

Kouluvaikutus ja tuloksellisuusrahoitus lukiokoulutuksessa

Laura Ansala
Jenni Pääkkönen

Laura Ansala, Valtion taloudellinen tutkimuskeskus ja Aalto-yliopiston
kauppakorkeakoulu

Jenni Pääkkönen, Valtion taloudellinen tutkimuskeskus

ISBN 978-952-274-068-7 (nid.)
ISBN 978-952-274-069-4 (PDF)

ISSN 1798-0305 (nid.)
ISSN 1798-0313 (PDF)

Valtion taloudellinen tutkimuskeskus
Government Institute for Economic Research
Arkadiankatu 7, 00100 Helsinki

Edita Prima Oy
Helsinki, maaliskuu 2013

Kansi: Niilas Nordenswan

Kouluvaikutus ja tuloksellisuusrahoitus lukiokoulutuksessa

Valtion taloudellinen tutkimuskeskus VATT Valmisteluraportit 16/2013

Laura Ansala – Jenni Pääkkönen

Tiivistelmä

Tässä tutkimuksessa arvioidaan lukioihin suunnitellun tuloksellisuusrahoituksen perusteita useista näkökulmista. Aluksi käydään läpi kannustejärjestelmien teoriaperusteita ja sen jälkeen perehdytään niiden käytännön sovellutuksiin kouluissa. Lisäksi selvitetään eri maiden rahoitusjärjestelmiä ja tutustutaan muutamaiin tunnetuimpiin tuloksellisuusjärjestelmiin, jotka ovat olleet suosittuja lähinnä Yhdysvalloissa. Lopuksi käydään läpi Suomen ammatillisen peruskoulutuksen järjestelmää ja siitä siirrytään lukioiden tietopohjan arvioimiseen ja laadun estimoimiseen. Tutkimus osoittaa, että lukioiden väliset estimoidut vaikutuserot ovat vähäiset, eikä kouluja voida panna sen perusteella järjestykseen luotettavasti kuin jakauman ääripäissä. Jos lukiota halutaan palkita ”laadun” perusteella, se kannattaa tehdä usean vuoden havaintojen pohjalta, jolloin satunnaisuuden vaikutusta tuloksiin voidaan vähentää.

Asiasanat: koulun vaikutus, lukiokoulutus, tuloksellisuusrahoitus

JEL-luokittelu: H75, I22, I28

Abstract

This study evaluates the performance-based incentives that are planned to be implemented in general upper secondary education in Finland. First we review the principal-agent and tournaments literature and then survey their use at schools. Second, we will present the different funding systems and evaluate performance-based programmes that are used mainly in the US. Lastly, we will evaluate the performance-based funding of vocational education and then move on to the evaluation of the existing data for rewarding and estimating the quality of general upper secondary schools. Our results show that the performance differences between schools are surprisingly small, and one can't reliably rank the schools according to it. If one wishes to rank the schools based on their estimated performance, we recommend quality to be estimated using several year's data to minimize the influence of noise to rankings.

Key words: school effect, general upper secondary education, performance-based funding

JEL classification numbers: H75, I22, I28

Yhteenveto

Erilaiset tuloksellisuusrahoitusjärjestelmät ovat olleet suosittuja koulujen ohjausjärjestelmiä erityisesti anglosaksisissa maissa. Tuloksellisuuteen sidottu rahoitus perustuu kahteen pääperiaatteeseen: ensinnäkin riskien tehokkaaseen jakoon yhtäältä päämiehen ja agentin välillä ja toisaalta erilaisten agenttien kesken. Toinen pääperiaate on, että koulujärjestelmä ei toimi toivotulla tavalla vaan esimerkiksi sen laadussa on puutteita, joita opettajille tai kouluille laadittujen kannusteiden avulla pyritään korjaamaan. Tuloksellisuusrahoituksen tavoitteena on siten tehostaa koulujärjestelmää ja pyrkiä seulomaan, ehkä jopa karsimaan, tehtävästään heikosti suoriutuvat koulut ja opettajat.

Suomessa tuloksellisuusrahoitus on ollut arkipäivää ammatillisen peruskoulutuksen järjestäjille viimeisten kymmenen vuoden ajan. Koulutuksen laadusta ja laadun parantamisesta palkitsevaa osuutta ollaan nyt laajentamassa lukiokoulutukseenkin. Ennen rahoitusperusteiden muuttamista on paikallaan selvittää, mitä tuloksellisuusrahoituksen käytöstä tiedetään. Lisäksi täytyy selvittää, mitä mahdollisuuksia tuloksellisuuteen perustuvan rahoitusjärjestelmän rakentamiseen nykyinen tietopohja tarjoaa. Tässä selvityksessä pureudutaan kannustejärjestelmien soveltavuuteen lukiokoulutuksessa.

Selvitystä tehdessä on tullut ilmeiseksi, että hyvän kannustejärjestelmän perusta ovat selkeästi asetetut ja perustellut tavoitteet. Kun järjestelmää laaditaan ja myöhemmin, kun se markkinoidaan toimijoille, päämiehen täytyy ilmaista mitä ongelmia ja puutteita nykyisessä järjestelmässä sillä halutaan korjata. Vasta kun toimijat, koulut ja opettajat, ymmärtävät ja sisäistävät kannustejärjestelmän tavoitteet, voidaan toivottuja vaikutuksia saada aikaan. Sen lisäksi palkittamisen kohteeksi valittavan toiminnan pitää olla sellaista, että sen kehittämällä voidaan ajatella saavutettavan järjestelmän tavoitteet ja kyetään korjaamaan havaittuja puutteita. Lisäksi, vaikka toimijoita olisi hyvä palkita monipuolisesti, järjestelmä ei saisi olla liian monimutkainen, koska silloin aikaansaadut parannukset jäävät helposti vähäisiksi suhteessa niiden eteen nähtyyn vaivaan ja järjestelmän ylläpidon kustannuksiin. Jos jossakin näistä epäonnistutaan, rahoituksen perusteet ja itse järjestelmä muuttuvat epäselviksi. Vastoin hyvän hallinnon periaatteita järjestelmä ei olisi läpinäkyvä eikä se tulisi aikaansaamaan toivottua muutosta. Pahimmillaan järjestelmää ei koettaisi sen enempää oikeudenmukaiseksi kuin kannustavaksikaan.

Kannustejärjestelmien hyötyihin kuuluvat riskin jakaminen uudelleen eri osapuolten kesken sekä toiminnan tehostaminen ohjaamalla huomio järjestelmän heikosti suoriutuviin osiin. Siinä on myös puutteensa. Se voi kannustaa toimijoita panemaan liikaa painoa palkittavalle toiminnalle muiden tehtävien kustannuksella. Jos palkinta kohdistuu koko kouluun, se voi kannustaa vapaamatkustamiseen eli siihen, että kaikki opettajat eivät ahkeroidakaan yhtä paljon yhteisen tuloksen

eteen. Toimijoilla on myös kannustin manipuloida signaalia, eli huijata. Usein palkitseminen on kuitenkin helpompaa toteuttaa kuin sille käänteinen toiminta eli rankaiseminen matalampien monitorointikustannusten takia.

Hieman yleistäen voidaan todeta, että oppimistuloksista tai niiden parantamisesta palkitsevilla järjestelmällä ei ole aikaansaatu toivottuja tuloksia. Koska useimpien palkitsemista käyttävien maiden yhteiskuntajärjestelmät, koulutusjärjestelmät ja kouluissa havaitut ongelmat poikkeavat huomattavasti Suomesta, tästä ei voida vielä päätellä, että tuloksellisuusrahoitus on tuomittu epäonnistumaan Suomessa. Esimerkiksi Israelin lukiotasoissa kouluissa tehdyt kokeet ovat osoittaneet, että erilaisilla kannusteilla voidaan saada aikaan parannuksia.

Analyysissämme emme löytäneet merkittäviä eroja lukiodien oppimistuloksissa eri koulujen välillä. Tämä ei tarkoita sitä, ettei laadusta palkitsevalla rahoituksella voitaisi saada aikaan laatuparannuksia, mutta tehostamispotentiaalia vain on niukalti. Käytännön soveltamisen ongelmaksi voi muodostua se, että lukion estimoidun vaikutuksen vaihteluun näyttää ainakin pienemmällä kouluilla liittyvän jokin satunnainen komponentti. Vain aivan jakauman ääripäissä koulujen oppimistulosten muutoksella määritellyt sijoitukset pysyivät vakaina vuodesta toiseen. Jotta kouluja ei päädytä rankaisemaan tai palkitsemaan liian satunnaisesti, voi olla hyödyllistä estimoida koulun vaikutusta oppimistuloksiin usealta vuodelta ja palkita kouluja sen perusteella.

Tuloksissamme on viitteitä siitä, että koulun vaikutus oppimistuloksiin vaihtelee taustaltaan erilaisten oppilasryhmien kesken. Koska meiltä kuitenkin puuttui lähes kaikki oppilaan taustaa kontrolloiva informaatio, esimerkiksi vanhempien koulutus- ja tulotaso, voivat saamamme tulokset johtua näiden kontrollien puutteesta. Tätä seikkaa olisi siksi syytä arvioida tarkemmin. Jos tulokset säilyvät näiden huomioon ottamisen jälkeenkin, kouluilla on potentiaalia tehostaa eri oppilasryhmien suoriutumista ylioppilaskirjoituksissa. Tällöin politiikka, joka muiden maiden esimerkin mukaan kiinnittää huomiota tiettyihin, osaamispotentiaaliinsa nähden alisuoriutuviin oppilasryhmiin voi olla tuloksekasta. Israelissa toteutettujen kokeiden toistaminen joissakin Suomen lukioissa voisi tuottaa päättöksentekijöille kaivattua lisäinformaatiota.

Sisällys

1 Johdanto	1
2 Päämies-agentti -ongelma ja sen soveltaminen koulutukseen	6
2.1 Päämies-agentti -kehikon perusteet	6
2.2 Turnajaisteorian perusteet	12
2.2.1 Homogeeniset ja riskineutraalit kilpailijat	14
2.2.2 Homogeeniset ja riskiä kaihtavat kilpailijat	16
2.2.3 Heterogeeniset ja riskineutraalit kilpailijat	16
2.2.4 Muita teoreettisia lähtökohtia	18
3 Palkitseminen käytännössä	21
3.1 Palkittava yksikkö	21
3.2 Tuotosten mittaaminen	24
3.3 Arvonlisämallit	26
3.4 Standardijärjestelmän mukainen palkitseminen	29
4 Empiirinen tutkimus koulutuksen taloustieteessä	31
4.1 Eksperimenttien tuloksia Israelista	31
4.2 Vastuuohjelmien ja palkitsemisen seurauksia Yhdysvalloissa	35
5 Kokemuksia palkitsemisohjelmista maailmalta	38
5.1 Yhdysvallat	38
5.1.1 Dallas Independentin koulupiiri, Texas	38
5.1.2 Charlotte-Mecklenburgin koulupiiri, North Carolina	40
5.1.3 Kentuckyn osavaltio	42
5.1.4 Kalifornian osavaltio	44
5.2 Chile ⁴⁷	
6 Lukioiden rahoitus- ja arviointijärjestelmiä Euroopassa	51
6.1 Iso-Britannia	51
6.1.1 Lukioiden arviointi	51
6.1.2 Lukioiden rahoitus	52
6.2 Manner-Eurooppa	53
6.2.1 Lukioiden arviointi	53
6.2.2 Lukioiden rahoitus	57

7 Suomi	60
7.1 Suomen lukioiden rahoituksen nykytila	60
7.2 Suomen ammattikoulujen tuloksellisuusrahoitus	61
7.2.1 Mittariston osa-alueet	62
7.2.2 Mittariston soveltuvuus lukioihin	65
7.3 Lukioiden palkitsemiseen soveltuvan aineiston läpikäyminen	67
7.3.1 Koulun vaikutus oppimistuloksiin	68
7.3.2 Rajauksia tuloksiin ja niiden käytettävyyteen	78
8 Johtopäätöksiä ja suosituksia	80
Lähteet	83

1 Johdanto

Koulutuspolitiikka on, ehkä hieman yllättäen, noussut talouspolitiikan keskiöön monissa maissa. Ilmiötä voidaan selittää sillä, että useimmissa teollistuneissa maissa työvoiman kasvu on pysähtynyt ja työvoiman määrä on lähivuosina kääntymässä jopa laskuun. Silloin talouskasvun ainoaksi lähteeksi jää tuottavuuden kasvu, jossa yhtenä keskeisenä tekijänä on inhimillinen pääoma. Lisäksi koulutettu työvoima on yksi talouden kilpailukyvyyn tekijöistä laadukkaiden instituutioiden ja korkean sosiaalisen pääoman ohella. Suomessa, missä koulutuspolitiikka on sitoutunut vahvasti alue- ja sosiaalipolitiikkaan, ollaan vasta heräämässä tähän todellisuuteen. Ja koska koulutus on Suomessa järjestetty pääosin julkisin varoin, tulevat julkisen talouden uudistamispaaineet näkymään myös tällä sektorilla. Siksi huolellisesti harkittu ja pohjustettu politiikka on entistäkin tärkeämpää Suomen tulevaisuuden kannalta.

Vaikeassa taloustilanteessa muodostettu uusi hallitus ottaa kantaa koulutuspolitiikkaan hallitusohjelmassaan (Valtioneuvoston kanslia, 2011). Hallitus kaavaillee, että lukioiden rahoituksen perusteisiin tullaan kuluvan hallituskauden aikana sisällyttämään tuloksellisuusrahoitusosuus, ”joka palkitsee koulutuksen laadusta ja laadun parantamisesta, esimerkiksi yksilön oppimistulosten paranemisesta ja vähäisistä keskeyttämisistä”. Ennen rahoitusperusteiden radikaalia muuttamista on paikallaan selvittää, mitä tuloksellisuusrahoituksen käytöstä tiedetään. Lisäksi täytyy selvittää mitä mahdollisuuksia tuloksellisuuteen perustuvan rahoitusjärjestelmän rakentamiseen nykyinen tietopohja tarjoaa. Tässä selvityksessä pureudutaan kannustejärjestelmien soveltuvuuteen lukiokoulutuksessa.

Kouluihin ulotettu tulosvastuu ei tietenkään ole mikään uusi tai erityisen suomalainen ilmiö, vaan kouluja on asetettu vastuuseen niiden tuloksista vuosikymmenten ajan useissa eri maissa. OECD:n ”Education at a Glance 2011” -raportissa on esitelty laajasti eri maissa käytössä olevia tuloksellisuusjärjestelmiä, jotka voivat perustua kolmeen erilaiseen ohjausmekanismiin: tuotosten arviointiin perustuvat järjestelmät, hallinnolliseen sääntelyyn perustuvat järjestelmät tai markkinamekanismiin, nk. kouluvalintaan, perustuvat järjestelmät. Useissa OECD:n jäsenmaissa lukiotasoisessa koulutuksessa on käytössä näiden kolmen järjestelmän jokin yhdistelmä.

Suomea voitaneen pitää koulujen arvioinnin suhteen jonkinlaisena poikkeuksena, sillä Suomessa sääntely ja valvonta näyttävät olevan muita maita vähäisempää (tarkemmin Eurydice, 2012). Suomessa keskushallinto ei arvioi lukioita, sen paremmin kuin muitakaan kouluja, lainkaan niiden tulosten perusteella, vaikka tietoa tähän olisi tarjolla.¹ Myöskään sääntöjen noudattamisen ulkoista arviointia ei

¹ Sen sijaan mediat tuottavat joka kevät omia lukioiden ”ranking-listojaan”, jotka perustuvat lähinnä ylioppilaskirjoitusten tuloksiin. Suomessa lukioiden arviointi on siis jätetty alan ammattilaisten sijaan mediataloille!

Suomessa tehdä, vaikka se on sangen yleistä muissa maissa (26 maata). Sen sijaan vastuu koulutuksen arvioinnista on koulutuksen järjestäjällä, mutta arvioinnin sisällöstä tai tiheydestä ei ole määräyksiä. Käytännössä kunnat seuraavat kouluja vuosittain, kun koulut laativat budjettinsa ja liittävät siihen tietoja oppilaiden määristä ja opettajien kelpoisuudesta. Luultavasti sen enempää koulujen kuin kuntienkaan johtajilla ei ole kannusteita arvioida koulutyön laatua kriittisesti. Ja vaikka kuntien valtuustoilla ja oppilaiden vanhemmilla olisikin intressi seurata koulun opetusta, ei hajautettu valvonta useinkaan toimi.

Vuoden 2010 kustannustietojen mukaan Suomen lukioiden ylläpidosta koituvat menot olivat noin 630 miljoonaa euroa, joista valtio korvasi kunnille noin 42 prosenttia. Vuoden 2013 talousarvioesityksessä lukiokoulutuksen menoiksi oli arvioitu 714 miljoonaa euroa, josta valtion osuus tulee olemaan noin 290 miljoonaa euroa arvion toteutuessa (VM, 2012). Muista maista poiketen Suomessa satojen miljoonien eurojen vuosittaista rahankäyttöä ja sen kohdentamista ei valvota keskushallinnon toimesta, vaikka auditointi ja läpinäkyvyys ovat hyvän hallinnon kulmakiviä.²

Keskitetyn valvonnan sijaan Suomessa lukioiden laadunvalvonta perustuu tällä hetkellä vahvimmin ”markkinamekanismiin”, sillä koululaiset saavat vapaasti hakeutua haluamaansa lukioon. Käytännössä lukiovalintaa rajoittaa suurten asutuskeskusten ulkopuolella se, että paikkakunnalla on tyypillisesti yksi lukio ja etäisyys seuraavaan lähimpään lukioon on pitkä, jolloin koulujen välistä kilpailua oppilaista ei näillä paikkakunnilla synny. Markkinamekanismi toimii siis lähinnä suurissa asutuskeskuksissa, joissa on useita lukioita ja hyvät julkiset liikenneyhteydet. Oppilaan oikeuksien toteutumisen ja hyvän hallinnon kannalta keskushallinnon pohtima tuloksellisuusrahoitus, johon liittyy laadunarviointi, voi hyvinkin olla askel parempaan suuntaan.

Yhteiskunta on antanut koulutustehtävän lukioille, mistä syntyy taloustieteissä hyvin tunnettu päämies-agentti -suhde. Myös kannustejärjestelmät, millaiseksi tuloksellisuusrahoituskin luetaan, pohjautuvat päämies-agentti -malleihin. Siksi tässä selvityksessä pohditaan koulutuksen tuottavuuden mittaamisen filosofiaa nimenomaan päämies-agentti -ongelman kehikossa. Se onkin luontevaa, sillä urauurtavassa artikkelissaan Holmström ja Milgrom (1991) perustelevat päämies-agentti -kehikon hyödyllisyyttä muun muassa koulutuksen sovelluksilla, erityisesti tuolloin USA:ssa pinnalla olleella ajatuksella sitoa opettajien palkat oppilaiden koemenestykseen. Näin myös myöhemmin tehtiin, ja aiheesta saadut empiiriset tutkimustulokset ovat hyvin mielenkiintoisia.

Vaikka koulutuksessa päämies-agentti -suhde voisi syntyä myös koulun ja oppilaiden välille, tämän selvityksen näkökulmasta päämies on ensisijaisesti yhteis-

² Koko koulutussektorille ohjautui julkisia varoja lähes 12 miljardin edestä vuonna 2010. Hyvän hallinnon määrittelystä lisää YK:n UNPANin internet -sivuilta.

kunta (julkinen sektori, valtio, kunta) ja agentti on koulu (joskus opettaja). Tämä rajaus on valittu siksi, että koulutus maksetaan yhteiskunnan varoista ja keskeiset koulutusta ohjaavat päätökset tehdään keskitetysti. Todellisuudessa lukiot joutuvat myös kilpailemaan potentiaalisista oppilaistaan, jolloin koulun asiakkaina toimivat sen oppilaat. Jotta ongelma ei olisi liian yksinkertainen, voidaan oppilaat nähdä myös yhteistuottajina (co-production, mm. Ostrom, 1996), joiden oma motivaatio ja aktiivisuus vaikuttavat oppimistuloksiin. Tällöin oppilaan oma työpanos ja opettajan antama opetus ovat epätäydellisiä substituutteja; ideaalitilanteessa sekä oppilas että opettaja panostavat oppimiseen. Myös kannustejärjestelmä voidaan rakentaa yhteistuottaja-näkökulma huomioiden. Tällöin ei siis riitä, että koulu palkitaan menestyksestä, vaan palkitseminen on syytä ulottaa myös oppilaisiin. Samalla tavalla, koska opettaminen on opettajan tehtävä, on mahdollista ja oikeastaan todennäköistä, että ideaalitapauksessa kannustejärjestelmä palkitsee suoraan opettajaa, ei koulua.

Tässä raportissa esiteltävä (kansan-) taloustieteellinen tutkimus rakentuu pääasiassa koulutuksen tuotantofunktion estimoimisen varaan. Koulutuksen tuotokset voidaan johtaa kouluille lailla annetuista tavoitteista ja tehtävistä ja panokset, joista keskustelemme tarkemmin jäljempänä, ovat esimerkiksi koulutustehtävän toteuttamiseen annetut resurssit. Vuonna 1998 annetun lukiolain (21.8.1998/629) mukaan lukiokoulutuksen tavoite on ”tukea opiskelijoiden kasvamista hyväksi, tasapainoiseksi ja sivistyneiksi ihmisiksi ja yhteiskunnan jäseniksi sekä antaa opiskelijoille jatko-opintojen, työelämän, harrastusten sekä persoonallisuuden monipuolisen kehittämisen kannalta tarpeellisia tietoja ja taitoja. Lisäksi koulutuksen tulee tukea opiskelijoiden edellytyksiä elinikäiseen oppimiseen ja itsensä kehittämiseen elämänsä aikana.”

Suurin osa lukioille asetetuista tavoitteista on edellä mainitun esimerkin tavoin abstrakteja, jolloin asetetun tavoitteen saavuttamista on vaikea arvioida.³ Karkeasti ottaen lukioille asetetut tavoitteet voidaan jakaa tiedollisiin ja oppilaan henkistä kehitystä tukeviin tavoitteisiin, joista erityisesti jälkimmäisiä on erittäin vaikea mitata. Mittaamisen vaikeudesta johtuen lukion onnistumista tyypillisesti arvioidaan niillä tiedoilla, jotka on helppo saada. Käytännössä koulujen onnistumista on arvioitu ensisijaisesti oppimistuloksilla, eli koulun menestymisellä ylioppilaskirjoituksissa. Tämä näkökulma pelkistää lukion tavoitteeksi suuria tietomääriä ja asiakokonaisuuksia hallitsevien ihmisten tuottamisen.⁴ Näkökulma on suppea ja siinä on puutteensa, mutta selkeytensä vuoksi pitäydymme siinä.

Koska lukiokoulutus tähtää jatko-opintokelpoisuuden saavuttamiseen, toinen lukion onnistumisen mittareista on sen oppilaiden sujuva siirtyminen jatko-opintoihin. Lukion suorittaneiden jatkokoulutusvalmiuksia korkeakoulutuksen

³ Esimerkiksi, mitä tarkoittaa kasvamisen ”hyväksi, tasapainoiseksi ja sivistyneeksi” ihmiseksi?

⁴ Oppilaiden kasvamista tasapainoiseksi yhteiskunnan jäseniksi voitaneen edistää oppilashuollon toimilla, mutta näiden panostusten vaikutuksia ei nykytiedon valossa ole mahdollista arvioida.

näkökulmasta on hiljattain selvittänyt muun muassa koulutuksen arviointineuvosto (Hautamäki ym. 2012). Vuonna 2007 ylioppilastutkintoon ilmoittautuneista vastaajista oli vuonna 2011 toteutetussa kyselyssä 45 prosenttia sitä mieltä, että lukio-opetusta pitäisi muuttaa, jotta se valmentaisi paremmin opintoihin korkeakouluissa. Korkeakoulujen henkilökunnan haastattelut osaltaan vahvistavat tätä näkemystä, sillä korkeakouluissa on jouduttu madaltamaan tavoitteita ja supistamaan opetusmateriaalin sisältöjä opiskelijoiden heikkojen pohjatietojen vuoksi. Vaikka tähän tulokseen oletettavasti vaikuttavat myös korkeakoulujen kasvaneet aloituspaikkamäärät ja korkeakoulujen halu täyttää ne hakijoiden taitotasosta riippumatta, lukio-opintojen ohjauksessa ja opintojen sisällöissä entisten lukio-laisten mielestä on aidosti kehittämisen varaa.⁵ Hautamäki ym. tuottivatkin arvokasta tietoa lukion opetuksen kehittämiseen ja koulutuksen laadun parantamiseen.

Yllä mainitut oppimistuloksia ja jatko-opintoihin siirtymistä kuvaavat mittarit on mainittu hallitusohjelmassa potentiaalisina tuotosmittareina. Tässä selvityksessä pohditaan tavoitteiden mittaamisen problematiikkaa erityisesti siitä näkökulmasta, riittääkö näiden helposti mitattavien tavoitteiden mittaaminen ja niihin perustuva palkitseminen ohjaamaan lukioiden työtä. Vähemmälle huomiolle tässä työssä jää se, mitä palkitsemisella halutaan saavuttaa ja ohjaako palkitseminen kouluja haluttuun suuntaan. Palkitsemisen tavoitteisiin tai ongelmiin, joita palkitsemisella halutaan korjata, ei oteta kantaa siitä syystä, että niitä ei ole lausuttu ääneen ainakaan hallitusohjelmassa eikä tiedossamme ole Suomen lukiokoulutuksessa ilmenneitä puutteita, joita ”laadusta” palkitsevalla ohjelmalla voitaisiin korjata. Kuitenkin selkeä tavoitteenasettelu ja ohjelman motivointi mahdollisilla koulutuksessa ilmenneillä puutteilla on ensiarvoista, jotta ohjelma olisi vaikuttava ja sitä alettaisiin kouluissa toteuttaa.

Selvitys rakentuu seuraavasti. Toisessa luvussa käydään lyhyesti läpi päämies-agentti -problematiikkaa, sillä alan teoreettinen kirjallisuus tarjoaa useita hyödyllisiä näkökulmia kannustejärjestelmien suunnittelijoille. Selvityksen kolmannessa luvussa paneudutaan optimaalisen palkitsemisjärjestelmän käytännön kysymyksiin ja neljännessä luvussa käydään läpi aihetta sivuavaa empiiristä tutkimusta. Tätä aihetta koskevaa laajaa tutkimusta käydään läpi niiltä osin, kun sillä on tarjottavaa Suomen koulutuksen kentälle.

Selvityksen viidennessä luvussa käydään läpi Yhdysvaltojen ja Chilen kokemuksia tuloksellisuusrahoituksesta. Vertailukelpoisia järjestelmiä löytyi niukasti, sillä suunnitellun kaltaiset tuloksellisuusrahoitusjärjestelmät eivät näytä lopulta olleen kovinkaan suosittuja, eivätkä harvat tuloksellisuusrahoituskokeilut välttämättä olleet onnistuneita. Esimerkiksi Chilen kuuluisa SNED -järjestelmä lähinnä jakoi rahoitusta uudelleen ilman merkittäviä oppimistulosparannuksia edes palkituissa kouluissa. Eri maiden järjestelmissä ilmenneitä hyötyjä ja haittoja arvioitaessa

⁵ Tarkemmin Hautamäki ym. (2012).

täytyy pitää mielessä, millaisia ongelmia näillä järjestelmillä on yritetty korjata. Jos kyseisiä puutteita tai ongelmia ei havaita suomalaisessa koululaitoksessa, eivät esitetyt hyödyt tai haitat silloin päde täällä. Toisaalta ratkaisu, joka toimii tietyssä yhteiskunnassa, ei välttämättä ole patenttiratkaisu, joka toimisi kaikissa yhteiskunnissa. Järjestelmien hyödyt ja haitat ovat siis riippuvaisia taustalla olevasta yhteiskuntajärjestelmästä ja sen ongelmista. Järjestelmän eroista johtuvat tuloksiin ja johtopäätöksiin tehtävät rajaukset mainitaan niiden esittelyn yhteydessä.

Joitakin lukijoita voi erityisesti kiinnostaa tieto siitä, miten arvioinnin kohteeksi joutuneet opettajat ovat ottaneet vastaan tuloksista palkitsevat tai rankaisevat järjestelmät. Näitä tuloksia on lyhyesti esitelty luvussa 4 sekä erityisesti luvussa 5 esiteltävissä Charlotte-Mecklenburgin koulupiirin ja Kentuckyn osavaltion järjestelmien tulosten kuvauksissa. Muutamaa poikkeusta lukuun ottamatta tulokset eivät ole kovinkaan rohkaisevia.

Kuudennessa luvussa esitellään rahoitus- ja arviointijärjestelmiä muualla Euroopassa ja seitsemännessä luvussa paneudutaan Suomen järjestelmän kuvailuun ja arviointiin. Tässä yhteydessä esitellään Suomen ammattikouluissa käytössä olevaa tuloksellisuusrahoitusta ja pohditaan lyhyesti sen toteutusta. Sen lisäksi luvussa tarkastellaan lukioiden tuloksellisuusrahoituksen perustaksi soveltuvia, olemassa olevia aineistoja ja niiden asettamia rajoitteita kannustejärjestelmän rakentamiseen. Työn tärkein anti lienee tässä luvussa tehtävä rekisteriaineistoon perustuva analyysi lukioiden välisistä oppimistulosten eroista ja niiden mittaamisesta. Tarkastelu nojautuu sitä edeltävien lukujen tärkeimpiin johtopäätöksiin ja siihen, mitä vastaavista järjestelmistä voidaan oppia.

Luku kahdeksan kokoaa yhteen joitakin suosituksia tuloksellisuusrahoitusjärjestelmän laatijoille: mitä hyötyjä ja mitä mahdollisia ongelmia lukion rahoituksen sitominen sen tuloksiin voi aiheuttaa. Lisäksi pohditaan mahdollisia muita tapoja kannustaa lukioita ja oppilaita.

2 Päämies-agentti -ongelma ja sen soveltaminen koulutukseen

2.1 Päämies-agentti -kehikon perusteet

Miten saadaan henkilö eli agentti (tässä opettaja tai koulu) tekemään jotain, josta hänelle yksilönä aiheutuu kustannuksia, mutta joka hyödyttää toista osapuolta (päämiestä eli valtionhallintoa)? Tämä kannusteisiin liittyvä kysymys on yksi taloudellisen toiminnan keskeisimmistä kysymyksistä ja sitä on tutkittu ja mallinnettu laajasti. Tässä luvussa luodaan katsaus teoriamallien tuloksiin sekä kannusteiden käytöstä saatuun empiiriseen evidenssiin. Koska aiheesta on kirjoitettu paljon, olemme rajoittaneet katsauksen keskeisiin artikkeleihin, jotka ovat julkaistu alan johtavissa lehdissä. Empiirisessä kirjallisuudessa paino on niin ikään koulutuksen sovelluksissa ja niiden tuottamissa tuloksissa. Aluksi tehdään lyhyt katsaus mallien taustalla vaikuttaviin käsitteisiin.

Päämies-agentti -mallin lähtökohta on, että päämies palkkaa itselleen agentin tekemään jonkin työsuorituksen, joka tuottaa hyötyä tai tuloa päämiehelle ja josta koituu agentille vaivaa. Koska ponnistelusta aiheutuu agentille kustannuksia, on luontevaa olettaa agentin haluavan työskennellä vähemmän kuin olisi muutoin optimaalista.

Informaatio agentin ja päämiehen välillä on epäsymmetristä ja agentin tuotokseen liittyy epävarmuutta. Epävarmuus tarkoittaa sitä, että agentin tuotokseen vaikuttaa jokin hänestä riippumaton tekijä, jota agentti ei pysty ennakoimaan. Tällöin agentti arvioi *odotettua* hyötyään perinteisessä tilastotieteen odotusarvon merkityksessä. Agentti tekee päätöksensä ennen kuin hänestä riippumaton tekijä havaitaan. Agentin *todellinen* hyöty realisoituu hänen toimimensa ja havaitsemattoman tekijän toteutumisen jälkeen. Lisäksi, koska tuotokseen vaikuttaa joku agentista johtumaton tekijä, päämies ei voi määrittää täsmällisesti agentin osuutta hyvään tai huonoon tulokseen. Esimerkiksi koulutuksessa yksittäisen opettajan tuotosta arvioitaessa tällainen hänen toimistaan riippumaton tekijä on vaikkapa oppilaskohortin kyvykkyys. Tällöin pelkästään tuotokseen perustuva palkkaus siirtää koko tuotoksen vaihtelusta aiheutuvan riskin agentin kannettavaksi, mikä ei välttämättä ole tehokas ratkaisu. Kiinteä palkka taas siirtää tuotoksen vaihtelusta aiheutuvan riskin päämiehen kannettavaksi, mikä sekään ei ole tehokas ratkaisu.

Useamman agentin tapauksessa epävarmuus tuotoksesta syntyy siitä, että kunkin yksittäisen agentin kontribuutiota yhteiseen tuotokseen ei havaita. Kouluesimerkissä tämä tarkoittaisi sitä, että oppilaiden opettamiseen ja motivoimiseen osallistuvat kaikki opettajat, mutta yksittäisen opettajan panosta oppilaan oppimistuloksissa ei voitaisi erottaa muiden panoksesta.

Informaation epäsymmetrisyys tarkoittaa sitä, että toisella osapuolella on enemmän informaatiota kuin toisella ja hän voi käyttää tätä informaatiota omaksi edukseen. Yhden agentin tapauksessa tämä tarkoittaa sitä, että agentti tietää oman panoksensa, mutta päämies ei. Usean agentin tapauksessa informaation epäsymmetrisyys tarkoittaa sitä, että päämies ei tiedä tarpeeksi agenttiensa kyvykkyydestä, ahkeruudesta tms., joten hän pyrkii rakentamaan järjestelmän, jossa agentit paljastavat tyyppinsä. Useamman kuin yhden agentin mallissa ahkerat tai kyvykkäät agentit haluavat palkkion toivossa, tai rangaistuksen välttämiseksi, erottautua laiskoista ja lahjattomista kilpailijoistaan. Jälkimmäiset puolestaan haluavat peittää todellisen tyyppinsä. Pahimmillaan, jos päämies ei pysty erottelemaan hyviä ja huonoja agenteja toisistaan, voidaan ajautua tilanteeseen, jossa ahkeratkin agentit laiskottelevat. Kyvykkäiden ja vähemmän kyvykkäiden erottelu voidaan toteuttaa kilpailun avulla.

Perinteisiä esimerkkejä koulumaailman kannusteista ja oppilaiden erottelusta ovat kokeet ja pääsykokeet, joissa oppilaiden kyvykkyyttä pyritään mittaamaan.⁶ Koulujen kokeet eivät siis pelkästään mittaa oppilaiden oppimista, vaan panevat oppilaat järjestykseen sen suhteen, miten hyvin he ovat asiat omaksuneet. Samalla tavalla lahjakkuuksien seulominen oli ylioppilaskirjoitusten alkuperäinen tarkoitus, kun Turun Akatemia alkoi sen avulla arvioida hakijoitaan. Näin ollen suomalainen ylioppilastutkinto itse asiassa on jo kannustejärjestelmä, nk. high stake -test, jossa pärjäämisestä seuraa palkkio.

Kuten aikaisemmin mainittiin, kansantaloustieteessä riskillä tarkoitetaan tuloon liittyvää vaihtelua ja vaihtelun tuomaa epävarmuutta: mitä suurempi vaihtelu, sitä korkeampi riski. Riskiaversiivisuus taas kertoo agentin tavasta suhteuttaa varmaa ja epävarmaa tuloa toisiinsa. Päämies-agentti -mallissa oleelliset riskikäsitteet riskineutraali ja riskiä kaihtava on helpointa ymmärtää esimerkin avulla. Ajatellaan, että agentti voi osallistua arvontaan, vaikkapa kolikon heittoon, jossa hän voi voittaa 100 euroa 50 prosentin todennäköisyydellä (kruuna) tai jäädä ilman palkintoa 50 prosentin todennäköisyydellä (klaava). Riskineutraali agentti olisi valmis osallistumaan peliin enintään panoksella, joka vastaa pelin odotettua tuottoa, ts. 50 eurolla. Riskiä kaihtava puolestaan ei koskaan ole valmis panostamaan peliin 50 euroa, vaan vähemmän. Hän siis pitää varmaa tuloa parempana kuin epävarmaa tuloa, vaikka summat ovat täsmälleen samat. Mitä vähemmän henkilö on valmis panostamaan peliin, sitä kaihtavampi hänen sanotaan olevan riskin suhteen. Sopimuksissa agentin kyky kantaa riskiä otetaan huomioon siten, että ideaalitulanteessa erilaiset agentit kantavat riskiä suhteessa riskinottohalukkuuteensa: suurin riski kohdentuu niille, joilla on eniten halua ottaa riskiä. Riskinottohalukkuuden ajatellaan kasvavan tulojen tai varallisuuden mukaan, joten

⁶ Taloustieteessä tunnetaan myös tilanne, jossa ”lahjaton” agentti pyrkii manipuloimaan päämiehen saamaa signaalia ahkeruudella ”liikaa”, jotta hän näyttäisi ”lahjakkaalta” agentilta. Yksinkertaisuuden vuoksi emme käsittele tässä teorialleja, joissa agentilla on mahdollisuus manipuloida signaalia. Esimerkki koulumaailmasta voisi olla oppilaan arvosanaan suotuisasti vaikuttava tuntiaktiivisuus, jolloin kokeessa heikosti menestynyt henkilö voi yrittää vaikuttaa lopulliseen arvosanaan viittaamalla ahkerasti.

varakas henkilö on valmiimpi ottamaan riskiä, koska hänen taloutensa kestää tappioita paremmin kuin vähävaraisen henkilön talous. Yleisesti päämies-agentti-malleissa oletetaan, että agentti on riskiä kaihtava ja päämies riskin suhteen neutraali. Hyvä sopimus ottaa huomioon agentin riskiaversiivisuuden ja sopeuttaa kiinteän korvauksen ja tulospalkkion suhteen sen mukaiseksi.

Mitä tulee riskin uudelleen jakamiseen eri osapuolten kesken, yksi hyvinvointitutkimuksen keskeisimmistä argumenteista, Pareto-parannus ja Pareto-optimi, ovat käsitteinä olennaisia. Pareto-parannus tarkoittaa sitä, että jonkun tilannetta voidaan parantaa heikentämättä toisen asemaa, eli alkuperäinen allokaatio oli tehoton. Jos jonkin sopimuksen avulla riskiä voidaan jakaa uudelleen tehokkaammin niin, ettei kenenkään sopijaosapuolen asema heikkene sen tuloksena, voidaan silloin puhua Pareto-parannuksesta, koska tulos ilman sopimusta olisi huonompi. Pareto-optimaalinen sopimus on sellainen, jossa Pareto-parannuksia ei voida saavuttaa.

Valtaosa teoriamalleista ja niistä johdetuista tuloksista perustuvat staattisiin sopimuksiin eli tilanteisiin, joissa päämiehen ja agentin välinen suhde tai sopimus on kertaluontoinen. Dynaamisissa malleissa sopimus muotoillaan uudelleen joka periodilla, jolloin sopimuksen sisältöön vaikuttaa agentin historia, ts. menestyminen aiemmilla periodeilla. Sopimuksen uusimisesta tehdyillä oletuksilla on luonnollisesti vaikutusta optimaaliseen sopimusratkaisuun, sillä epävarmuudella on taipumus vähentyä toistojen myötä. Toisin sanoen toistettujen sopimusten maailmassa päämiehen on helpompi erottaa ahkerat ja laiskat agentit toisistaan, sillä tuotokseen vaikuttavan muun informaation merkitys vähenee ajan mittaan. Tyyppillisesti on ajateltu, että julkisella sektorilla tehdyt sopimukset ovat luonteeltaan pikemminkin staattisia kuin dynaamisia (mm. Gersbach ja Keil, 2005). Siksi dynaamisia sopimuksia tai niiden ratkaisuja ei tässä yhteydessä käsitellä tämän enempää. On kuitenkin syytä huomata, että koulumaailmassa mikään ei estä siirtymistä käyttämään dynaamisia sopimuksia, jos uskotaan, että sillä tavoin voidaan tehostaa koulujen tai opettajien kannustimia. Tämä tarkoittaisi käytännössä sitä, että koulun toimilupa olisi määräaikainen tai opettajien virat olisivat määräaikaisia ja ne olisivat ehdollisia koulun tai opettajan onnistumiselle.

Optimaalinen palkitsemisjärjestelmä olisi helppo rakentaa, jos päämies pystyisi havaitsemaan agentin työpanoksen. Tällöin agentille voitaisiin maksaa hänen panoksensa mukaan ja häntä voitaisiin rangaista laiskottelusta. Koko päämies-agentti -ongelman ydin on nimenomaan siinä, että päämies ei havaitse agentin panosta, hän havaitsee vain tuotoksen. Holmströmin (1979) yhden agentin mallissa oletetaan, että tuotokseen vaikuttavat agentin työpanoksen ohella muutkin tekijät ja niiden on luonteeltaan satunnaista. Päämies ei siis varmuudella tiedä, johtuuko tuotoksessa havaittu vaihtelu agentin työpanoksen vaihtelusta vai jonkin satunnaisen, agentista riippumattoman tekijän vaihtelusta. Varmistaakseen, että agentti työskentelee tehokkaasti eikä laiskottele, hän laatii kannustejärjestelmän, eli ns. kannustesopimuksen. Siinä agenttia palkitaan jonkun havaitun sig-

naalin perusteella siten, että kun signaali on positiivinen, agentti saa osan tuotosta tai voitosta. Näin tuotoksen vaihtelusta aiheutuvaa riskiä jaetaan sopimuksen avulla uudelleen agentin ja päämiehen välillä.

Yksi päämies-agentti -mallin keskeisistä tuloksista liittyy signaaleihin ja informaatioon. Holmström (1979) osoittaa, että kaikkia, myös informaatioarvoltaan heikkoja, signaaleja voidaan käyttää kannustesopimuksen perusteena kunhan ne sisältävät *jotain* informaatiota agentin toiminnasta. Tällainen sopimus parantaa sekä päämiehen että agentin hyvinvointia, koska se allokoii riskiä uudelleen. Tätä tulosta heikentää hieman se, että jos signaali on hyvin heikko, toisin sanoen sisältää runsaasti agentin toimista riippumatonta informaatiota ja jos sen hankkiminen on päämiehelle kallista, niin silloin kustannushyöty-suhde voi osoittautua negatiiviseksi. Tällöin tosin voi tulla kyseeseen informaation hankinnan ehdollistaminen tuotokselle: jos tuotos x on matalampi kuin joku \bar{x} , niin silloin päämies voi hankkia lisäinformaatiota agentin toiminnasta. Tämä kriittinen havainto kustannusten ja hyötyjen suhteesta kannattaa pitää mielessä, jos Suomen lukioihin aletaan rakentaa palkitsemisjärjestelmää. Mitä informaatiota päätöksentekijöillä on saatavilla, onko se luotettavaa, saako sen kohtuullisella vaivalla kohtuulliseen hintaan?

Edellä kuvattua Holmströmin (1979) tuloksen soveltamisalaa rajoittaa se, että mallissa agentilla on yksi ainoa työtehtävä, kun todellisuudessa yhdellä agentilla on useita työtehtäviä, joiden kesken hänen täytyy jakaa työaikaansa. Toinen rajoite on, että päämies voi palkita häntä vain yhden (havaitun) suorituksen perusteella. Agentin kustannukset työnteosta riippuvat kokonaistyöpanoksesta, jolloin jos palkitsemista yhdestä toiminnasta kasvatetaan, kannustaa se agenttia suuntaamaan huomiotaan enemmän tähän työhön muiden tehtävien kustannuksella. Holmström ja Milgrom (1991) osoittavat, että kiinteä palkka, jossa kannustepalkalle ei jätetä sijaa, voi olla tällöin optimaalinen. Erityisesti, mitä vaikeampaa on arvioida agentin suoriutumista muissa tehtävissä, sitä vähemmän suositeltavaa kannustepalkkauksen käyttäminen on.

Holmströmin ja Milgromin oma käytännön esimerkki koskee koulumaailmaa ja nimenomaan opettajien kannustepalkkoja. Kouluissa perustaitoja ja vaativia taitoja tuotetaan samalla panoksella, mutta vaativien taitojen opettaminen edellyttää suurempaa panostusta kuin perustaitojen opettaminen. Jos palkitseminen sidotaan helpommin mitattaviin perustaitoihin, palkitseminen voi johtaa siihen, että opettajat keskittyvät perustaitojen opettamiseen vaativampien taitojen opettamisen kustannuksella. Tämä ilmiö tunnetaan yleisesti nimellä ”teaching to the test”, eli opettajat keskittyvät opettamaan perustaitoja koetta varten. Näissä tilanteissa ei kannustepalkkauksen käyttö ole suositeltavaa.

Holmström (1982) rakentaa myös mallin, jossa päämiehellä on useita agenteja tekemässä yhteistyötä. Koska tiimityössä yksittäisen agentin panosta ei havaita, mutta jokaisen agentin työpanos vaikuttaa yhteiseen tuotokseen, yksittäisellä

agentilla ja siten kaikilla agenteilla, on kannuste vapaamatkustaa ja nauttia muiden työn hedelmistä, jolloin yhteenlaskettu työpanos on liian vähäinen optimaaliseen tilanteeseen nähden. Holmströmin (1982) mallissa havaittuun tuotokseen liittyy siten epävarmuutta kahdella tapaa: yksittäisen agentin panosta ei voida erottaa muiden panoksesta ja tuotokseen voi sen lisäksi liittyä agenttien toimista riippumatonta vaihtelua. Mallin soveltamisessa oleellista on agenttien yhteinen vastuu tuotoksesta; jos kunkin agentin tuotos on täysin riippumaton muiden agenttien toimista – ja yksittäisen agentin tuotos siten voidaan mitata – palautuu ongelma myöhemmin esiteltävään turnajaiskehikkoon.

Holmströmin (1982) tiimityömallissa havaittuun tuotokseen liittyvä epävarmuus johtaa kahteen erilaiseen päätöksentekosääntöön. Jos havaittuun tuotokseen ei liity muuta epävarmuutta kuin yksittäisen agentin osuus sen tuotannossa, päämiehen ainoa tehtävä on palkita tai rangaista *kaikkia* agenteja havaitun yhteisen tuotoksen mukaan (nk. ryhmärangaiseminen tai -palkitseminen). Päämiehen ei siis kannata kerätä informaatiota agenttien toimista vaan hän voi palkita näitä aivan satunnaisesti. Koulumaailmassa tämä tarkoittaisi esimerkiksi sitä, että jos ryhmän eli opettajien tuotokset – oppilaiden oppimistulokset – olisivat tiettyä ennalta määrättyä raja-arvoa korkeammat, päämiehen tehtävä olisi jakaa ylijäämä opettajien kesken. Teorian mukaan tällainen palkitseminen on omiaan lisäämään agenttien ja päämiehen hyvinvointia. Jos taas havaittuun tuotokseen liittyy vielä joku havaitsematon komponentti, johon agentit eivät toimillaan voi vaikuttaa, ryhmän satunnainen palkitseminen tai rankaiseminen toimii vain siinä tapauksessa, että agentit ovat neutraaleja riskin suhteen.⁷ Jos agentit ovat riskiä kaihtavia, päämiehen kannattaa kerätä informaatiota agenttien toimista ja sitoa palkitseminen tähän informaatioon. Erityisesti, vaikka kannustejärjestelmällä ei kyettäisi vaikuttamaan agenttien työpanokseen, se jakaa riskiä uudelleen agenttien kesken, sillä agenttien kyky sietää riskiä vaihtelee. Kannustejärjestelmän avulla on mahdollista saavuttaa riskin tehokkaampi allokaatio nk. Pareto-parannus.

Päämies-agentti -mallit ovat hyviä työkaluja, kun pohditaan kannusteiden luomista, mutta kuten teoriat yleensä, ne sisältävät paikoin vahvoja oletuksia, jotka eivät välttämättä päde käytännössä. Aikaisemmin kunkin mallin esittelyn yhteydessä mainittiin joitakin kyseisen mallin soveltamiseen liittyviä rajoituksia, mutta mallikehikkoon liittyy myös kaikille malleille yhteisiä rajoittavia ominaisuuksia. Yksi esimerkki tällaisista rajoitteista on malleihin sisältyvä oletus agentin rationaalisuudesta ja erityisesti hänen tapansa arvottaa voittoja (palkkioita) ja tappioita (sanktioita). Perinteisen mallintamisen näkökulmasta sanktiot ovat palkitsemiselle käänteistä toimintaa.⁸ Työstään kansantaloustieteen Nobelin

⁷ Esimerkiksi, jos agentilla on alkuvarallisuutta ”vähän”, optimaalinen rangaistus voi olla suurempi, kuin se omaisuus, jonka agentilta voidaan takavarikoida. Koulun tai opettajan tapauksessa suurin rangaistus voi olla toimiluvan tai viran menettäminen.

⁸ Rangaistusten rikoksia ehkäisevästä vaikutuksesta keskustellaan paljon (esim. Ehrlich, 1975; ja Katz ym., 2005).

muistopalkinnolla vuonna 2002 palkittu Daniel Kahneman on yhdessä edesmenneen kollegansa Amos Tverskyn kanssa jo 70-luvulta lähtien esittänyt useita poikkeamia rationaalisuudesta (mm. Kahneman ja Tversky, 2000). Palkitsemisjärjestelmän suunnittelun kannalta on olennaista agenttien tapa arvottaa rahallisesti samansuuruisia tappioita ja voittoja eri tavoin: tappion tuottama haitta tuntuu isommalta kuin vastaavankokoisen voiton tuottama hyöty. Jos verrataan onnistumisista palkitsevaa ja epäonnistumisesta rankaisevaa käytäntöä, niiden aikaansaamat vaikutukset voivat olla hyvinkin erilaiset. Toisin sanoen, joissakin tilanteissa kannattaa valita pelkkä monitorointi ja rankaiseminen, toisissa taas palkitseminen. Koulumaailmaan sovellettuna tulos tarkoittaa sitä, että onnistuneiden koulujen palkitsemisen sijaan saattaisikin olla kannattavampaa rangaista huonosti menestyneitä kouluja.

Käytännössä palkitsemisjärjestelmiä, miksei sanktioitakin, laadittaessa on pitkään uskottu, että onnistumisesta seuraavan palkkion (tai epäonnistumisesta seuraavan sanktion) täytyy olla riittävän suuri, jotta kannusteilla saavutettaisiin haluttu vaikutus. Jos palkkio ei ole riittävän suuri suhteessa sen eteen nähdystä vaivasta koituvaan haittaan, agentilla ei ole todellista kannustetta lisätä työnte-koa. Viimeaikaiset kokemukset vuosien 2008–2009 pankkikriisin aiheuttaneesta pankkien liiallisesta riskinotosta ja tähän kannustaneista palkitsemisjärjestelmistä ovat kuitenkin asettaneet nämä uskomukset kyseenalaisiksi (esim. Bebchuk ym., 2010a; Bebchuk ym., 2010b). Suuret palkkiot voivat johtaa joko liialliseen riskinottamiseen tai tulosten manipulointiin Holmströmin ja Milgromin (1991) mallin mukaisesti. Tulosten manipulointi ja liiallinen riskin ottaminen johtivat siihen, että liikkeenjohtajat (agentit) eivät enää ajaneet osakkeenomistajien (päämies) etua, vaan omaa etuaan.⁹ Vähitellen kannustejärjestelmiä koskevasta tutkimuksesta on kertymässä evidenssiä siitä, että paikoin rahaa parempia kannusteita voivat olla erilaiset kunniakirjat ja vastaavat ei-rahalliset palkinnot (Ashraf ym., 2012).¹⁰ Myöhemmin luvussa 4.2 esitetään esimerkki epäonnistuneesta kannustekokeilusta Yhdysvalloissa, jossa kannusteiden seurauksena opettajat päätyivät manipuloimaan oppilaidensa koetuloksia.

Myös sanktioiden vaikutukset voivat olla varsin odottamattomia. Eräs esimerkki tästä on Gneezyn ja Rustichinin (2000) israelilaisissa päiväkodeissa toteutettu tutkimus, jossa lapsensa myöhässä hakevia vanhempia alettiin ”sakottaa”. Ns. karkotushypoteesin (deterrence hypothesis) mukaan vanhempien olisi sikon uhalla pitänyt alkaa hakea lapsensa ajoissa, mutta sen sijaan myöhässä tulevien vanhempien määrä lisääntyi merkittävästi. Tämä oli seurausta siitä, että päiväko-

⁹ Koska johdon palkitseminen konkurssiin menneessä Bear Stearnissa ja Lehmanilla perustui edellisen vuoden tulokseen ja palkkiot sidottiin osakkeisiin, järjestelmä kannusti suosimaan nopeita voittoja tuottavaa toimintaa pitkän aikavälin optimoinnin sijaan.

¹⁰ AIDSin riivaamassa Zambiassa tehdyssä kokeessa vapaaehtoiset kampaamoalan työntekijät möivät kondomeja. Ne vapaaehtoiset, jotka palkittiin työstä kunniakirjalla, möivät kondomeja selvästi enemmän kuin ne, jotka saivat osansa kondomikaupan tuotosta. Rahaa tärkeämpiä syitä menestymiselle olivat henkilön sisäinen motivaatio ja pärjääminen sosiaalisessa vertailussa.

tien ja vanhempien välinen sopimus lasten hoidosta jäi vanhemmille epäselväksi. Esimerkiksi vanhemmille ei määritelty kerran, kahdesti tai useammin myöhässä tulemisen kaikkia seurauksia. On mahdollista, että kun tähän epäselvään sopimukseen liitettiin sakko, vanhemmat alkoivat uskoa, että sakko oli myöhästymisen ainoa ja pahin seuraus. Täten he saattoivat testata päiväkodin reaktioita useampiin myöhästymisiin, ja kun sakon lisäksi muita reaktioita ei syntynyt, he lisäsivät myöhästelyään. Kun sakko poistettiin käytöstä, vanhemmat jatkoivat tiheämpää myöhästelyään, sillä heillä ei ollut mitään syytä odottaa sakkoja pahempia seurauksia.

Toinen mahdollisuus on, että ennen sakon käyttöönottoa vanhemmat pitivät päiväkodin hoitajia anteliaina ja valmiina uhraamaan omaa aikaansa ja kärsivällisyyttään, mikäli vanhemmat olivat myöhässä. Sakon käyttöönotto puolestaan saattoi saada vanhemmat ajattelemaan, että hoitajien ajalla on itse asiassa hinta (sakko), ja että he saattoivat ostaa tätä aikaa niin paljon kuin he tarvitsivat. Toisin sanoen, siinä missä ennen sakon käyttöönottoa vanhemmat saattoivat kokea syyllisyyttä myöhästymisestään, sakon maksamiseen ei tätä häpeää liittynyt. Gneezy ja Rustichini (2000) kuitenkin uskovat, että suurempi sakko olisi lopulta vähentänyt myöhästymisiä. Heidän tutkimuksensa osoittaa, että esimerkiksi sakon käyttöönotto voi muuttaa sopimusten osapuolten käsityksiä tilanteesta ja täten saada aikaan odottamattomiakin seurauksia. Siksi on tärkeää tutustua muiden maiden kokemuksiin palkitsemisesta ja rankaisemisesta.

2.2 Turnajaisteorian perusteet

Turnajaisteorian avulla voidaan mallintaa palkitsemisjärjestelmä, jossa päämies palkitsee agenttiaan absoluuttisen tuotoksen sijasta hänen suhteellisen tuotoksensa mukaan. Tällaisessa järjestelmässä agentit kilpailevat toisiaan vastaan, ja ennalta määritellyn palkkion pyrkimyksenä on kannustaa agenteja voittamaan kilpailu. Agenttia ei siis palkita sen perusteella, onko hän hyvä, vaan onko hän *parempi* kuin muut palkkiota tavoittelevat agentit. Esimerkkejä tällaisista turnajaisasetelmista on lukuisia: ylennyksestä kilpailevat työntekijät, opiskelupaikasta yliopistossa kamppailevat ylioppilaat tai olympiamitalista kamppailevat urheilijat.

Lukioissa mahdollisesti sovellettava kannustejärjestelmää, joka voi perustua lukioiden (agenttien) suhteelliseen tuotokseen, voidaan mallintaa turnajaisteorian avulla. Tällöin lukiot tavoittelisivat valtion tai kuntien (päämiehien) määrittämiä palkkioita, joiden lukiokohtainen suuruus määräytyisi sen mukaan, kuinka hyvin kukin lukio pärjäisi suhteessa muihin lukioihin mitattuna oppilaiden oppimistuloksilla tai niiden parantamisella. Tämänkaltaisia järjestelmiä on käytössä erityisesti Yhdysvaltojen eri osavaltioissa, joissa lukiot (high school) asetetaan paremmuusjärjestykseen opintotulostensa tai niiden parantamisen perusteella. Vertailussa parhaiten pärjänneet lukiot saavat muita suuremman palkkion (tarkemmin luvussa 5.1).

Edellisessä luvussa mainittiin, että keskeistä tällaisen palkitsemisjärjestelmän hyödyllisyydelle ja soveltavuudelle on agentin työpanoksen ja tuotoksen monitorointikustannukset. Lazear ja Rosen (1981) toteavat, että kun agentin työpanoksen seuranta on edullista ja luotettavaa tietoa hänen absoluuttisesta tuotoksestaan on saatavilla, palkitsemisen tai palkan on syytä perustua agentin työpanoksen suuruuteen. Tällöin myös agentin kantama riski on pieni, sillä hänellä ei ole epävarmuutta työpanoksensa vaikutuksesta palkkion suuruuteen. Jos seuranta on vaikeaa, ja agentti pystyy muuttamaan työpanostaan ilman, että sitä huomataan, työpanokseen perustuva palkitseminen kannustaa agenttia pinnaamaan. Kun agentin työpanoksen seuraamisesta tulee niin kallista, että agentin moraalinen heikkenemisestä tai moraalikadosta aiheutuu merkittäviä ongelmia, palkitsemisesta agentin suhteellisen tuotoksen mukaan tulee kannattavampaa. Vaikka suhteelliseen tuotokseen perustuvassa järjestelmässä agentti ei voi olla varma saamansa palkkion suuruudesta ja täten hänen kantamansa riski on suurempi kuin työpanokseen perustuvassa palkitsemisessä, suhteelliseen tuotokseen perustuvan palkitsemisen tehokkuuden tuomat hyödyt ovat sen kustannuksia suuremmat. Käytännössä on usein halvempaa monitoroida tuotoksien suhteellista kuin absoluuttista paremmuutta.

Agenttien suhteellisiin tuotoksiin perustuva palkitsemisjärjestelmä on varsin joustava (Nalebuff ja Stiglitz, 1983). Esimerkiksi, jos jokin muutos työympäristössä helpottaa työntekoa, jokaisen agentin on helpompaa tuottaa enemmän. Tästä seuraa, että muutoksen jälkeisessä tilanteessa kaikki agentit tavoittelevat ahkerammin samaa palkkiota, ja palkkio tuotosyksikköä kohden pienenee automaattisesti. Lisäksi kunkin agentin todennäköisyys voittaa riippuu siitä, millainen hänen panoksellaan aikaansaatu tuotos on suhteessa toisten kilpailijoiden tuotokseen (O'Keefe ym., 1984).

Turnajaisasetelmassa sekä päämiehen että agentin kantama riski on pienempi kuin absoluuttiseen tuotokseen perustuvassa palkitsemisessä eli tulospalkkauksessa (O'Keefe ym., 1984). Kun palkkio määritellään etukäteen eikä perustu agentin tuotoksen suuruuteen, päämies tietää tarkalleen palkitsemisen kustannukset. Myös agentti voi olla varmempi saamansa palkkion suuruudesta, joka rajautuu vähimmäis- ja enimmäispalkkion suuruuden välille. Täysin ongelmattomat turnajaiset eivät kuitenkaan ole. Turnajaiset voivat tietyissä tilanteissa saada aikaan vääransuuruisen panostuksen agentilta ja ne voivat houkutellessa vääranlaisia agentteja osallistumaan. Seuraavassa luvussa esitelläänkin tarkemmin turnajaisien teoreettisia lähtökohtia ja tilanteita, joissa turnajaisasetelman mukaista palkitsemista kannattaa soveltaa.

Turnajaisteoriaa selvitetään ensin tilanteessa, jossa kilpailijat ovat kyvykkyydeltään samanlaisia ja neutraaleja riskin suhteen. On kuitenkin mahdollista, että kilpailijat pyrkivät ainakin jonkin verran välttämään riskiä, minkä vuoksi tarkastellaan myös tilannetta, jossa samankaltaiset kilpailijat ovat riskiä kaihtavia. On myös todennäköistä, että kilpailijat eivät ole yhtä kyvykkäitä, joten kol-

manneksi esitellään turnajaisteorian implikaatioita tilanteessa, jossa kilpailijat ovat riskin suhteen neutraaleja, mutta osaamistasoltaan erilaisia. Päämies oletetaan aina riskineutraaliksi. Viimeiseksi esitellään muutamia turnajaisteorian tärkeimpiä näkökohtia, jotka on syytä ottaa huomioon palkitsemisjärjestelmien käytännön suunnittelussa.

2.2.1 Homogeeniset ja riskineutraalit kilpailijat

Taloustieteessä turnajaisia ovat ensimmäisinä teoreettisesti mallintaneet Lazear ja Rosen (1981), Green ja Stokey (1983) sekä Nalebuff ja Stiglitz (1983). Mallien lähtökohtana on päämies-agentti -asetelma, jossa agentti tuottaa päämiehelle jonkin havaittavan tuotoksen, joka riippuu agentin henkilökohtaisesta työpanoksesta, jostakin kaikkiin agenteihin yhteisesti vaikuttavasta satunnaisuudesta eli yhteissatunnaisuudesta (esimerkiksi asetetun työtehtävän vaativuudesta), ja jostakin agentista riippuvasta satunnaisuudesta eli yksilösatunnaisuudesta (esimerkiksi agentin kyvykkyydestä suoriutua työtehtävästä). Agentti pystyy itse määrittämään työpanoksensa, mutta yhteissatunnaisuuteen ja yksilösatunnaisuuteen hän ei voi vaikuttaa. Esimerkiksi lukioden tuotokseen, eli oppimistuloksiin vaikuttava yhteissatunnaisuus voisi tarkoittaa ylioppilaskokeiden vaativuutta, kun taas yksilösatunnaisuudessa voisi olla kyse kunkin lukion opettajien kyvykkyydestä.

Agentti pyrkii maksimoimaan *odotetun* hyötynsä ja valitsee sen mukaisesti työpanoksensa suuruuden, mutta tästä panostuksesta koituu agentille kustannuksia. Päämies pyrkii maksimoimaan odotetun voittonsa, eli agenttinsa (tai agenttiansa) tuotosten odotetun summan ja hänelle (heille) maksettavan korvauksen (palkkion) välisen erotuksen. Mallissa päämies asettaa ensin palkat, minkä jälkeen agentit sitoutuvat omaan työpanokseensa saatuaan jonkin signaalin yhteissatunnaisuudesta. Informaatio on epäsymmetristä kahdella tavalla: Päämies ei havaitse agenttien työpanoksia tai yhteissatunnaisuutta, eikä yksittäinen agentti havaitse toisten agenttien työpanoksia.

Lazear ja Rosenin (1981) yhden päämiehen ja kahden agentin turnajaismallissa, jota tässä ja seuraavissa luvuissa pääasiassa selostetaan, turnajaisten voittaja saa suuremman ennalta määrätyn palkkion kuin häviöjä. Kun agentit ovat riskineutraaleja, he päätyvät valitsemaan samansuuruiset työpanokset, jolloin heistä molemmat voivat voittaa turnajaiset samalla 50 prosentin todennäköisyydellä. Valittu työpanosten suuruus riippuu siitä, kuinka suuri erotus voittajan ja häviäjän palkkioiden välillä on: mitä suurempi palkkioiden välinen ero on, sitä enemmän agentit ovat valmiita panostamaan työntekoon. Kukin agentti haluaa parantaa todennäköisyyttään voittaa, koska voitosta koituva hyöty määräytyy palkkioiden erotuksesta. Palkkioiden absoluuttinen suuruus vaikuttaa vain siihen, osallistuvatko agentit turnajaisiin. Päämies puolestaan haluaisi aina kasvattaa palkkioiden välistä erotusta saadakseen agenteiltaan mahdollisimman suuret työpanokset, mikä kasvattaisi yrityksen voittoa. Suuremmat työpanokset kuitenkin

kasvattavat myös päämiehen kustannuksia, mikä rajoittaa hänen asettamansa palkkioerotuksen suuruutta.

Koska agentit antavat samansuuruiset työpanokset, heidän odotetut tuotoksensa ovat samat. Agenttien *realisoituneet* tuotokset kuitenkin ovat satunnaismuuttujia niihin liittyvästä yhteis- ja yksilösatunnaisuudesta johtuen. Ennalta määrätty palkkiot maksetaan realisoituneiden tuotosten mukaan siten, että se agentti, jonka realisoitunut tuotos on suurempi saa suuremman palkkion.

Verrattuna muihin palkitsemisjärjestelmiin, riskineutraaleille agenteille sekä turnajaiset että agenttien absoluuttiseen tuotokseen perustuva tulospalkkaus saavat aikaan Pareto-optimaalisen resurssien allokaation. Samaan tulokseen päästäisiin myös käyttämällä ns. standardijärjestelmää, jossa agentit tavoittelevat jotakin ennalta määrättyä tuotoksen tasoa. Tällä on samanlaiset kannustevaikutukset kuin toista agenttia vastaan kilpailemisella. Kaikki nämä järjestelmät saavat aikaan samat työpanokset, samat odotetut palkkiot, ja siten samat odotetut hyödyt agentteille.

Se, mikä erottaa järjestelmät käytännössä toisistaan, ovat erilaisiin järjestelmiin liittyvät erisuuruiset tuotosten monitorointikustannukset. Mikäli päämiehelle on halvempaa havaita agentin *suhteellinen* kuin absoluuttinen tuotos, turnajaiset päihittävät järjestelmänä tulospalkkauksen ja standardijärjestelmän. Toisaalta, päämies voi säästää kustannuksissa, mikäli agenttien absoluuttinen tuotos on helposti havaittavissa.

Lukioiden palkitsemisjärjestelmää ajatellen agenttien eli lukioiden absoluuttiseen tuotokseen liittyy epävarmuutta. Ensiksi, on vaikeaa määrittää, kuinka paljon oppilaiden oppimistulokset riippuvat koulusta tai sen opettajista, kuinka paljon taas oppilaiden omasta kyvykkyydestä ja motivaatiosta. Toiseksi, absoluuttisia tulospalkkauksia ei lopullisissa ylioppilaskokeiden arvosanoissa havaita, sillä ne sovitetaan kunakin vuonna samalle jakaumalle. Lisäksi joka vuosi koetta on suorittamassa eri oppilaat, jolloin arvosanat eivät tänäkään puolesta ole vertailukelpoisia vuodesta toiseen. Koska koulujen absoluuttisen tuotoksen paranemista ei voi havaita tai mitata ajassa, turnajaisjärjestelmä toimii lukioissa paremmin kuin tulospalkkaus. Toisaalta, turnajaisjärjestelmässä keskenään samanlaiset ja yhtä suuren panoksen antavat agentit saavat keskenään erisuuruiset palkkiot sen mukaan, miten yksilösatunnaisuus vaikuttaa heidän tuotokseensa. Jos yksilösatunnaisuus mittaa puhtaasti opettajan kyvykkyyttä, siitä palkitseminen on aivan oikein. Jos yksilösatunnaisuus kuitenkin mittaa jotain muuta, vaikka oppilaiden kyvykkyyttä, siitä palkitseminen voi vaikuttaa epäreilulta. Työn luvussa 7.3. hahmotellaan näiden tekijöiden suhdetta.

2.2.2 Homogeeniset ja riskiä kaihtavat kilpailijat

Tilanteessa, jossa agentit ovat riskiä kaihtavia, edellä saadut tulokset palkitsemisjärjestelmien yhdenvertaisuudesta eivät enää päde, ja useimmiten yksi järjestelmä antaa agentille suuremman odotetun hyödyn kuin toinen. Agentit pyrkivät edelleen maksimoimaan odotetun hyötynsä: heidän työpanoksensa ovat samansuuruiset ja heillä on yhtäläiset mahdollisuudet voittaa turnajaiset. Riskiä kaihtavien agenttien työpanos kuitenkin riippuu muun muassa agentin absoluuttisesta riskiaversiovisuudesta: mitä enemmän agentti haluaa kaihtaa riskiä, sitä pienemmän työpanoksen hän valitsee.

Lazear ja Rosen (1981) päättelevät, että agentit, joilla on enemmän varallisuutta ja jotka täten kaihtavat riskiä vähemmän, suosivat todennäköisemmin turnajaisjärjestelmää, kun taas agentit, joilla on vähemmän varallisuutta ja kaihtavat riskiä enemmän, suosivat tulospalkkausta. Mikäli erot agenttien varallisuuksien välillä ovat riittävän suuret, päämiehen voi olla kannattavaa palkita toisia heidän absoluuttisesta, toisia heidän suhteellisesta tuotoksestaan. Lisäksi, tietyillä riskiaversion tasoilla työpanos ja agentin odotetut tulot (palkkio) ovat pienemmät turnajaisjärjestelmässä kuin tulospalkkauksessa. Toisaalta, mikäli yhteissatunnaisuus ja sen varianssi (vaihtelevuus) on suuri, turnajaisjärjestelmä päihittää tulospalkkausjärjestelmän, sillä tulospalkkauksessa satunnaisuus lisää epävarmuutta, josta riskiä kaihtavat agentit eivät pidä. Yhteissatunnaisuudella ei puolestaan ole merkitystä, kun vertaillaan agenttien suhteellisia tuotoksia.

Standardijärjestelmään liittyy vähemmän varianssia kuin turnajaisjärjestelmään, joten riskiä enemmän kaihtavat agentit pitävät standardeista enemmän kuin turnajaisista. Näissäkin vertailuissa tulee kuitenkin muistaa erilaisten järjestelmien monitorointikustannukset. Lisäksi, mikäli yhteissatunnaisuus on suuri, päämiehen on vaikeaa tietää, jäikö jokin standardi saavuttamatta siksi, ettei agentti tehnyt riittävää työpanosta, vai siksi, että jokin yhteissatunnaisuus esti standardin saavuttamisen. Turnajaisjärjestelmässä agentit lisäävät hyötyään kilpailemalla toisiaan vastaan ja siten eliminoivat yhteissatunnaisuuden vaikutukset.

2.2.3 Heterogeeniset ja riskineutraalit kilpailijat

Eräs Lazearin ja Rosenin (1981) tärkeimmistä tuloksista liittyy tilanteeseen, jossa useat riskineutraalit agentit eivät ole yhtä kyvykkäitä hoitamaan heille annettuja työtehtäviä, eli he edustavat eri tyyppisiä (A ja B). Päämies ei tiedä, ketkä agenteista on kyvykkäämpiä, mutta agentit itse tietävät taitonsa, eli kyseessä on informaation epäsymmetrisyyden ongelma. Tyypin A agenttien työpanoksen rajakustannus on pienempi kuin tyypin B agenttien, mikä tarkoittaa, että tyypin A agenteille on helpompaa lisätä työpanostaan kuin tyypin B agenteille. Lazear ja Rosen (1981) osoittavat, että kyvyiltään eritasoiset agentit eivät omatoimisesti valikoidu kilpailemaan omiensa kesken, vaan haluaisivat kaikki kilpailla tyypin

A agenttien muodostamassa ”liigassa”.¹¹ Turnajaiset, joissa eri tyyppisiä edustavat agentit kilpailevat keskenään eivät kuitenkaan ole tehokkaita: riippuen siitä, kummantyyppisiä agenteja on kilpailussa mukana enemmän, agentit tekevät joko liian suuren tai pienen työpanoksen.

Käytännössä tästä ongelmasta päästään eroon ”lajittelemalla” agentit heidän aiempien näyttöjensä perusteella heille sopiviin liigoihin, mikä voi olla paitsi kannattavaa, myös agenttien kannalta tarpeellista. Ensiksi, voidaan olettaa, että kyvykkäimmän agentin on tehokasta toimia vaikeimmassa työtehtävässä. Tällöin päämiehelle voi olla kustannustehokasta järjestää peräkkäisiä ”eliminointiturnajaisia”, joissa lopulta kyvykkäimmät agentit päätyvät kilpailemaan toisiaan vastaan. Toiseksi, agentit eivät välttämättä tiedä tarkkaan omia kykyjään tai kustannusfunktioitaan ja toivovat niihin liittyvää informaatiota ennen kuin valitsevat mielestään sopivan työpanoksen. Päämies voi tässä tapauksessa järjestää agenteille ”karsintoja” tarjotakseen heille informaatiota optimaalisista työpanosstrategioista.¹²

Yllä esitetyt huomiot ovat tärkeitä palkitsemisjärjestelmää valittaessa: kun riskineutraalit agentit eivät ole yhtä kyvykkäitä ja informaatio on epäsymmetristä, *tulospalkkauksella* päästään aina tehokkaaseen ratkaisuun, toisin kuin turnajaisilla. Kuitenkin tapauksissa, joissa agenttien lajittelu on kannattavaa ja joissa agenteille on syytä tarjota informaatiota heidän tyypistään, *turnajaisjärjestelmä* voi hyvinkin johtaa parempaan resurssien allokaatioon. Tästä näkökulmasta esimerkiksi oppilaiden lajittelu lukioihin heidän yläasteen päästötodistustensa perusteella voi olla tehokas ratkaisu. Samalla tavalla lukioiden palkitsemisjärjestelmää suunniteltaessa voisi olla kannattavaa jakaa lukiot niiden oppilaiden oppimistulosten perusteella eri liigoihin. Tähän huomioon palaamme vielä myöhemmin luvussa 7.3.

Lajittelun sijaan tyyppin B agenteille voidaan antaa tasoitusta eri tyyppien välisessä kilpailussa, jos agenttien tyypit tunnetaan (Lazear ja Rosen, 1981). Tyyppin A (tyypin B) agenteille koitua hyötyä tällaisessa kilpailussa on sitä pienempi (suurempi), mitä suurempi tasoitus on. Mutta mikäli tasoituksen suuruus asetetaan ns. kilpailulliselle tasolle, joka kompensoi osan kyvykkyyseroista, kummankaan tyyppin agenteille ei koidu hyötyä tai tappiota kilpailemisesta toisiaan vastaan. Toisaalta, kilpailullisen tason tasoitus ei johda ns. reiluun kilpailuun, sillä se on pienempi kuin tyyppin A ja tyyppin B agentin optimaalisten työpanosten välinen erotus. Tyyppin A agenteilla on kilpailuetu ja he tuottavat suuremman osuuden kokonaistuotoksesta kuin tyyppin B agentit. Esimerkkejä tasoituksen antamisesta

¹¹ Tämä johtuu siitä, että liigassa A agentin odotetut tulot (palkkio), ovat kaikilla työpanosten suuruuksilla suuremmat kuin liigassa B, eli liigassa A kilpaileminen on molemmantyyppisille agenteille aina kannattavampaa kuin liigassa B.

¹² Koulujen pääsykokeet ovat tällaisia karsintoja, jossa myös koulutukseen hakeutuva agentti saa itselleen lisää informaatiota omista kyvyistään ja kustannuksistaan.

ovat käänteinen diskriminaatio esimerkiksi sukupuolen tai rodun mukaan sekä nk. mahdollisuuksien tasa-arvo.

Tasoituksen vaihtoehtona eri tyyppien agentteja voidaan palkita erisuuruisilla palkkioilla, jolloin myös päästään tehokkaaseen ratkaisuun. Palkkio tyyppin A agenteille olisi siis pienempi kuin tyyppin B agenteille, mutta ensiksi mainittujen odotettu tulo (palkkio) olisi kuitenkin suurempi, sillä tyyppin A agenttien todennäköisyys voittaa on suurempi. Käänteinen diskriminaatio, jossa vähemmän kyvykkäille agenteille annetaan tasoitusta tai heitä palkitaan enemmän, jos he sattuvat vastoin odotuksia voittamaan kilpailun, ei välttämättä ole ristiriidassa tehokkaiden kannustemekanismien kanssa.

Heterogeenisten kilpailijoiden turnajaisten tarkastelu on oleellista lukioiden palkitsemissen kannalta. Sekä päämies, eli valtio ja kunnat, että agentit, eli lukiot, olettavat lukioiden edustavan eri tyyppisiä, ainakin mikäli lukion tyyppi määritetään sen oppilaiden menestyksellä ylioppilaskokeissa. Täytyy kuitenkin ottaa huomioon, että lukiot ovat heterogeenisiä myös sen suhteen, millaisia oppilaita ne ottavat opiskelemaan. Tämän vuoksi tärkeäksi muodostuukin palkkioiden jakamisen, eli lukioiden kilpailusijoitusten peruste: siinä missä palkkioiden myöntäminen esimerkiksi ylioppilaskirjoituksissa saatujen arvosanojen perusteella suosisi lukioita, joissa jo lähtökohtaisesti opiskelee paremmin menestyviä oppilaita, palkkioiden myöntäminen arvosanojen parantamisen perusteella tasoittaisi kilpailuasetelmaa ”parempien” ja ”huonompien” lukioiden välillä (tarkemmin luvussa 7.3).

2.2.4 Muita teoreettisia lähtökohtia

Green ja Stokey (1983) jatkavat turnajaisten tarkastelua Lazearin ja Rosenin (1981) viitoittamassa kehikossa ja päätyvät samoihin johtopäätöksiin mallintamalla tilanteen, jossa samanlaisia, riskiä kaihtavia agentteja on useampi kuin kaksi. He painottavat erityisesti, että mikäli yhteissatunnaisuutta ei ole, päämiehen ja yksittäisen agentin välille on aina löydettävissä jokin optimaalinen sopimus, joka päihittää optimaaliset turnajaiset. Tällöin muiden agenttien tuotokset eivät sisällä mitään informaatiota yksittäisen agentin työpanoksesta. Toisaalta, mikäli yhteissatunnaisuutta on ja sen varianssi on riittävän suuri, optimaaliset turnajaiset toimivat paremmin kuin optimaaliset yksilölliset sopimukset. Lisäksi Green ja Stokey (1983) osoittavat, että agenttien lukumäärän kasvaessa riittävän suureksi, optimaaliset turnajaiset päihittävät optimaaliset yksilölliset sopimukset eli saavat aikaan suurimman odotetun hyödyn agenteille. Tällöin turnajaisilla saavutetaan lähes sama lopputulos, kuin täydellisen informaation tilanteessa ja päämies pärjää yhtä hyvin kuin, jos hän havaitsisi yhteissatunnaisuuden suoraan.

Agenttien lukumäärän kasvattaminen muuttaa palkkioita niin, että agentille voi olla houkuttelevampaa jättäytyä kokonaan pois kilpailusta, sillä agentin odotettu hyöty kilpailun ulkopuolella voi olla suurempi kuin odotettu hyöty kilpailuun

osallistumisesta (Nalebuff ja Stiglitz, 1983). Rankaisun palkitsemiselle käänteisestä luonteesta johtuen huonoiten menestyvän agentin rankaiseminen on tuolloin tehokkaampaa kuin parhaiten suoriutuneen agentin palkitseminen. Tällöin kaikki agentit valitsevat optimaalisen työpanoksen sen sijaan, että eivät lainkaan osallistuisi kilpailuun.

Nalebuff ja Stiglitz (1983) huomauttavat, että perinteisillä turnajaisilla voi päästä perusasetelmaa parempaan lopputulokseen, mikäli päämies toteaa paremman agentin paremmaksi ja palkitsee hänet vain, jos tämän agentin tuotos on tietyn, ennalta määrätyn määrän kilpailijan tuotosta suurempi. Tällaisen tuotosvaatimuksen asettaminen voi myös parantaa turnajaisten ajallista kestävyyttä, jonka kannalta on olennaista, että voittaminen määräytyy ennemmin agentin kyvykkyyden kuin satunnaisuuden perusteella. Jos korkeisiin asemiin edenneet mutta vähemmän kyvykkäät agentit tietävät tullessa ylennetyiksi perinteisissä turnajaisissa, he tietävät myös voivansa helposti menettää asemansa, mikäli satunnaisuuteen perustuvan turnajaisjärjestelmän käyttö jatkuu. Jos taas jotkut agentit todella valitsevat suuremman työpanoksen kuin toiset, voittajat määräytyvät heidän ylimääräisen työpanoksensa perusteella, ja turnajaisten järjestäminen tulevaisuudessakin on kannattavaa ja haluttavaa.

Nalebuff ja Stiglitz (1983) tekevät tärkeitä huomioita turnajaisjärjestelmän soveltamisesta käytäntöön. He huomauttavat muun muassa rangaistusten ongelmallisuudesta, sillä rangaistuksista voi tulla itseään ruokkivia: häviäjän motivaatio laskee, hän ei kykene enää kilpailemaan ja siten jatkaa häviämistään. Lisäksi häviäjät ovat usein vähemmän kyvykkäitä ja joutuvat siksi antamaan suuremman työpanoksen kuin keskimääräinen työntekijä voidakseen kilpailla vakavasti. Varman häviäjän läsnäolo myös tuhoaa kaikkien muiden kilpailijoiden kannustimet panostaa työhön. Toisaalta, kun agentteja on paljon, pelkät palkkiot voivat olla tehottomia kannustimia suurempiin työpanoksiin, sillä tällöin agentin odotettu palkkio on liian pieni kannustaakseen pinnaamisen lopettamiseen. Työyhteisössä voi kuitenkin olla mahdollista turvautua myös ns. sosiaalisiin kannustimiin taloudellisten kannustimien sijasta: jos agentit pystyvät monitoroimaan toistensa työpanosta, pinnaajia voidaan rankaista muiden agenttien paheksunnalla, ja agentit valitsevat optimaalisen työpanoksen vaikka agentin odotettu palkkio olisikin pieni. Turnajaisjärjestelmän varjopuoli on, että kilpailutilanteessa agenteilla ei ole minkäänlaisia kannusteita tehdä yhteistyötä. Tulospalkkausjärjestelmässä puolestaan agentteja kannustetaan yhteistyöhön, mikäli se on kaikkien osapuolien kannalta hyödyllistä.

Kaikki tässä luvussa mainitut teoreettiset näkökulmat herättävät kysymyksiä siitä, millainen palkitsemisjärjestelmä olisi lukioille optimaalisin ja saisi aikaan toivotunlaisia tuloksia. Tulisiko palkitsemisjärjestelmä ulottaa kaikkiin lukioihin, ja jos näin tehdään, voisiko sillä olla myös negatiivisia kannustevaikutuksia? Mitä ehtoja palkkion saamiselle tulisi asettaa ja mitkä sen perusteet olisivat? Tulisiko esimerkiksi yhden lukion voittaa toinen tietyllä ”marginaalilla” saadakseen palk-

kion ja tulisiko palkkion suuruus sitoa tämän marginaalin suuruuteen? Millaisia palkkioita muille lukioille kuin voittajalukiolle tulisi tarjota, jotta niiden kannustevaikutukset olisivat optimaaliset? Edelleen, tulisiko huonosta menestyksestä koitua lukioille taloudellisia sanktioita? Vai pitäisikö lukiot jakaa ”liigoihin” oppilaiden taustan mukaan? Lienee selvää, että mikäli lukioiden rahoitukseen halutaan rakentaa toimiva palkitsemiskomponentti, sen suunnittelun on syytä perustua tarkoin analysoituihin teoreettisiin ja empiirisiin päätelmiin ja huomioihin.

3 Palkitseminen käytännössä

Kuten jo edellä on mainittu, turnajaisjärjestelmän mukaisen palkitsemisen yksi etu ovat sen alhaisemmat informaatiokustannukset verrattuna esimerkiksi tulospalkkaukseen tai standardijärjestelmään. Kun palkkioiden suuruudet sovitaan etukäteen, sekä palkkioita jakavat valtio ja kunnat että vastaanottavat lukiot ovat tietoisia odotetuista kustannuksista ja hyödyistä. Sen sijaan, jos palkitsemisjärjestelmä olisi rakennettu niin, että kaikilla lukioilla olisi mahdollisuus saada palkkio jonkin sovitun standardin saavuttamisesta, valtiolle ja kunnille koituvat kustannukset olisivat epävarmat, ja mahdollisesti myös huomattavasti suuremmat.

Turnajaisjärjestelmän käytännön soveltamiseen koulutuksessa liittyy kuitenkin monia haasteita, jotka tulee ottaa huomioon palkitsemisjärjestelmän suunnittelussa ja implementoinnissa. Tässä luvussa selvitetään erityisesti palkittavaan yksikköön ja tuotosten mittaamiseen ja ns. arvonlisämalleihin liittyviä näkökohtia nimenomaan koulutuksen näkökulmasta sekä esitellään vaihtoehtoinen, standardeihin perustuva tapa palkita lukioita. Aluksi täytyy kuitenkin määritellä koulutuksen tuotantofunktio, johon mittaaminen perustuu.

Määritelmän mukaan tuotantofunktio kuvaa suurinta mahdollista tuotantomäärää, joka on saavutettavissa erilaisilla mahdollisilla panoskombinaatioilla tuottamalla tehokkaasti annetulla teknologialla. Koulutuksen tuotoksia ovat oppimistulokset ja panoksiin luetaan annettu opetus, mutta myös perhe ja luokkakaverit tai luokan koko, voidaan ymmärtää panoksiksi. Jos siis halutaan mitata opettajan vaikuttavuutta, täytyy olla mittaustulos oppimistuloksista tai niiden muutoksista. Sen lisäksi opettajan vaikutus pitäisi voida eristää muiden panosten vaikutuksesta, mikä tekee mittaamisesta erityisen haasteellista. Usein opettajan vaikutusta ei voida erottaa esimerkiksi luokkakavereiden vaikutuksesta, koska niistä ei ole saatavilla havaintoaineistoa.

3.1 Palkittava yksikkö

Tähän asti on oletettu, että turnajaisjärjestelmän mukainen palkitseminen tapahtuu koulutasolla ja palkitsemiseen kytkeytyvä vastuu oppimistuloksista on kouluilla. Turnajaisjärjestelmää voi kuitenkin soveltaa myös yksittäisten opettajien palkitsemisessa. Palkitsemisjärjestelmää suunniteltaessa onkin mietittävä, missä määrin halutaan kannustaa yksittäisten opettajien, missä määrin kokonaisten koulujen tuotosten parantamista.

Seuraavaksi käydään läpi palkitsemisjärjestelmän käytöstä esitettyjä hyviä ja huonoja puolia. Aivan kaikki esitetyistä argumenteista eivät sovellu sellaisenaan Suomen lukioihin, vaan ne ovat käytössä hyvin toisenlaisissa koulujärjestelmissä ja yhteiskunnissa. Väärien analogioiden välttämiseksi myös Suomeen sopimattomiin tuloksiin on syytä perehtyä ja arvioida ne.

Opettajien palkitsemista on tutkittu laajasti niin teoreettisista kuin empiirisistä näkökulmista. Kun palkittavaksi yksiköksi valitaan yksittäiset opettajat, kukin opettaja ottaa täyden vastuun omasta työpanoksestaan, ja yhteys saadun palkkion suuruuden ja oman työpanoksen välillä on vahvimillaan. Podgursky ja Springer (2007) toteavat, että on myös todennäköistä, että tämänkaltaisen palkitseminen muokkaa opettajakuntaa: tehokkaimmat opettajat pysyvät viroissaan, ja ei-palkitut opettajat joutuvat joko lisäämään työpanostaan tai jättämään tehtävänsä. Lisäksi sellaiset tulevat opettajat, jotka todennäköisimmin saivat palkkion, myös todennäköisimmin alkavat harjoittaa opettajan ammattia. Toisin sanoen koko opettajakunnan laatu voi parantua erottelevalla rekrytoinnilla ja tehokkaimpien opettajien ammatissa pysymisenä. Myös Lazearin (2000) ja OECD:n (2009) mukaan opettajien palkkion sitominen heidän tuotoksiinsa on tärkeää yleisesti vallitsevien opettajien tehostomien palkkauskäytäntöjen vuoksi. Opettajia palkattaessa rehtoreilla ei ole tarkkaa tietoa palkattavien opettajien tehokkuudesta, mutta tätä informaation puutetta voi paikata opettajien tuotoksia evaluoimalla ja sitomalla tuotokset heidän palkkoihinsa ja ylennyksiinsä. Siksi useissa maissa käytössä oleva virkakäytäntö, jossa opettajien työsopimusten katkaiseminen on tiukasti säädeltyä, on ristiriidassa joustavien palkitsemisjärjestelmien hyötyjen kanssa.

Yksittäisten opettajien palkitsemiseen liittyy kuitenkin tiettyjä ongelmia, joita muun muassa Neal (2009) kartoittaa. Opettajien tuotoksia mittaavien mittareiden kehittäminen on hankalaa. Ensiksi, arviot opettajan vaikutuksesta oppimistuloksiin voivat olla melko epätasaisia, silloinkin kun niissä yritetään ottaa huomioon oppilaiden ja luokan taustaominaisuuksia. Lisäksi, nämä ominaisuudet eivät pysty kattamaan kaikkia oppilaan kyvykkyyteen ja oppimistuloksiin vaikuttavia tekijöitä. Toiseksi, opettajien suhteellista tuotosta arvioivat mittarit usein olettavat, että oppilaat valikoituvat satunnaisesti opettajille, annettuna tavanomaiset oppilaan taustaa kuvaavat tekijät. Joissakin tilanteissa tiettyjen opettajien osoittaminen tietyille oppilaille heijastaa rehtoreiden tekemiä valintoja, jotka perustuvat sellaiseen informaatioon, jota mittareiden laatijoilla ei ole.¹³ Oppilaiden epäsatunnaista valikoitumista opettajille ei siis voida perustellusti sivuuttaa, ja satunnaisuuden oletukseen perustuvat mittarit tuottavat harhaisia arvioita opettajien vaikutuksesta oppimistuloksiin.

On esitetty, että jos kahden tai useamman opettajan toiminta vaikuttaa mitattavaan ja palkittavaan tuotokseen, jotkut opettajat voivat kääntyä tuottavimpia opettajia vastaan, sillä he vaikuttavat siihen standardiin, jolla kaikkia opettajia arvioidaan (Bowles ja Gintis, 2002).¹⁴ Neal (2009), Burgess ym. (2001) ja Mil-

¹³ Esimerkiksi jotkin opettajat ovat erittäin hyviä opettamaan lapsia, joilla on oppimisvaikeuksia tai ongelmia kotona, toiset taas parempia opettamaan luonnostaan ahkeria lapsia.

¹⁴ Katsauksessaan OECD korostaa, että opettajien keskuudessa voi alkaa esiintyä mielipahaa ja kateutta, jos jotkin ei-palkitut opettajat kokevat vaikuttaneensa sellaisten oppilaiden menestykseen, joiden opettaja saa palkkion (OECD, 2009).

grom ja Roberts (1990) toteavatkin, että yksittäisten opettajien turnajaisjärjestelmän mukainen palkitseminen koulujen sisällä voi johtaa opettajien väliseen haitalliseen kilpailuun, mikä vahingoittaa työilmapiiriä ja heikentää koko koulun oppimistuloksia. Pahimmillaan opettajat eivät enää suostu tarjoamaan toisilleen apua tai jakamaan informaatiota, vaan päätyvät sabotoimaan toistensa suorituksia.

Koska opettajien välisen yhteistyön hyödyt voivat olla suuret suhteessa sen kustannuksiin, palkitsemisen tulisi kannustaa yhteistyöhön ja informaation jakamiseen eikä rapauttaa sitä (Neal, 2009). Jos kilpailua käydään koulujen välillä, yksittäisillä opettajilla on selkeä kannuste auttaa kollegojaan, koska kukaan heistä ei saa minkäänlaista palkkiota ellei koulu pärjää paremmin kuin muut koulut. Tällaiset ryhmätason kannustimet voivat siis edistää sosiaalista koheesiota ja reiluuden kokemusta, mikä puolestaan kannustaa tiedon jakamiseen ja kaikkien osapuolten oppimiseen (Pfeffer, 1995; Lazear, 1998; Che ja Yoo, 2001). Kouluja palkittaessa on kuitenkin otettava huomioon, että kaikki opettajat eivät välttämättä viitsi lisätä työpanostaan eivätkä siten kannata asianmukaista vastuuta koko koulun tuotosten parantamisesta. On siis todennäköistä, että kouluja palkittaessa osa resursseista allokoitetaan tehottomasti, sillä pinnaavat opettajat tietävät saavansa palkkion koska muiden opettajien työpanokset kompensoivat heidän laiskottelunsa (OECD, 2009). Muiden opettajien luoma sosiaalinen paine saattaa kuitenkin ratkaista vapaamatkustamisen ongelman: opettajien keskinäinen monitorointi ja sosiaalisten ”rangaistusten”, kuten häpeän tai syyllistämisen, mahdollisuus voi saada yksittäiset opettajat kantamaan osuutensa vastuusta (Kandel ja Lazear, 1992).

Rivkin ym. (2005) huomauttavat, että kokonaisia kouluja palkittaessa on mahdollista, ettei tehokkaimpia opettajia kouluissa, tai koko koulujärjestelmässä, tunnusteta, jos heidän tuotoksiaan ei eroteta tehottomampien opettajien tuotoksista. Jos tehokkaimpien opettajien työpanokset eivät saa tunnustusta, he voivat pahimmillaan jättää koko ammatin kannustimien puutteen vuoksi. Koulujen palkitseminen voi tässä mielessä olla hyvinkin tehotonta: palkitsemisjärjestelmien yleinen tarkoitus on parantaa työvoiman laatua ja tuottavuutta, mutta tämä ei voi tapahtua, mikäli yksittäisten opettajien välisiä tuottavuuseroja ei tunnusteta.

On kuitenkin kolmas, ”hybridi” vaihtoehto, joka yhdistää ryhmä- ja yksilötason palkitsemisen. Nealin (2009) mukaan järjestelmät, joissa palkkiot jaetaan kouluille ja joissa koulut jakavat nämä palkkiot oman henkilöstöpolitiikkansa mukaisesti opettajien kesken, tukevat kilpailua opettajamarkkinoilla: koska koulujen resurssit määräytyvät niiden aikaansaamien oppimistulosten mukaan. Näin jokaisella koululla on kannuste valita sellainen henkilöstöpolitiikka, joka houkuttelee, säilyttää ja palkitsee sellaisia opettajia, jotka eniten myötävaikuttavat koko koulun laadun parantamiseen. Lisäksi, mikäli koulujen rehtoreilla on vapaus jakaa saatu palkkio opettajien kesken omien oppimistulosaggregaattien ja subjektiivisten opettaja-arvioidensa perusteella, rehtorit voivat paitsi palkita henkilökohtai-

sista saavutuksista myös opettajien välisestä yhteistyöstä. Tähän liittyen Jacob ja Lefgren (2008) ovat tutkineet, kuinka hyvin rehtorit ylipäättään erottavat tehokkaammat ja vähemmän tehokkaat opettajat toisistaan. Tutkijat vertaavat rehtoreiden subjektiivisia arvioita koulujensa opettajista opettajien aikaansaamiin oppimistulosparannuksiin ja tulevat siihen tulokseen, että rehtorit tunnistavat hyvin kaikkein tehokkaimmat ja vähiten tehokkaat opettajat, mutta eivät kykene erottamaan opettajia, jotka sijoittuvat tehokkuudellaan jakauman keskivaiheille.

Hybridipalkitsemismallikaan ei ole ongelmaton. Ensiksi, hybridijärjestelmä edellyttää että sekä yksittäisten opettajien että koulujen tuotoksia on teknisesti mahdollista monitoroida, mikä voi aiheuttaa merkittäviä kustannuksia. Päämäärät voivat myös hybridijärjestelmää sovellettaessa menettää selkeyttään, mikä voi heikentää yksittäisten opettajien motivaatiota: paineet saavuttaa lukuisat eri tavoitteet voivat heikentää opettajien kykyä allokoida työpanoksiaan tehokkaasti (Richardson, 1999).

3.2 Tuotosten mittaaminen

Kun turnajaisjärjestelmän soveltamista lukioiden (tai opettajien) palkitsemiseen aletaan suunnitella, tarvitaan metodi, jolla koulut asetetaan paremmuusjärjestykseen niiden tuotosten – usein oppilaiden koetulosten – perusteella sekä palkkioita ja rangaistuksia eri sijoituksen saavuttaville kouluille. McCaffrey ym. (2009) ja Neal (2009) kuitenkin huomauttavat, että on vaikeaa arvioida tarkasti kaikkia koulujen tuottamia tuotoksia. Opettamiseen liittyy suuri joukko tehtäviä, joihin vaikuttavat opetettävien oppilaiden lukumäärä, luokkataso, opetettavat kurssit ja kurssien suhde kokeisiin ja koetuloksiin. Vaikka saatavilla olisikin mittari kaikille tehtäville ja tuotoksille ja mittarit olisivat luotettavia ja ilmaistavissa intervaaliasteikolla, on hyvin epäselvää, kuinka parannuksia näillä eri asteikolla tulisi arvottaa rahallisesti. Koska koulujen (opettajien) tuotokset ovat moniulotteisia, niiden palauttaminen joksikin yksiulotteiseksi mittariksi ei ole mitenkään itsensä selvä prosessi.

Lisäksi oppiminen on useiden ns. yhteistuottajien aikaansaannosta. Tyypillisesti opettamiseen osallistuu useampia opettajia, mutta myös oppilaan omalla tai hänen vanhempiansa panoksella on merkitystä. Lähtökohtaisesti ei ole selvää, miten koulun lisääntynyt panostus opettamiseen vaikuttaa oppilaan tai hänen vanhempiansa panostukseen; se voi joko vähentää tai lisätä sitä. Erityisesti vanhempien panostuksista on vähän tutkimusta, mutta Gelber ja Isen (2011) osoittavat, että ainakin esikoululaisten vanhemmat lisäsivät muun muassa lukemistaan lapselle, kun nämä olivat tulleet valituksi kansalliseen Head Start -ohjelmaan. Jos oppilaat tai heidän vanhempansa muuttavat käytöstään ohjelman myötä, mutta sitä ei mitata tai havaita, tulee opettajan aikaansaama muutos väärin arvioiduksi.

OECD (2009) huomauttaa, että jos palkitsemisen yksiköksi valitaan yksittäiset opettajat, niin näiden opettajien oppilaiden koetulosten monitorointi ja opettajien

palkitseminen niiden parantamisesta kattaa vain osan koulun koko henkilöstön panostuksesta ja ylipäänsä koulun missiosta. Tukeutuminen ainoastaan objektiivisiin tuotosten mittareihin, kuten koetuloksiin, voi myös kannustaa opettajia keskittymään liiallisesti palkittavaan toimintaan muiden tärkeiden koulutuksen tehtävien kustannuksella. Holmströmin ja Milgromin (1991) mallin ennusteen mukaisesti opettajat voivat keskittyä opettamaan vain kokeissa tarvittavia tietoja ja taitoja sen sijaan, että kannustaisivat oppilaitaan kriittiseen ja luovaan ajatteluun. Opettajat voivat myös keskittää suhteettoman paljon huomiotaan sellaisiin oppilaisiin, jotka todennäköisimmin maksimoivat opettajan mahdollisuudet voittaa palkkio. Lisäksi, jos palkitsemisjärjestelmä ei ota huomioon jotakin merkittävää osaa opettajien työnkuvasta tai jos asetetut tavoitteet ovat epäselvät, opettajia on vaikea saada hyväksymään kyseinen järjestelmä ja toimimaan järjestelmän pyrkimysten mukaisesti (Murnane ja Cohen, 1986). Samalla tavalla oppilaiden koetuloksiin voivat vaikuttaa sellaiset tekijät, joihin opettajat eivät voi vaikuttaa. Tällaisia tekijöitä ovat esimerkiksi oppilasotoksien ominaispiirteistä, kuten ahkeruudesta tai lahjakkuudesta, aiheutuva variaatio ja erilaiset satunnaiset koetilanteeseen vaikuttavat tekijät (Kane ja Staiger, 2001). Mitä useampia tällaisia opettajista riippumattomia riskitekijöitä vaikuttaa koetuloksiin, sitä heikommät ovat palkitsemisjärjestelmän kannustevaikutukset.

Kokonaisten koulujen palkitsemisesta Neal (2009) tarjoaa kolme pääperiaatetta, joiden tulisi ohjata koulujen tuotoksia mittaavan järjestelmän rakentamista. Ensiksi, palkitsemisjärjestelmää kuvaavien dokumenttien tulisi selkeästi määrittää, millaisia parannuksia järjestelmän tulisi saada aikaan, ja selvittää päättäjien prioriteetit sen suhteen, missä aineissa ja millaisten oppilaiden keskuudessa näitä parannuksia halutaan saada aikaan. Esimerkiksi lukiolaissa lukioille asetetut tavoitteet ovat hyvin abstrakteja ja lukiodien odotetaan edistävän tuhansien oppilaiden lukuisia eri taitoja. Mikäli selkeitä perusteita palkitsemiselle ei löydetä, koulujen asettaminen paremmuusjärjestykseen objektiivisella tavalla ei tule onnistumaan.

Toiseksi, Nealin (2009) mukaan em. prioriteettien tulisi selkeästi ja täsmällisesti linkittyä niihin toimenpiteisiin, joilla koulujen paremmuusjärjestys saadaan luotua. Palkitsemisjärjestelmät, joiden taustalla ei ole selkeää linkkiä tuotosten määrittelyyn ja niiden mittaamisen välillä, jäävät väistämättä vaille uskottavuutta. Tuotosten mittarit voivat esimerkiksi painottaa parannuksia tietyllä koetulosjakautuksen osalla ja täten arvottaa tiettyjen oppilasryhmien oppimista korkeammalle kuin toisten. Tällaisissa tapauksissa on ensiarvoisen tärkeää, että palkitsemisjärjestelmän suunnitellut taho osaa perustella tämänkaltaiset valintansa. Erityisesti, mikäli järjestelmässä kouluja palkitaan niiden tuotosten lisäksi myös opettajien työpanosten ja/tai opettajien kokemusvuosien perusteella ja näitä komponentteja painotetaan järjestelmässä jollakin tavalla, niin painotukset lähettävät vahvan signaalin oppilaille, heidän vanhemmilleen ja koulutuksen alalla työskenteleville siitä, mitä järjestelmän suunnitellut taho koko koulutusjärjestelmässä arvostaa (OECD, 2009).

Kolmanneksi, Nealin (2009) mukaan palkitsemisjärjestelmien tulisi jakaa koulut liigoihin sen mukaan, millaisia oppilaita ja perheitä ne palvelevat. Tällöin kunkin koulun tuotoksia verrattaisiin vain sellaisten koulujen tuotoksiin, jotka palvelevat oppilaita, joilla on samankaltainen sosioekonominen tausta. Mikäli kouluja ei jaeta tällaisiin liigoihin, on mahdotonta saada luotettavaa tietoa siitä, mikä on kunkin koulun tai opettajan merkitys oppilaiden oppimistulosten parantamisessa, sillä oppilaiden tekemät parannukset heijastavat heidän lähtötasonsa ja opetuksen laadun vuorovaikutusta. Esimerkiksi, kaksi koulua voi kilpailla keskenään mutta opettaa oppilaita, joilla on täysin erilainen tausta. Jos koulu A todetaan paremmaksi kuin koulu B, niin herää kysymys, missä koulu A oikeastaan on parempi kuin koulu B. Perusteluja koulujen jakamisesta eri liigoihin löytyy myös turnajaisteoriasta. Lazearin ja Rosenin (1981) mukaan agenttien lajitteleminen heidän kyvykkyytensä perusteella sopiviin liigoihin on sekä kannattavaa että tarpeellista, sillä turnajaiset, joissa eri tyyppisiä edustavat agentit kilpailevat keskenään, eivät ole tehokkaita.

On toki mahdollista, että liigoihin jakamisen seurauksena huonompiosaisten yhteisöjen tai kuntien koulujen tuotokset eivät parane, sillä näissä yhteisöissä toisten koulujen voittaminen onnistuu ilman, että tuotokset olisivat kovinkaan korkeita. Neal (2009) kuitenkin huomauttaa, että jos on olemassa tällaisia ns. helppoja liigoja, joissa on ”halvempaa” voittaa palkkio tai välttää rangaistus, niin lahjakkailta rehtoreilla ja opettajilla on kannustimet siirtyä kouluihin, jotka kilpailevat tällaisissa liigoissa. Ennen kaikkea sellaisilla rehtoreilla ja opettajilla, jotka parhaiten soveltuvat opettamaan tällaisissa kouluissa, on suurimmat kannustimet siirtyä näihin kouluihin. Määrittämällä erisuuruiset palkkiot eri liigoille voidaan oikeat rehtorit ja opettajat houkuttaa huonompiaisisiin yhteisöihin. Ekstrapalkkioon reagoivat eivät ole vain niitä, jotka suostuvat opettamaan huonompiaisissa yhteisöissä, vaan myös niitä, jotka luottavat siihen, että he menestyvät tässä tehtävässä. Toisaalta vastoin Nealin väitettä Ladd ja Walsh (2002) toteavat, että palkitsemisjärjestelmät voivat myös luoda vääränlaisia kannusteita koulujen opettajille: esimerkiksi Pohjois- ja Etelä-Carolinassa palkitsemisen perusteina käytetyt arvonlisämitat kannustivat opettajia ja hallintohenkilökuntaa (entisestään) välttämään kouluja, joiden oppilaat pärjäsivät keskimääräistä huonommin. Kannustimilla saatiin aikaan lopputulos, joka oli vastoin odotuksia. On siis mahdollista, että koulun huonous johtuu siinä määrin tekijöistä, joihin opettajat eivät voi työllään vaikuttaa, että he eivät ennusteen mukaan hakeudu ”helppo- ja voittoja” tuottaviin kouluihin.

3.3 Arvonlisämallit

Oli kyseessä sitten opettajien tai koulujen palkitseminen, niin ns. arvonlisämallit ovat tulleet yhä suosittummiksi palkitsemisen perusteena. Arvonlisämallit määritellään malleiksi, jotka analysoivat oppimistulosten parantumista tai oppilaiden tiedon lisääntymistä tietyllä aikavälillä (ks. esim. Amrein-Beardsley, 2008; Raudenbush, 2004; Hanushek, 2003; Ladd ja Walsh, 2002). Muista malleista arvon-

lisämallit erottaa se, että arvonlisämalleihin sisältyy jokin mittaustulos oppilaan aiemmasta koulumenestyksestä, mikä mahdollistaa arvion koulun vaikutuksesta oppilaan oppimistulosten kehitykseen. Arvonlisämallit ottavat hyvin huomioon myös ne ei-havaitut tekijät, jotka vaikuttavat oppilaan lähtökohtaiseen koulumenestykseen, sillä ne sisältyvät aiempaan mittaustulokseen (esim. oppilaan kyvykyys). Usein malleihin sisällytetään myös oppilaiden sosioekonomisia ominaisuuksia kuvaavia tekijöitä. Tällaisia malleja OECD (2008) nimittää kontekstuaalisiksi arvonlisämalleiksi. Näiden tekijöiden tarkoituksena on ottaa huomioon koulujen erilaiset opetukselliset haasteet, jotka johtuvat oppilaiden erilaisista taustoista. Käytännössä tämä vastaa Nealin (2009) peräänkuuluttamaa koulujen jakamista eri liigoihin niiden palvelemien oppilaiden ominaisuuksien perusteella.

Amrein-Beardsleyn (2008) mukaan arvonlisämallien suosio perustuu muun muassa siihen, että opettajien tai koulujen tehokkuuden tutkiminen niiden aikaansaamien oppimistulosparannusten perusteella on perustellumpaa kuin luottaminen perinteiseen oppimistulosten *tasoon* tietyllä hetkellä. Arvonlisämalleja käyttämällä opettajia ja kouluja ei palkita tai rangaista väärin perustein kerta-luontoisesti erittäin hyvin tai erittäin huonosti pärjäävistä oppilasryhmistä, vaan perustuen siihen, kuinka paljon lisäarvoa ne ovat tuoneet oppilaidensa oppimiseen. OECD (2008) puolestaan linjaa, että arvonlisämallit voivat parhaimmillaan tuottaa informaatiota opetuksen jatkuvan kehityksen tueksi, sillä niiden avulla voidaan tunnistaa, mitkä alueet kouluissa ja koulutusjärjestelmissä tuottavat eniten lisäarvoa ja millä alueilla tarvitaan kehitystä. Mallit voivat myös auttaa luomaan sellaisia koulutusrahoitusjärjestelmiä, jotka allokoivat resursseja tehokkaammin niille alueille, jotka resursseja eniten kaipaavat. Lisäksi arvonlisämallien avulla voidaan tasoittaa sosioekonomisia epätasa-arvoisuuksia, joita kouluissa voi umpimähkäisten ja epätarkkojen tulostittareiden alla piillä.

Vaikka arvonlisämalleja pidetään tärkeinä työkaluina koulujärjestelmien hyvien toimintatapojen tunnistamiseksi, mallien käytännön suunnittelu ja soveltaminen on haasteellista (OECD, 2008). Keskeisin vaikeus liittyy mittaamiseen: arvonlisämallisissa käytettyjen koetulosten tulee mitata oikeita asioita ja tulosten tilastoinnin tulee olla laadukasta. Koetulosten pitäisi esimerkiksi riittävästi heijastaa sitä, mitä oppilaat todella osaavat. Lisäksi, arvonlisämallit eroavat toisistaan siinä, missä määrin niissä otetaan huomioon oppilaiden sosioekonomisia ominaisuuksia ja mitä ominaisuuksia malleihin sisällytetään. Näiden ominaisuuksien mittaamiseen sisältyy jonkin verran virhettä, mikä voi aiheuttaa harhaa estimointituloksiin etenkin silloin, kun koulun käytännöt vaikuttavat näihin ominaisuuksiin. Tällaisia ominaisuuksia ovat esimerkiksi oppilaan suhtautuminen kouluun ja keskimääräinen viikoittaisten kotitehtävien määrä. Lisäksi Todd ja Wolpinin (2003) mukaan mallit perustuvat usein monille vahvoille oletuksille oppilaiden oppimisen luonteesta, joten jos jokin näistä oletuksista ei pidä paikkaansa, koulu-

jen vaikutuksista oppimiseen voidaan vetää hyvinkin virheellisiä johtopäätöksiä.¹⁵

Myös oppilaiden sosioekonomisten ominaisuuksien huomioon ottamisesta arvonlisämalleissa on kiistelty, vaikka OECD (2007) onkin PISA-tutkimuksissaan havainnut, että oppilaiden vanhempien sosioekonominen asema on yksi vahvimmista koulujen menestystä ennustavista tekijöistä. Mutta koska arvonlisämalleihin sisältyy jo mittaustulos oppilaan aiemmasta koulumenestyksestä, sosioekonomisten ominaisuuksien sisällyttäminen malleihin ei välttämättä lisää mallien kykyä selittää oppilaiden erilaisia oppimistuloksia (Ballou ym., 2004; Hægeland ja Kirkebøen, 2008). Ballou ym. (2004) tähdentävät, että arvonlisämallit mittaavat oppimistulosten paranemista oppilaan omasta lähtötasosta, eli ne implisiittisesti kontrolloivat niitä vaikutuksia, joita sosioekonomisilla tai muilla taustatekijöillä voi olla oppimistuloksiin. Toisaalta McCaffrey ym. (2003) huomauttavat, että mikäli tietoa oppilaan aiemmasta koulumenestyksestä ei ole saatavilla, edes oppilaan sosioekonomista taustaa kuvaavien muuttujien lisääminen malliin ei riitä täysin poistamaan taustaominaisuuksien vaikutuksia. Erityisesti Mizala ym. (2007) osoittavat, että ainakaan Chilen kaltaisessa järjestelmässä pelkkä oppilaan taustan kontrolloiminen ei riitä, ja vuosittainen satunnainen vaihtelu tuo liikaa satunnaisuutta koulujen paremmuusjärjestykseen, jotta mallin tuloksia voitaisiin käyttää luotettavasti palkitsemisen perusteina.

On tärkeää ymmärtää, mitä sosioekonomista taustaa kuvaavien muuttujien lisäämisellä nimenomaan arvonlisämalleihin saadaan selville ja ei saada selville. Ballou ym. (2004) huomauttavat, että oppilaat eivät välttämättä valikoidu satunnaisesti kouluihin: mikäli paremmat opettajat saavat viran kouluissa, jotka palvelevat taustaltaan vaurasta oppilaspopulaatiota, tai mikäli vauraat vanhemmat pyrkivät saamaan lapsensa parhaisiin kouluihin, sosioekonomisista muuttujista tulee opettajien ja koulujen laatua kuvaavia muuttujia (proxy-muuttujia). Koska nämä muuttujat ovat korreloituneita ei-mitattavien koulun ja opettajien laadun kanssa, niiden kertoimiin sisältyy osittain se, mitä estimointien virhetermeillä pyritään mittaamaan. Täten, kun huomataan, että vaikeammista oloista tulevat oppilaat eivät paranna oppimistuloksiaan yhtä paljon kuin vauraampien vanhempien oppilaat, on vaikeaa sanoa, johtuuko se yksinomaan heidän taustoistaan vai heidän koulutuksensa laadusta. Sosioekonomisten muuttujien lisääminen arvonlisämalleihin ei paljasta, kumman vaikutuksista on kyse.

Tekwe ym. (2004) vertaavat erilaisia koko koulua koskevia arvonlisämalleja käytännössä. Käyty keskustelu kulminoituu sosiodemografisten taustakontrollien käyttöön ja rooliin eri malleissa. Mallin valinta riippuu siitä, pitäisikö koulut asettaa vastuuseen sosiodemografisten tekijöiden merkittävistä vaikutuksista. Jos nämä tekijät lisätään malliin, koulut käytännössä vapautetaan vastuusta niiden vaikutuksista. Niin sanotuissa high stakes -järjestelmissä, joissa opettajien palkat

¹⁵ Esimerkiksi opettajan vaikutus oppimiseen voi tulla esiin viiveellä, jolloin se ei näy mittaustuloksessa.

ja koulujen budjetit riippuvat oppimistulosten parantamisesta, sosiodemografisten tekijöiden huomiotta jättäminen voi kannustaa opettajia ja hallintohenkilökuntaa jättämään koulut, joissa on paljon köyhiä tai vähemmistöihin kuuluvia oppilaita. Toisaalta näiden tekijöiden huomioon ottaminen voi rajoittaa heikoin päijävien koulujen mahdollisuuksia saavuttaa täyttä potentiaaliaan, kun niiltä jo lähtökohtaisesti odotetaan vähemmän kuin paremmin pärjäviltä kouluilta.

Suomen järjestelmän kannalta oleellinen kriittinen havainto on se, että tietoaineiston keräämiseen ja sen muokkaamiseen käytettävään muotoon voi liittyä merkittäviäkin kustannuksia.¹⁶ Koska Suomessa ei ole peruskoulujen valtakunnallisia päättökokeita ja opettajien arvostelussa on huomattava subjektiivinen osuus, arvonlisämallissa oppilaan lähtötason mittari on epätarkka. Jos peruskoulun päättötodistuksen arvosanojen ”virheet” ovat jokseenkin satunnaisesti jakautuneita, ongelma ei ole suuri. Sen sijaan, jos joidenkin koulujen oppilaiden arvosanat ovat järjestelmällisesti parempia (heikompia) kuin niiden pitäisi olla, lukion tuottama arvonlisäys tulee aliarvioiduksi (yliarvioiduksi). Kustannuksia voi aiheutua myös uusien tulostittareiden esittelemisestä ja käyttöön otosta koulutusjärjestelmässä.

3.4 Standardijärjestelmän mukainen palkitseminen

Standardijärjestelmä tarjoaa vaihtoehdon turnajaisjärjestelmälle. OECD:n (2009) mukaan standardijärjestelmässä määritetään jokin tuotoksen eli oppimistulosten taso tai standardi, joka koulujen tulee saavuttaa saadakseen palkkio. Näitä standardeja voidaan määrittää yksi tai useampia, ja palkkio voi kasvaa myös lineaarisesti tietyn tason saavuttamisen jälkeen. Tämä eroaa ratkaisevasti turnajaisjärjestelmästä, sillä standardin saavuttava koulu saa palkkion huolimatta muiden koulujen tuotoksista. Standardijärjestelmän mukaisen palkitsemisen edut näkyvät ennen kaikkea silloin, kun järjestelmää sovelletaan yksittäisten opettajien palkitsemiseen, sillä toisin kuin turnajaisjärjestelmällä, standardijärjestelmällä vältetään haitallinen opettajien välinen kilpailu.

Standardijärjestelmän etu on, että se asettaa kouluille (opettajille) selkeät tavoitteet, mikä puolestaan auttaa suuntaamaan koulujen huomiota ja toimintaa kohti tärkeäksi arvioitua tehtävää. (Locke, 1968; Locke ym., 1990; Locke ym., 1981). OECD (2009) huomauttaakin, että verrattuna turnajaisjärjestelmään, standardijärjestelmä voi saada paremman vastaanoton kouluissa, jotka voivat pitää suhteellisen tuotostavoitteen saavuttamista mahdottomana. Koulut voivat myös pitää turnajaisjärjestelmän mukaista tuotoksen mittaamis-/arviointijärjestelmää epäoikeudenmukaisena, koska koulut eivät saa etukäteen tietää standardia, joka niiden tulisi saavuttaa palkkion saamiseksi. Standardijärjestelmässä palkitsemiseen liittyvä epävarmuus on turnajaisjärjestelmää vähäisempää.

¹⁶ Tarkemmin Pursiainen, Pääkkönen ja Seppälä (2011).

Standardijärjestelmän soveltamisen ongelmana on kuitenkin sen kustannuksiin liittyvä epävarmuus, sillä mikäli ei etukäteen määritellä kiintiötä palkkioita saavien koulujen lukumäärälle, on mahdotonta etukäteen arvioida, voiko järjestelmää soveltaa ja ylläpitää kestäväällä pohjalla annetulla rahoituksella. Mikäli koulut (opettajat) kyseenalaistavat rahoituksen kestävyys, ne eivät enää yhtä todennäköisesti kasvata työpanostaan saadakseen palkkion. Toisaalta opettajilla ja kouluilla on kannuste lobata päätöksentekijöitä standardin alentamisen puolesta ja ylipäänsä vaikuttaa järjestelmän sääntöihin itselleen suotuisasti, mikä kasvattaisi palkkion saavien koulujen (opettajien) lukumäärää ja siten palkitsevan tahon kantamaa rahoituksellista riskiä (Heneman ym., 2007). Siksi, kun standardia asetetaan, täytyy varmistaa, ettei valittu standardi suosi tiettyjä kouluja (opettajia).

4 Empiirinen tutkimus koulutuksen taloustieteessä

Koulutuksen vaikuttavuuden ja tehokkuuden mallintaminen rakentuu tuotantofunktioon (Hanushek, 2010). Tavallisimpia tuotoksia ovat oppilaiden menestyminen opinnoissa ja läpäisyasteet. Erilaisia panoksia sen sijaan on useampia, kuten koulujen (taloudelliset) resurssit, opettajan laatu, oppilaan sosioekonominen tausta, luokkakoko ja luokkakaverit (Hanushek, 2002). Koska osa resurssimuuttujista on sellaisia, että poliittiset päätöksentekijät voivat vaikuttaa niihin, on taloustieteellistä tutkimusta käytetty runsaasti päätöksenteon taustalla. Poliitiikasta johtuva ”tarveharkinta” voi ikävä kyllä tuottaa tutkimustuloksiin harhaa; jos toimenpiteet kohdistetaan tiettyyn ryhmään, joka vielä mahdollisesti itse valikoituu toimenpiteen kohteeksi, voivat politiikan vaikutukset näyttäytyä todellista suuremmilta.

Osa hyvissäkin jurnaaleissa julkaistusta empiirisestä tutkimuksesta on ollut laadultaan heikkoa, useista lähinnä menetelmistä ja/tai aineiston puutteista johtuvista syistä. Hanushekin (2003, 2010) mukaan hyvin tyypillinen virhe on ollut sivuuttaa koulutuksen kumulatiivinen luonne: koska aiemmat panostukset vaikuttavat myöhempien vuosien tuloksiin, täytyy niiden vaikutus kontrolloida esimerkiksi mittaamalla arvonlisäystä. Myös jatkuvasti kehittyvät menetelmät ja rekisteriaineistojen yleistyvä käyttö ovat mahdollistaneet koulutuksen tuotantofunktion aikaisempaa paremman mittaamisen ja tutkimuksen laatu on parantunut viime vuosikymmeninä.

Edellä mainituista puutteista johtuen tässä luvussa keskitytään tutkimuksiin, joiden menetelmällinen laatu on nykytiedon valossa korkea. Lisäksi on keskitytty tutkimuksiin, joissa arvioinnin kohteena oleva järjestelmä on riittävän samankaltainen Suomen koulutusjärjestelmän kanssa tai joissa järjestelmien välisten erojen ei pitäisi vaikuttaa tulosten yleistettävyyteen. Tällainen järjestelmä on esimerkiksi Israelissa, jossa lukioiden oppilaat suorittavat opintojensa päätteeksi Suomen ylioppilaskoetta vastaavan päättökokeen, Bagrutin.

4.1 Eksperimenttien tuloksia Israelista

Lavyn (2002) tutkimus Israelin lukioista tarjoaa hyvän esimerkin turnajaisjärjestelmän käytöstä koulujen palkitsemisessa. Lavy vertailee kahta Israelin lukioissa 1990-luvun puolessavälissä tehtyä koetta: toisessa kouluja palkittiin sen mukaan, kuinka paljon ne paransivat oppimistuloksiaan suhteessa muihin kouluihin (turnajaiskokeen). Toisessa koulut saivat oppimistuloksista riippumattomia lisäresursseja opetusajan lisäämiseksi, opetusryhmien pienentämiseksi ja heikoimpien oppilaiden avustamiseksi (resurssiekperimentti). Kummankin kokeen tavoitteena oli vähentää pudokkaiden määrää ja parantaa oppilaiden tuloksia ylioppilaskokeissa.

Lavy (2002) tutki näiden kahden kokeilun vaikutuksia neljänlaisiin oppimistuloksiin: käytyjen kurssien keskimääräiseen vaatavuustasoon, ylioppilaskokeissa saatuihin keskimääräisiin pisteisiin, ylioppilastutkinnon suorittaneiden osuuteen sekä pudokkaiden osuuteen. Turnajaiskokeessa osallistuvien koulujen oppimistuloksia seurattiin kahdessa vaiheessa. Ensimmäisessä vaiheessa koulujen keskimääräiset tulokset (kolmella ensimmäisellä oppimistulosmittarilla) normalisoitiin suhteessa laskennalliseen odotettuun perustasoon, joka otti huomioon oppilaiden sosioekonomisen taustan. Toisessa vaiheessa koulut asetettiin paremmuusjärjestykseen sen mukaan, kuinka ne onnistuivat parantamaan tuloksiaan em. mittareilla. Vain paras kolmannes kouluista sai palkkion ja koulukohtainen palkkio määräytyi vain koulun sijoituksen perusteella.

Kouluissa 75 prosenttia palkkiosta jaettiin opettajien kesken bruttotulojen suhteessa ja 25 prosenttia käytettiin parantamaan muun muassa henkilöstön yhteisiä tiloja. Kyseessä oli siis kokonaista ryhmää palkitseva ohjelma, jossa ryhmän yhteinen tuotos määrittää lopullisen kannustepalkkion. Palkkio puolestaan jaettiin ryhmän jäsenten kesken riippumatta heidän henkilökohtaisesta työpanoksestaan ja tuotoksestaan. Kuten palkittavaa yksikköä käsittelevässä luvussa 3 kävi ilmi, niin verrattaessa ryhmien kannustamista yksilöiden kannustamiseen, esille nousee kysymyksiä erityisesti yksilöiden työpanoksen mittaamisesta, vapaamatkustamisesta, useista keskenään kilpailevista työtehtävistä sekä moraalikadosta.

Lavyn (2002) tulosten mukaan turnajaiskokeella oli positiivinen ja tilastollisesti merkitsevä vaikutus suoritettujen kurssien vaatavuustasoon sekä ylioppilaskokeissa saatuihin pisteisiin. Vaikutus oli positiivinen, mutta ei tilastollisesti merkitsevä ylioppilastutkinnon suorittaneiden osuuteen. Kokeella oli vaikutusta erityisesti heikompien oppilaiden oppimistuloksiin ja sillä oli merkitsevä vaikutus sosioekonomiselta taustaltaan muita heikompien oppilaiden ylioppilaskokeen suorittaneiden osuuteen. Kokeella oli puolestaan negatiivinen ja tilastollisesti merkitsevä pienentävä vaikutus pudokkaiden osuuteen. Tässä tapauksessa turnajaisilla oli siis halutunlaiset vaikutukset. Toisaalta myös resurssikokeella oli samankaltaisia myönteisiä vaikutuksia suoritettujen kurssien vaatavuustasoon ja ylioppilaskokeissa saatuihin pisteisiin, mutta ei juuri vaikutusta pudokkaiden osuuteen. Lisäksi, toisin kuin turnajaiskokeessa, resurssikokeen vaikutukset näkyivät erityisesti parempien oppilaiden tuloksissa. Kun valitaan vaikuttamisen tapaa, nämä erot tuloksissa ovat oleellisia: turnajaiskoe vaikutti heikompien oppilaiden oppimistuloksiin ja resurssikoe puolestaan parempiin oppilaisiin.

Lavy (2002) vertaa turnajaiskoetta Lazearin ja Rosenin (1981) teoreettisiin hypoteeseihin, joiden mukaan symmetrisen informaation, riskineutraaliuden ja yhtä kyvykkäiden kilpailijoiden tapauksessa kaikkien turnajaisisten kilpailijoiden pitäisi antaa samansuuruinen työpanos, ja palkkion saamisen todennäköisyyden pitäisi olla yhtä suuri kuin palkkion saavien koulujen osuus. Lavyn (2002) turnajaiskokeessa palkittavien koulujen osuus oli yksi kolmasosa. Kaiken kaikkiaan palkkiota jaettiin kolmena vuonna (1995–1997), mutta ensimmäisenä vuonna niiden

tarkoituksena oli vain luoda kokeelle julkisuutta ja luottamusta. Palkkion saamisen todennäköisyys kahtena perättäisenä vuonna, jolloin kokeella olisi voinut olla kausaalisia vaikutuksia (1996–1997), olisi pitänyt siis olla 0,11 ($0,33 \cdot 0,33$). Todellisuudessa kuitenkin 53:sta turnajaisiin osallistuneesta koulusta 14, eli 28 prosenttia, sai palkkion kahdesti, mikä osoittaa, että turnajaisten voittamiseen liittyi jotakin systeemistä ja epäsatunnaista. Lavy (2002) ei kuitenkaan havaitse mitään eroja voittaja- ja häviäjäkoulujen koetta edeltävien ominaisuuksien, kuten oppilaiden koulumenestysten tai sosioekonomisten taustojen tai opettajien ja oppilaiden lukumäärien välillä.

Lavy (2002) arvioi kahden kokeensa kustannustehokkuutta. Vaikka turnajaiskokeen kokonaiskustannukset olivat suuremmat kuin resurssikokeen, koulukohtaiset kustannukset olivat jälkimmäisessä lähes kaksinkertaiset turnajaiskokeeseen verrattuna, sillä turnajaiskoe vaikutti lähes kolminkertaiseen määrään kouluja resurssikokeeseen verrattuna. Jotta eksperimenteistä toinen voitaisiin todeta tehokkaammaksi, koulukohtaisten kustannusten välistä eroa tulisi verrata kokeiden vaikutusten väliseen eroon. Vaikka Lavyn (2002) tulokset osoittavat, että resurssikokeella oli keskimäärin 50–70 prosenttia suuremmat vaikutukset kolmenlaisiin oppimistuloksiin (suoritettuihin opintopisteisiin, keskimääräisiin pisteisiin ylioppilaskokeissa ja ylioppilaskokeisiin osallistuneiden oppilaiden osuuteen) kuin turnajaiskokeella, sillä oli pienempi vaikutus kolmeen muunlaiseen tuotokseen (tiedeoppiaineissa suoritettuihin opintopisteisiin, pudokkaiden osuuteen ja ylioppilastodistuksen saaneiden oppilaiden osuuteen) ja se oli lähes kaksi kertaa kalliimpi kuin turnajaiskoe. Täten turnajaiseksperimentti osoittautui kustannustehokkaammaksi kuin resurssieksperimentti.¹⁷

Opettajien palkitsemisen vaikutusta oppimistuloksiin on tutkittu sangen laajasti erityisesti Yhdysvalloissa¹⁸, mutta Suomen lukioihin läheisimmin rinnastettavat tulokset on saatu edellä mainitun tutkimuksen tavoin Israelista, jossa toteutettiin vuonna 2001 koe, jossa aineopettajia palkittiin heidän tuloksistaan suhteessa muihin saman aineen opettajiin (Lavy 2009). Tutkimuksessa arvioitiin opettajan palkitsemisen vaikutusta matematiikan ja englannin oppimistuloksiin kouluissa, jotka aiemmin olivat menestyneet kokeissa kansallista keskiarvoa heikommin. Tutkijalla oli käytössään laaja joukko koulun ja oppilaan taustoja kuvaavia muuttujia, joiden avulla ennustettiin oppilaiden menestymistä kokeessa.

¹⁷ Lavy (2002) myös huomauttaa, että turnajaiseksperimentti alkoi vuotta myöhemmin kuin resurssiohjelma ja sen vaikutukset olivat hyvin rajalliset ensimmäisenä soveltamisvuonna. Lisäksi resurssieksperimentti kasvatti merkittävästi koulujen opetusresursseja – keskimäärin yli kahdella opettajalla koulua kohden – kun taas opettajille jaetut bonukset olivat varsin pieniä suhteessa heidän vuotuisiin tuloihinsa: suurin bonus, 715 dollaria, oli vain noin 2,5 prosenttia opettajien keskimääräisistä vuotuisista tuloista. Turnajaiseksperimentin saaman julkisuuden ja voittajakoulujen parantuneen maineen vaikutuksia ei kuitenkaan sovi aliarvioida tässä yhteydessä.

¹⁸ Jacob ja Ludwig (2009) käyvät lyhyesti läpi aiheeseen liittyvää kirjallisuutta. Koska suuri osa tutkimuksista on tehty peruskouluikäisillä ja hyvin erilaisessa institutionaalisessa ympäristössä, eivät tuloksia välttämättä voi yleistää Suomen lukioihin.

Opettajat palkittiin kahden mittarin perusteella: kuinka suuri osuus oppilaista läpäisi kokeen ja mikä oli heidän koearvosanojensa keskiarvo. Vain ne opettajat arvioitiin, joiden oppilaat pärjäsivät molemmilla mittareilla mitaten ennustettua paremmin. Parhaan opettajan saama palkinto oli huomattava, jopa puolet opettajien vuoden keskiansiosta. Huomattava palkkio takasi sen, että opettajilla oli aito kannuste panostaa opetukseen.

Opettajien panostukset todella kannattivat. Suhteessa kontrolliryhmään koeryhmän oppilaat osallistuivat useammin kokeeseen, suorittivat sen hyväksytysti ja heidän koekeskiarvonsa oli selvästi kontrolliryhmän oppilaita korkeampi. Muutoksia oli havaittavissa molempien tutkittujen aineiden tuloksissa, mutta ne olivat suurempia matematiikassa. Suurimmat vaikutukset havaittiin lähtötasoltaan heikommilla oppilailla: matematiikassa kolmessa ensimmäisessä kvartiilissa ja englannissa kahdessa ensimmäisessä kvartiilissa. Heikoimmassa ryhmässä (1.kvartiili) vaikutus syntyi suuremman osallistumisasteen seurauksena, kun taas toisessa ja kolmannessa kvartiilissa efekti syntyi läpäisyasteen ja keskiarvosanan noususta. Toisaalta, jos verrataan heikointa kvartiilia kontrolliryhmään, kokeeseen osallistuminen nousi matematiikassa (englannissa) 32 (24) prosenttia, läpäisyaste 57 (23) prosenttia ja koekeskiarvo 46 (17) prosenttia. Erityisesti heikoimmat oppilaat hyötyivät saamastaan huomiosta, ja opettajille tehty kysely vahvisti tuloksen: opettajat muuttivat opetuskäytäntöjään. He muun muassa kasvattivat opetukseen käytettyä aikaa, lisäsivät pienryhmäopetusta ja sopeuttivat opetusmetodejaan oppilaiden tasoon. Aiemmissä tutkimuksissa havaittua näyttöä opettajien harrastamasta huijaamisesta ei sen sijaan löydetty (Jacob ja Levitt, 2003).

Israelin kokeilla saadut tulokset osoittavat, että opettajien kannusteilla on todella merkitystä. Erityisesti heikoimmat oppilaat hyötyvät siitä, että opettajat panostavat enemmän opettamiseen. Tällöin koulujärjestelmässä on oltava tehottomuutta, jossa oppilaiden potentiaalia hukataan järjestelmän toimesta. Tutkimus ei kuitenkaan kykene vastaamaan siihen, mitkä ovat järjestelmän pitkän aikavälin vaikutukset. Ensinnäkin, olisiko tulokset pystytty toistamaan, jos koe olisi jatkunut useamman vuoden ajan? Toisaalta, mikä vaikutus parantuneilla koetuloksilla on oppilaiden osaamiseen tulevaisuudessa: Muuntuvatko paremmat oppimistulokset pysyväksi osaamiseksi, vai kuvaavatko ne kykyä vastata kokeeseen aikaisempaa paremmin?¹⁹

Myös oppilaille suunnattujen kannusteiden vaikutuksia on tutkittu jonkin verran.²⁰ Oppilaan palkitsemisen voidaan arvioida kannustavan opiskelemaan useis-

¹⁹ Esikouluissa positiivisten vaikutusten pysyvyydestä sen sijaan on jonkin verran näyttöä, mm. Gelber ja Isen (2011).

²⁰ Hiljattain esimerkiksi Fryer (2011) löysi hyvin vähäisiä vaikutuksia oppilaiden palkitsemiselle. Tuloksen yleistettävyyttä heikentää se, että osa palkinnoista tuli toiminnasta, joka ei suoraan edistänyt oppimista ja sen lisäksi suuri osa oppilaista oli mahdollisesti liian nuoria, jotta he olisivat pystyneet edistämään oppimistaan omatoimisesti. Kokeen yhteydessä tehty kysely mm. osoittaa, että oppilaat eivät ymmärtä-

ta syistä. Ensinnäkin nuoret ovat lyhytnäköisiä, eli he haluavat tehdä asioita, joista on hyötyä nyt (tv:n katselu) eikä myöhemmin (opiskelu). Rahalla palkitseminen voi muuttaa tehtävien tärkeysjärjestyksen käänteiseksi. Palkitseminen myös helpottaa oppilaan budjettirajoitetta: rahoittaakseen opintojaan nuoren ei välttämättä tarvitse käydä ansiotyössä. Lisäksi joissakin nuorison alakulttuureissa hyvä koulumenestys ei ole ansio, joten rahan ansaitseminen voi olla ainoa hyväksytty syy hankkia parhaat mahdolliset arvosanat.

Angrist ja Lavy (2009) tutkivat Israelissa tehdyn kokeen tuloksia. Kokeessa opinnoissaan heikosti suoriutuvia oppilaita palkittiin, jos he suorittivat Bagrutin ja/tai pärjäivät jossakin sen yksittäisessä kokeessa hyvin. Korkein mahdollinen palkinto oli 10000 Israelin shekeliä (noin 2000 euroa), noin kolmannes tulosta, minkä lukion keskeyttänyt nuori keskimäärin verottomana ansaitsee. Angrist ja Lavy osoittivat, että tytöt reagoivat kannustimiin odotetusti ja paransivat suoritustaan. Erityisen voimakkaasti rahallisiin kannustimiin reagoivat tytöt, jotka jo ennestään menestyivät suhteellisen hyvin. Tällainen tulos on odotettava, sillä heikolla osaamistasolla on vaikeaa itsenäisesti kuroa umpeen kahden ensimmäisen vuoden aikana syntynyttä osaamisvajetta opintojen viimeisenä vuonna. Sen sijaan pojille kannustimilla ei näyttänyt olevan mitään vaikutusta, mikä on hyvin yleinen tulos koulutukseen liittyvissä tutkimuksissa. Angrist ja Lavy selittävät tulosta sillä, että aiemman tutkimuksen mukaan nuorilla tytöillä on vastaavanikäisiä poikia kehittyneempi kyky työskennellä kurinalaisesti ja siirtää palkintoa tuonnemmaksi.²¹

4.2 Vastuuohjelmien ja palkitsemisen seurauksia Yhdysvalloissa

Vastuuohjelmat ovat olleet laajassa käytössä jo vuosikymmenten ajan Yhdysvalloissa, joten valtaosa tuloksista on sieltä. Esimerkiksi Jacob (2005) ja Jacob ja Levitt (2003) ovat saaneet negatiivisia tuloksia Chicagon peruskouluista. Positiivisena tuloksena voidaan mainita, että kun koulujen oppilaiden menestystä alettiin tarkemmin seurata ja koulut saatettiin entistä tiukemmin vastuuseen oppilaidensa oppimistuloksista, oppilaiden tulokset todella paranivat ja oppimistuloksilla mitattuna politiikalla saatiin vastaava hyöty kuin aiemmissa luokkoon pienentämistä arvioivissa kokeissa (Jacob, 2005). Mutta nuorempien oppilaiden tulokset paranivat vain valtakunnallisissa kokeissa, eivätkä lainkaan osavaltion yhteisissä kokeissa. Lisäksi koetulosten tarkempi tarkastelu osoitti, että suurelta osin parantunut koemenestys johtui oppilaiden omasta aikaisempaa suuremmasta panoksesta oppimiseen ja kokeen kannalta tärkeiden taitojen kehittymisestä. Lisäksi saatiin jonkin verran todisteita siitä, että opettajat reagoivat kannusteisiin strategisesti jättämällä heikompia oppilaita luokalleen tai sijoitta-

neet miten edistää omaa oppimistaan esimerkiksi oppikirjaa lukemalla tai pyytämällä apua opettajalta tai vanhemmilta.

²¹ Tämä tulos viittaa siihen, että heikoimpien poikien ongelmat eivät ole hoidettavissa pelkästään opetusta lisäämällä tai kannustamalla rahallisesti, vaan ongelmat ovat syvemmällä.

malla heitä erityisluokkiin, jolloin he eivät kuuluneet testattuun ryhmään. Sen lisäksi, aivan Holmströmin ja Milgromin (1991) mallin ennusteen mukaisesti, opettajat käyttivät enemmän aikaa testiaineiden (äidinkieli ja matematiikka) opettamiseen muiden aineiden kustannuksella. Koska opettajien luonteva vastaus kannusteisiin oli lisätä sitä toimintaa, josta heitä palkittiin, se ei suinkaan tapahtunut kokonaistyöpanosta lisäämällä vaan vähentämällä muuhun opetukseen käytettyä aikaa, mikä oli tavoitteiden vastaista.

Järjestelmän sääntöjen kanssa pelaaminen voi johtaa edellistä hurjempaan tulemiin. Edellä kuvailun tutkimuksen kanssa osin samaa aineistoa käyttäen Jacob ja Levitt (2003) osoittavat, että ainakin neljässä prosentissa Chicagon peruskoulujen luokista opettajat huijasivat standardoitujen kokeiden arvostelussa. Huijaaminen oli yhteydessä myös opettajien kannusteiden muuttumiseen: vuonna 1996 tehdyn uudistuksen myötä kouluja alettiin asettaa koeajalle, mikäli alle 15 prosenttia oppilaista saavutti kansalliset tavoitteet lukutaidon kokeessa. Jos koeajalla olevat koulut eivät saaneet aikaan riittävästi parannuksia, ne saatettiin jopa sulkea. Lisäksi, 3.-, 6.- ja 8.-luokan oppilailta alettiin vuonna 1996 vaatia tietyn vähimmäisstandardin saavuttamista lukutaito- ja matematiikan kokeissa, jotta heidät voitiin päästää luokalta. Jacob ja Levitt (2003) näyttävät, että erityisesti huonoiten pärjäävissä luokissa huijaamisen todennäköisyys kasvoi molempien uudistusten myötä.

Edellä kuvatuunlaisia tuloksia on saanut myös Fryer (2011) New Yorkin kaupungin kouluista: opettajien palkitseminen ei vaikuttanut oppilaiden oppimistuloksiin, poissaoloihin tai valmistumiseen. Sen sijaan Fryer ja kumppanit (2012) osoittivat, että Kahnemanin ja Tverskyn lanseeraama tappioiden välttäminen, jota käsiteltiin lyhyesti luvussa 2.1., tuotti tulosten paranemisia. Kaksi ryhmää sai palkkion perinteiseen tapaan vuoden lopuksi, ensimmäinen opettajan omasta ja toinen ryhmän keskimääräisestä suorituksesta. Kahdelle ryhmälle jaettiin bonus etukäteen ja sen sai pitää, jos opettajan oma suoritus tai opettajan ryhmän suoritus oli palkitsemisen arvoinen. Palkitsemisella ei ollut merkitystä perinteisesti palkittujen opettajien oppilaiden testituloksiin, mutta palkintonsa etukäteen saaneet ja sen menettämistä pelkäävät opettajat saivat aikaan tilastollisesti merkitseviä oppimistulosparannuksia sekä yksin että ryhmässä.²²

Lyhytaikaisilla kokeiluilla ei välttämättä saada pysyviä käyttäytymismuutoksia, ts. opiskelu- tai opetustavat eivät muutu pysyvästi, vaan kouluissa palataan vanhoihin käytäntöihin heti, kun kannusteet poistuvat. Ainakin yksi tulos, joka puhuu pysyvämpien muutosten puolesta, on saatu Yhdysvaltojen alakouluista (elementary schools). Rouse ym. (2007) arvioivat Floridan järjestelmän vaikutuksia oppimistuloksiin ja opetusmetodeihin pitkällä aikavälillä. Floridassa kou-

²² Levitt ja kumppanit (2012) saavat samankaltaisia tuloksia oppilaiden rahallisesta palkitsemisesta. Oppilaat, jotka saivat 20 dollaria *ennen* kokeen alkua, saivat tilastollisesti merkitsevästi parempia koe-arvosanoja kuin muiden ryhmien oppilaat.

lut joutuivat vastuuseen oppilaidensa oppimistuloksista: jos koulu sai heikoimman arvosanan F (asteikolla A-F) useana vuonna peräkkäin, sen oppilaat saivat voucherin, joka oikeutti oppilaat siirtymään parempaan kouluun. Tutkimuksessa seurattiin heikoimpia kouluja kolme vuotta. Kävi ilmi, että opettajat todella panostivat opettamiseen monin eri tavoin. Kouluissa muun muassa lisättiin opetuksen määrää ja opettajat keskittivät huomionsa heikoimmin pärjääviin oppilaisiin. Näiden muutosten seurauksena oppimistulokset paranivat ja vaikutukset kestivät pidempään. Kannusteilla, myös rankaisevilla, voidaan siis saada aikaan muutoksia, mutta järjestelmästä täytyy mitä ilmeisimmin tehdä pysyvä. Myös palkkion tai rangaistuksen todennäköisyys ja vaikutus kouluun tai yksittäiseen opettajaan on oltava riittävän suuri, jotta sillä on merkitystä. Luultavasti suurin hyöty on saavutettavissa heikoimmissa kouluissa, joissa on eniten mahdollisuuksia parantaa oppimistuloksia.

Vaikka kouluvalinta ei ole tämän tutkimuksen varsinaisen mielenkiinnon kohteena, on tässä yhteydessä hyvä kommentoida joitakin siihen liittyviä tuloksia. Cullen ym. (2006) kävivät läpi Chicagon yläasteen valintaan liittyvää aineistoa. Chicagossa oppilas saa vapaasti hakeutua oman alueensa kouluun ja jos hakijoita on enemmän kuin paikkoja, paikan saavat valitaan arpomalla. Vertaamalla arvonnassa voittaneita ja hävinneitä, voidaan mitata ”paremman” koulun, ts. oppimisympäristön, vaikutusta oppimistuloksiin. Kävi ilmi, että arvonnassa voittaneet oppilaat eivät menestyneet hävinneitä paremmin yhdelläkään perinteisellä mittarilla mitattuna: esimerkiksi oppimistuloksissa, valmistumisasteessa, osallistumisasteessa tai suoritettujen kurssien lukumäärissä ei havaittu tilastollisesti merkittävää eroa. Itse asiassa tutkimuksessa saatiin tukea sille, että arvonnassa ”parempaan” kouluun päätyneet oppilaat menestyivät luokan sisäisessä arvosanavertailussa heikommin kuin muihin kouluihin päätyneet. Erityisesti ne oppilaat, jotka päätyivät omaa lähikouluun selvästi ”parempaan” kouluun, keskeyttivät opintonsa muita useammin.

Cullen ym. (2006) arvioivat negatiivisten tulosten johtuvan siitä, että oppilaat kokivat olevansa väärässä joukossa eivätkä olleet tottuneet olemaan luokkansa heikoimpia. Tutkimuksen ainoa positiivinen tulema oli, että arvonnassa voittaneet joutuivat selvästi harvemmin kurinpidollisten toimien kohteeksi tai kahnauksiin virkavallan kanssa. Heitä siis koulun vaihto auttoi pääsemään irti haitallisista käyttäytymismalleista. Nämä tulokset sotivat sitä Suomessa yleisesti esitettyä ajatusta vastaan, että oppilaat hyötyisivät mahdollisimman heterogeenisestä oppilasjoukosta.²³

²³ Itse asiassa Lavy ym. (2012) tutkivat heikkojen oppilaiden (luokan uusijoiden) osuuden vaikutusta muiden oppilaiden oppimistuloksiin. Mitä suurempi määrä uusijoita luokassa oli, sitä haitallisempi vaikutus sillä oli muiden oppilaiden koulumenestykseen.

5 Kokemuksia palkitsemisohjelmista maailmalta

Edellisessä luvussa esiteltiin joitakin eri koulutusasteiden kouluissa toteutettujen palkitsemisohjelmien vaikutuksia sekä niihin liittyviä ongelmia. Tässä luvussa esitellään Yhdysvaltojen ja Chilen kouluissa sovellettujen vastuu- ja palkitsemisohjelmien ominaisuuksia sekä niiden vaikutuksia oppilaiden oppimistuloksiin. Kaikkiin esiteltyihin ohjelmiin on kuulunut myös lukioita, joten näitä ohjelmia voi tarkastella siitä näkökulmasta, voisiko niitä, tai osia niistä, soveltaa myös Suomen lukioissa. Aivan kaikkia tuloksia ei kuitenkaan voi sellaisenaan soveltaa Suomeen, sillä joko esiteltyt järjestelmät eroavat Suomen järjestelmästä tai oppilaspopulaatio poikkeaa Suomen lukioiden oppilasaineksesta. Merkitystä on myös sillä, mitä ongelmaa palkitsemisohjelmilla on pyritty korjaamaan. Koska eri ongelmien korjaamiseen soveltuvat usein eri lääkkeet, on näissä ohjelmissa saatujen tulosten anti Suomen koulujärjestelmän kannalta rajallinen.

5.1 Yhdysvallat

Erilaisia palkitsemisohjelmia on ollut käytössä erityisesti Yhdysvalloissa, jossa on sovellettu mitä moninaisimpia käytäntöjä koulupiiritasolta maanlaajuisiin järjestelmiin. Tässä luvussa esitellään Yhdysvaltojen kokemuksia turnajaistyyppisistä palkitsemisohjelmista, joissa palkittavina ovat olleet kokonaiset koulut (peruskoulut ja/tai lukiot), ja koulut on asetettu paremmuusjärjestykseen ainakin osittain oppilaiden koetulosten perusteella. Ohjelmat eroavat kuitenkin melko merkittävästi toisistaan, ja luvun tarkoituksena onkin kuvata, millä eri tavoin koulujen palkitsemista on toteutettu.

5.1.1 Dallas Independentin koulupiiri, Texas

Dallasin koulupiirin koulujen palkitsemisohjelmaa alettiin soveltaa lukuvuonna 1991 – 1992. Palkitsemisen piiriin kuului sekä peruskouluja että lukioita, ja ohjelman pyrkimyksenä oli saattaa kaikki koulupiirin koulut vastuuseen oppilaiden oppimistuloksista ja vähitellen parantaa niitä. Webster ym. (1997) ja Ladd (1999) kuvailevat tarkkaan Dallasin palkitsemisohjelman, jossa koulut asetettiin kaksivaiheisesti paremmuusjärjestykseen sen mukaan, kuinka paljon koulut saivat parannettua oppilaidensa oppimistuloksia vuoden aikana suhteessa osavaltion keskimääräisiin oppimistuloksiin. Kyseessä oli siis eräänlainen arvonnäköalämalli. Ensimmäisessä vaiheessa kunkin koulun oppilaiden vuoden t ja vuoden $t-1$ koetulokset estimoitiin erilaisten oppilaiden ominaisuuksien, kuten sukupuolen ja etnisen taustan, funktiona. Tämä toimenpide tuotti jäännöstermejä, jotka voitiin tulkita koetuloksina, jotka oli ”puhdistettu” oppilaiden sosioekonomisen ja etnisen taustan vaikutuksista. Toisessa vaiheessa vuoden t jäännöstermejä selitettiin vuoden $t-1$ jäännöstermeillä. Tästä seuranneet jäännöstermit estimoivat oppimistulosten parantumista vuoden aikana. Nämä parannukset koottiin yhteen koulutasolle laskemalla parannusten keskiarvo kaikissa kouluaineissa jokaiselle vuonna t

koulussa opiskelleelle oppilaalle. Vaikka koulut asetettiin paremmuusjärjestykseen niiden suhteellisten tuotosten perusteella, ohjelmaan sisällytettiin myös tietty parannusstandardi, joka koulujen tuli saavuttaa tullakseen harkituiksi palkkion saajiksi. Tällä haluttiin varmistaa, että koulut, joissa absoluuttiset oppimistulokset eivät parantuneet, eivät myöskään saaneet palkkiota.

Websterin ym. (1997) ja Laddin (1999) mukaan oppilaiden sosioekonomisen taustan huomioon ottamisella pyrittiin siihen, etteivät arvonlisämittarit olisi suosineet parempiosaisten oppilaiden kouluja sekä siihen, että kaikilla kouluilla olisi ollut yhtäläinen mahdollisuus voittaa palkkio. Oppilaiden oppimistuloksia mitattiin lukuisilla eri mittareilla, jotta opettajilla ei olisi ollut kannusteita alkaa opettaa kapea-alaisesti jotakin tiettyä koetta varten. Esimerkiksi lukioissa otettiin huomioon oppilaiden tulokset ns. ACP-kokeessa, SAT-kokeessa, ACT-kokeessa sekä PSAT-kokeessa, joista kolme viimeistä on valtakunnallisia kokeita. Koetulosten lisäksi kouluja arvioitiin erilaisilla koulunlaajuisilla mittareilla, kuten oppilaiden läsnäoloasteella, kokeisiin osallistumisasteella ja pudokkaiden osuudella.

Jaettavien palkkioiden kokonaissumma sovittiin etukäteen. Websterin ym. (1997) mukaan koulut jaettiin luokkatasojen mukaan tiettyihin ryhmiin ja kullekin ryhmälle allokoitiin tietty osuus kokonaissummasta. Palkkiot jaettiin ryhmien parhaille kouluille kunnes kunkin ryhmän palkkiosumma oli käytetty. Koulujen välinen kilpailu ja vertailu tapahtuivat ryhmien sisällä, sillä oletettavasti eri luokkatasoja opettavat koulut ovat ominaisuuksiltaan erilaisia. Palkituissa kouluissa rehtori, kaikki opettajat ja muut työntekijät saivat palkkion, huolimatta heidän työpanoksiensa suuruudesta. Noin 20 prosenttia koulupiirin kouluista sai palkkion ja palkkioiden suuruudet olivat kaikissa kouluissa samat. Myöhemmin kuitenkin rehtorit ja opettajat niissä kouluissa, jotka saivat aikaan suurimpia parannuksia suhteessa odotettuihin parannuksiin, saivat myös suurempia palkkiota.

Ladd (1999) huomauttaa, että ohjelmalla oli sekä puolustajansa että vastustajansa. Jälkimmäiset kritisoivat erityisesti ohjelman monimutkaisuutta, sen perustamista koulujen suhteelliselle menestykselle sekä sitä, että ohjelmaa alettiin toteuttaa ilman muita perusteellisimpia koulutusreformia. Koulupiiri nimittäin päätti käyttää palkitsemisohjelmaansa ikään kuin ylhäältä ohjattuna katalyyttina muille muutoksille. Lyhyellä aikavälillä ohjelman tarkoituksena oli saada koulujen henkilöstö keskittymään oppilaiden oppimisen parantamiseen. Pitemmällä aikavälillä tavoitteena oli kannustaa opettajat vaatimaan parempia ammatillisia kehittymismahdollisuuksia, auttaa päättäjiä tunnistamaan tehottomat koulut sekä tehdä koulujen autonomian kasvattamisesta poliittisesti hyväksyttävämpää.

Ladd (1999) on tutkinut Dallasin ohjelman vaikutuksia ainoastaan 3.- ja 7.-luokkalaisten koetuloksiin, lukiopudokkaiden määrään ja rehtoreiden vaihtuvuuteen ohjelman neljän ensimmäisen vuoden aikana. Verrattuna muihin suuriin texasilaiskaupunkeihin ja muihin koulutusreformia soveltaviin kaupunkeihin,

ohjelmalla oli positiivisia vaikutuksia valkoisten ja latinalaisamerikkalaisten 7.-luokkalaisten koetuloksiin, muttei mustien 7.-luokkalaisten koetuloksiin. 3.-luokkalaisten oppimistuloksissa ei havaittu parannuksia. Lukiopudokkaiden osuus puolestaan laski Dallasissa suhteellisesti enemmän kuin muissa kaupungeissa ohjelman ensimmäisen kolmen vuoden aikana, mutta on epäselvää, missä määrin se oli palkitsemisohjelman ansiota. Kaikkien luokkatasojen kouluissa rehtoreiden vaihtuvuus kasvoi ohjelman käyttöönoton jälkeen, mikä kertoo siitä, että ohjelma ja korostunut vastuullisuus oppimistuloksista kannusti koulupiiriä vaihtamaan tehottomat rehtorit tehokkaampiin. Vaihtuvuus kasvoi ennen kaikkea tehottomimmissa kouluissa.

Dallasin ohjelman opetusten soveltuvuus Suomen lukioihin on kyseenalaista, sillä yhdysvaltalainen yhteiskunta ja järjestelmä poikkeavat Suomen vastaavista huomattavasti. Yhdysvalloissa väestöryhmien väliset koulutuserot ja tuloerot ovat maailman suurimpia, Suomessa vastaavasti maailman pienimpiä. Lisäksi ainoa lukioissa havaittu merkittävä muutos oli pudokkaiden osuuden lasku. Suomessa lukiopudokkaiden osuus on Tilastokeskuksen tietojen mukaan vain noin 2 prosenttia. Lisäksi, vaikka Dallasissa arvonlisän mittaaminen on tehty huolellisesti, suomalainen aineisto ei mahdollista vastaavaa, sillä ylioppilastutkinto on ensimmäinen valtakunnallinen koe, johon kaikki lukiolaiset osallistuvat. Sama kritiikki pätee osin seuraavassa kappaleessa esitettäviin palkitsemisohjelmiin ja niiden tulosten hyödyllisyyteen Suomen järjestelmälle.

5.1.2 Charlotte-Mecklenburgin koulupiiri, North Carolina

Charlotte-Mecklenburgin koulupiirin koulujen palkitsemisohjelma otettiin käyttöön lukuvuonna 1992 – 1993 ja sen tarkoituksena oli parantaa oppilaiden, erityisesti vähemmistöoppilaiden, heikkoja oppimistuloksia. Kelley ym. (2000) ovat kattavasti kuvanneet ohjelman piirteitä ja huomauttavat, että Charlotte-Mecklenburgin koulupiirissä alettiin toteuttaa yhtäaikaisesti myös muita koulutusreformejä, toisin kuin Dallasissa. Kouluille asetettiin tavoitteita (standardeja) yhdeksällä eri alueella, ja riippuen luokkatason koostumuksesta, kouluilla oli 14–44 alatavoitetta. Oppilaiden oppimistuloksia arvioitiin standardoiduilla monivalintakokeilla, mutta tavoitteita asetettiin myös muun muassa oppilaiden poissaoloasteelle, lukiopudokkaiden osuudelle ja vaikeampien kurssien osallistumisasteelle. Vuosina 1991–1992 määriteltiin lähtötasot kullekin yhdeksälle alueelle, ja kullekin alueelle määriteltiin parannustavoitteet jokaiselle seuraavalle kouluvuodelle. Joka vuosi asetettiin koko koulupiirinlaajuinen tavoite kullakin alueella, minkä jälkeen koulupiiri asetti numeeriset parannustavoitteet kullakin alueella erikseen kullekin koululle koulun oppilaiden edellisvuoden oppimistuloksiin perustuen. Charlotte-Mecklenburgissa sovellettiin siis arvonlisämallia, johon oli yhdistetty myös standardijärjestelmän piirteitä.

Koulut saivat pisteitä, mikäli saavuttivat eri alueiden tavoitteita. Kullekin koululle oli jaossa 100 pistettä, jotka jaettiin tasan alatavoitteille: jos koululla oli esi-

merkiksi 20 alataavoitetta, kunkin tavoitteen saavuttamisesta koulu sai viisi pistettä. Koulu sai 100 prosenttia kunkin alataavoitteen pisteistä, mikäli se saavutti tavoitteen täysin, 75 prosenttia, mikäli se saavutti tavoitteen osittain eikä yhtään pistettä saavuttamattomasta tavoitteesta. Tällä tavoin kerätyt pisteet laskettiin yhteen koulun yhteispistemääräksi. Koulut jotka saivat 75 pistettä tai enemmän luokiteltiin esimerkillisiksi ja jokainen niiden opetus- ja tukihenkilöstöstä sai palkkion. Koulut jotka saivat 60–74 pistettä luokiteltiin korkealaatuiseksi, ja jokainen niiden opetus- ja tukihenkilöstöstä sai palkkion, joka oli kuitenkin pienempi kuin palkkiot parempien koulujen henkilöstölle. Vaikka huonosti pärjääville kouluille ei määriteltykään erityisiä rangaistuksia, jatkuvasti huonosti menestyvät koulut saattoivat saada koulupiirin avustusta erilaisten oppimistulosten parantamiseen tähtäävien strategioiden ja resurssien muodossa.

Koska palkitsemisohjelman ohella käynnistettiin myös muita reformeja, pelkän palkitsemisen vaikutuksia oppilaiden oppimistuloksiin on vaikea erottaa muiden reformien vaikutuksesta. Oppimistulosten ja muiden tärkeiden indikaattoreiden kehitystä näiden reformien alkuvuosien (1992–1996) aikana ovat kuitenkin tutkineet Smith ja Mickelson (2000). Esimerkiksi lukio-oppilaiden tulokset SAT-kokeessa paranivat huomattavasti ohjelman alussa, mutta sittemmin tasoittuivat. Trendi oli kuitenkin samanlainen muidenkin koulupiirien kouluissa, joissa ei ollut käytössä palkitsemisohjelmaa. Charlotte-Mecklenburgin koulupiirin oppilaiden menestyminen lukion kursseilla muun muassa matematiikassa, biologiassa ja historiassa, ei yleisesti ottaen parantunut tai huonontunut verrattuna muiden koulupiirien oppilaisiin, ja Charlotte-Mecklenburgin oppilaat menestyivät heikommien verrattuna muiden koulupiirien oppilaisiin. Lukiopudokkaiden osuuksien muutostrendit puolestaan olivat samankaltaiset kaikissa kouluissa, mutta Charlotte-Mecklenburg'ssa osuus oli kuitenkin selvästi suurempi kuin muualla kautta tarkasteluajanjakson.

Charlotte-Mecklenburgin palkitsemisohjelman vaikutuksia on tutkittu ennen kaikkea sen kannalta, kuinka ohjelma onnistui motivoimaan opettajia ja kuinka opettajat suhtautuivat palkitsemiseen. Kelley'n ym. (2000) mukaan opettajat ymmärsivät ohjelman tavoitteet ja sitoutuivat niihin ja pitivät selkeitä päämäärien ja vastuullisuuden luomista ohjelman parhaina vaikutuksina. Opettajat myös muuttivat opetuskäytäntöjään paremmin vastaamaan ohjelman tavoitteita ja suurin osa heistä pyrki aktiivisesti saavuttamaan oppilaiden oppimistulostavoitteet. Toisaalta palkittujen koulujen opettajat pitivät seuraavan vuoden tavoitteiden ja palkkion saavuttamista selvästi vaikeampana kuin ei-palkittujen koulujen opettajat.

Vaikka monien opettajien mielestä palkkion saaminen koulukohtaisista parannuksista oli tarkoituksenmukaista, kaikki eivät kuitenkaan motivoituneet palkkion mahdollisuudesta. Monien opettajien mielestä palkkio oli muun muassa liian pieni suhteessa lisäponnisteluihin, joita palkkio edellytti. Monet muut asiat kuitenkin motivoivat heitä, kuten julkisen tunnustuksen saaminen, tavoitteiden saavuttaminen, oppilaiden kehityksen näkeminen ja muiden opettajien kanssa

työskenteleminen. Toisaalta opettajat pelkäsivät julkista kritiikkiä saavuttamattomista tavoitteista, ammattiympäryyden menettämistä ja työn menettämistä. Opettajat myös epäilivät sekä ohjelman että sen rahoituksen kestävyyttä. He eivät myöskään vakuuttuneet, että koulua palkittaisiin, vaikka se saavuttaisikin sille asetetut tavoitteet, ei edes silloin, kun koulu oli jo aiemmin kertaalleen palkittu. Kaiken kaikkiaan opettajat eivät pitäneet ohjelmaa uskottavana. Jotkin opettajista myös kokivat painetta saavuttaa tavoitteet, stressaantuivat ja tekivät ylitöitä. Opettajat kuitenkin pitivät kohtuullisena sitä, että heidät asetettiin vastuuseen oppilaidensa oppimistuloksista, mutta eivät olleet täysin yksimielisiä siitä, että ohjelma kohteli kouluja tasapuolisesti.

Kelley ym. (2000) tutkivat myös, oliko opettajien motivaatiolla mahdollisesti vaikutusta koulun saamien pisteiden määrään. Motivaatiota mitattiin sillä, kuinka vahvasti opettajat kokivat, että heidän työpanoksellaan oli merkitystä oppimistulostavoitteiden saavuttamiselle. Useiden eri estimointien mukaan näin mitatulla motivaatiolla oli tilastollisesti merkitsevä ja positiivinen vaikutus koulun saamiin pisteisiin. Estimoinneissa ei kuitenkaan havaittu, että koulut, joissa palkkion saamisen todennäköisyys oli suurempi ja täten opettajien motivaatio mahdollisesti korkeampi, olisivat saaneet enemmän pisteitä.

Charlotte-Mecklenburgin ohjelman arvioinneista saatu tieto opettajien tuntemuksista on arvokasta; monet opettajat reagoivat ohjelmaan muuttamalla opetuskäytäntöjään ja sitoutuivat siihen. Arvokas tieto on sekin, että motivoituneiden opettajien koulussa koepisteet olivat korkeammat. Silti opettajat eivät ilmeisesti juuri motivoituneet parempien palkkioiden toivossa, ehkä siksi, että eivät uskooneet koulunsa mahdollisuuteen tulla palkituksi. Lisäksi, oppilaiden osalta Charlotte-Mecklenburgin ohjelma epäonnistui tavoitteissaan, sillä se ei aikaansaanut mitattavia parannuksia oppimistuloksissa, mikä on ristiriidassa opettajien reaktioiden kanssa. Tämän kokemuksen perusteella kannustejärjestelmää suunniteltaessa on tärkeää kiinnittää huomiota palkkioiden suuruuteen, niiden mahdollisiin vaikutuksiin opettajien toimintaan ja motivaatioon sekä siihen, kuinka ohjelma ”myydään” kouluille ja opettajille niin, että sen toimivuuteen ja kestävyYTEEN luotetaan.

5.1.3 Kentuckyn osavaltio

Kentuckyn osavaltionlaajuinen koulujen palkitsemisohjelma käynnistettiin luvuonna 1991–1992 ja se oli osa laajempaa osavaltion koulutusreformia. Palkitsemisohjelmaa ovat kuvanneet Willis ym. (1999) sekä Kelley ym. (2000), joiden mukaan uudistusten tarkoituksena oli muun muassa motivoida opettajia saamaan aikaan parempia oppimistuloksia, keskittää oppilaiden ja opettajien huomio asetettuihin tavoitteisiin ja lisätä informaatiota koulujen toiminnasta ja menestyksestä niin opettajien, oppilaiden, vanhempien kuin päättäjienkin käyttöön. Koulut asetettiin vastuuseen nk. vastuuindeksin avulla, joka sisälsi sekä kognitiivisen (5/6) että ei-kognitiivisen (1/6) komponentit. Kognitiiviseen komponenttiin kuu-

lui 4.- 8.- ja 12.-luokkalaisten (lukion viimeisellä luokalla olevien oppilaiden) koetulokset erilaisissa aineryhmissä, ei-kognitiiviseen puolestaan pudokkaiden osuus, läsnäoloaste, jne. Oppilaiden saamat koetulokset luokiteltiin arvioinnissa neljään luokkaan – aloittelija, harjoittelija, pätevä ja etevä – joista koulu sai tietyn määrän pisteitä. Koko koulun vastuullisuusindeksin kognitiivinen komponentti laskettiin antamalla kuhunkin luokkaan kuuluvien oppilaiden osuudelle painokerroin, joka antoi enemmän arvoa paremmin menestyneiden oppilaiden koetuloksille.

Koulujen tekemiä parannuksia, toisin sanoen arvonlisää, mitattiin kahden vuoden ”vastuullisuusyhteisöissä”, jotta voitiin varmistua siitä, etteivät parannukset olleet kertaluonteisia. Ensimmäiset tilastotiedot kerättiin lukuvuonna 1991–1992, ja niistä muodostettu indeksiluku määritteli kunkin koulun lähtötason. Ensimmäinen vastuullisuusyhteisö sisälsi lukuvuodet 1992–1993 ja 1993–1994, toinen sykli lukuvuodet 1994–1995 ja 1995–1996, jne. Koulujen odotettiin parantavan indeksilukuaan 10 prosentilla erotuksesta koulun lähtötason (edellisvuoden indeksiluku) ja pitkän aikavälin tavoitteen (kaikkien oppilaiden koetulokset luokiteltavissa luokkaan ”etevä”) välillä. Jos koulut ylittivät tämän 10 prosentin tavoitteen ja saivat siirrettyä ainakin 10 prosenttia oppilaista aloittelijaluokasta harjoittelijaluokkaan, ne saivat palkkion. Palkkion suuruus määräytyi sen mukaan, kuinka selvästi koulu ylitti tavoitteen ja kuinka suuri sen henkilöstö oli. Palkkiotasoa määriteltiin kaiken kaikkiaan 52. Palkkion suuruudelle asetettiin kuitenkin katto, joka oli 10 prosenttia keskimääräisestä opettajan palkasta niissä Kentuckyn viidessä koulupiirissä, joissa keskimääräinen palkka oli suurin. Palkkion vähimmäissuuruus oli puolestaan puolet tästä enimmäissuuruudesta. Myöhemmin palkkiotasoa kuitenkin vähennettiin. Koulut saivat jakaa saamansa palkkion halumallaan tavalla, esimerkiksi palkan bonuksina.

Koulut luokiteltiin indeksilukujensa perusteella viiteen luokkaan – palkittavat, menestyvät, kehittyvät, heikentyvät ja kriisikoulut. Kahdessa viimeisessä luokassa koulut eivät saavuttaneet edes edellisen syklin indeksilukuaan. Heikentyvien koulujen tuli laatia ulkopuolisen teknisen avustajan johdolla kehityssuunnitelma ja ne saivat lisäresursseja. Kriisikoulujen kohtelu oli kutakuinkin samanlaista, mutta teknisellä avustajalla, joka tyypillisesti oli jokin kokenut opettaja tai hallintovirkamies, oli laajat oikeudet muun muassa erottaa opettajia.

Oppimistuloksista ja niiden kehityksestä Kentuckyn koulureformien aikana, Willis ym. (1999) toteavat, että kaikkien koulutustasojen kouluissa trendi oli yleisesti ottaen myönteinen. Lukioissa kognitiivisen komponentin eri aineryhmille lasketuissa indekseissä kaikki kehittyivät positiivisesti vuodesta 1993 vuoteen 1998: esimerkiksi lukemista mittaava indeksi parani 20,2:sta 51,4:ään, tiedeaineiden indeksi 27,1:stä 47:ään, ja taide- ja humanististen aineiden indeksi 14,3:sta 22,2:een. Indeksiluvut saattoivat tosin myös notkahtaa tai saada suurempia arvoja tarkasteluajanjaksolla. Kuten Charlotte-Mecklenburgin koulupiirin tapauksessa,

ei ole selvää missä määrin nämä parannukset olivat palkitsemisohjelman, missä määrin muiden reformien tulosta.

Kelley ym. (2000) ovat tutkineet ohjelman vaikutusta opettajien motivaatioon samalla tavoin kuin Charlotte-Mecklenburgin koulupiirissä, päätyen samansuuntaisiin tuloksiin. Kentuckyssa opettajat eivät olleet aivan yhtä mieltä siitä, auttoivatko ohjelman tavoitteet fokuoimaan opetusta paremmin kuin aiemmin. Kuitenkin selvästi pienempi osuus kentuckylaisista kuin Charlotte-Mecklenburgin opettajista ilmoitti, ettei tavoittele ohjelman asettamia tavoitteita. Tämä johtui mahdollisesti siitä, että Kentuckyssa koulun huonosta menestyksestä seurasi tiukempia rangaistuksia. Opettajat olivat hyvin tietoisia ohjelman rangaistuksista ja kokivat ne todellisena uhkana kouluille, jotka eivät parantaneet oppimistuloksiaan. Rangaistuksen uhka myös motivoi opettajia ammatillisen autonomian, ammattiylpeyden ja työpaikan menettämisen pelolla. Rangaistusten pelko sai opettajat myös uskomaan, että ohjelmaan osallistumisesta seuraisi lisääntyneitä työtunteja, paineita ja stressiä.

Näyttäisi siltä, että ainakin Kentuckyssä ennemmin rangaistukset kuin palkkiot motivoivat opettajia saavuttamaan kouluille asetetut tavoitteet. Opettajat kuitenkin uskoivat, että vaikka he *yksilöinä* yrittäisivät saavuttaa tavoitteet, heidän *koulunsa* ei saavuttaisi palkkion edellyttämiä tavoitteita edes 40 prosentin todennäköisyydellä. Lisäksi he uskoivat koulunsa välttävän rangaistuksen alle 45 prosentin todennäköisyydellä heidän omista ponnisteluistaan huolimatta, eli opettajat olivat sangen pessimistisiä. Palkitsemisohjelmasta motivoituneimmat opettajat olivat niitä, jotka olivat tyytymättömiä senhetkiseen palkkaansa, kokivat palkkion suuruuden sopivaksi ja palkitsemisen oikeudenmukaiseksi. Opettajat, jotka halusivat pois ohjelman piiristä eli toiseen kouluun tai tyystin toiseen ammattiin, olivat niitä, joiden mielestä ohjelma oli epäreilu ja heidän koulunsa mahdollisuus voittaa palkkio oli muita heikompi. Samoin kuin Charlotte-Mecklenburgin koulupiirissä, opettajien motivaatiolla oli positiivinen ja tilastollisesti merkitsevä vaikutus koulun saamaan indeksilukuun.

Kentuckyn ohjelman vaikutuksia oppilaiden osaamiseen on vaikea arvioida, sillä siitä puuttuu kontrolliryhmä. Emme siis tiedä, missä määrin havaitut muutokset olivat ohjelman ansiota, vai johtuivatko ne kokonaan joistakin muista samankaisista reformeista. Siksi palkitsemisohjelmia olisi syytä testata ensin pienellä, satunnaisesti valitulla joukolla kouluja, jotta niiden oppilaiden suorituksia voitaisiin verrata muiden koulujen oppilaiden suorituksiin. Suurten reformien huolellinen suunnittelu ja pilottihankkeet auttavat välttämään virheitä ja korjaamaan ohjelmaa tarvittaessa.

5.1.4 Kalifornian osavaltio

Kalifornian osavaltion koulujen palkitsemisohjelma käynnistettiin vuonna 1985 ja sen tavoitteena oli taata oppilaille mahdollisuus tasa-arvoiseen ja laadusta tin-

kimättömään koulutukseen. Kaliforniassa etnisesti erilaisten oppilasryhmien välillä on selkeitä eroja oppimistuloksissa ja näiden erojen kiinnikurominen on yksi osavaltion opetusministeriön päätavoitteista. Palkkion saavat parillisina vuosina peruskoulun 1.–6. -luokkia opettavat koulut (ala-asteet) ja parittomina vuosina 7.–9. -luokkia opettavat koulut (yläasteet) sekä lukiot.²⁴

Kaikilla lukioilla ei ollut mahdollisuutta saada palkkiota. Ensiksi, edellisenä palkitsemisvuonna voittaneet koulut eivät saa hakea palkkiota. Toiseksi, koulujen on täytynyt täyttää ns. AYP -vaatimukset (Adequate Yearly Progress), jotka perustuvat ainoastaan osavaltionlaajuisten kokeiden koetuloksiin. Vaatimukset kasvat vuoteen 2014 asti, jolloin kaikkien koulujen kaikkien oppilaiden tulee olla ns. pätevällä tai paremmalla tasolla koetuloksilla mitattuna. Koulujen on täytynyt täyttää myös ns. API-vaatimukset (Academic Performance Index). AYP- ja API-vaatimusten lisäksi koulujen täytyy täyttää jokin neljästä vaatimuksesta liittyen eri oppilasryhmien välisten oppimistuloserojen kaventamiseen. Esimerkiksi jonkin vähemmistöryhmän, kuten afroamerikkalaisten, latinalaisamerikkalaisten tai sosioekonomisesti heikompien, oppilaiden API-luvun tulee kasvaa enemmän kuin valkoisten tai aasialaisten oppilaiden API-luku tai jonkin vähemmistöryhmän API-luvun kasvun täytyy ylittää sillä asetettu kasvutavoite.²⁵

API on osavaltion vastuullisuusohjelman yksi keskeisimmistä kulmakivistä ja sen tavoitteena on mitata koulujen oppimistuloksia ja niiden muutosta. API on numeerinen indeksiluku, jonka arvot voivat vaihdella 200:n ja 1000:n välillä. API-luku perustuu koetuloksiin osavaltionlaajuisissa kokeissa sekä muihin indikaattoreihin. Kukin koulu raportoi API-lukunsa kaksi kertaa vuodessa, tammikuussa ja elokuussa. Koulun aikaansaamat parannukset lasketaan vähentämällä elokuun luvusta tammikuun luku. Jotta koulu täyttäisi API-vaatimukset ja sillä olisi mahdollisuus saada palkkio, sen elokuun API-luvun tulee olla vähintään 720, ja sen täytyy saavuttaa vuosittainen koulunlaajuinen kasvutavoitteensa sekä jokaiselle oppilasryhmälle erikseen asetetut tavoitteet. Koulunlaajuinen kasvutavoite on viisi prosenttia koulun tammikuun API-luvun ja osavaltion asettaman 800 pisteen tavoitteen välisestä erotuksesta, kuitenkin vähintään viisi pistettä. Koulujen, joiden API-luku on jo 800 tai enemmän, tulee vähintäänkin säilyttää saavuttamansa taso. Eri oppilasryhmille asetetut tavoitteet ovat samankaltaiset, mutta tammikuun API-luvut lasketaan kullekin ryhmälle erikseen. API-pisteidensä perusteella koulut asetetaan paremmuusjärjestykseen sekä osavaltiossa että samankaltaisten koulujen kesken. Jälkimmäisessä vertailussa koulun API-lukua verrataan 100 muuhun kouluun, joilla on ”samankaltaisia haasteita”.²⁶ Koulun saaman palkkion suuruus riippuu sen sijoituksesta, ja jokainen koulun henkilöstön jäsen saa samansuuruisen palkkion.

²⁴ <http://www.cde.ca.gov/ta/sr/cs/proginfo.asp>

²⁵ <http://www.cde.ca.gov/ta/ac/ay/cefayp.asp>; <http://www.cde.ca.gov/ta/sr/cs/eligibility09.asp>

²⁶ <http://www.cde.ca.gov/ta/ac/pa/cefpsaa.asp>

Kalifornian palkitsemisohjelman vaikutuksia oppimistuloksiin on tutkittu etenkin siitä näkökulmasta, onko sen peruseriaate perusteltu. Ohjelma nimittäin olettaa, että niille kouluille, joilla on jo lähtökohtaisesti korkea API-luku, on vaikeampi parantaa tuloksiaan, joten näiden koulujen ei tarvitse saada palkkion saadakseen yhtä suuria absoluuttisia parannuksia kuin heikompien koulujen. Driscoll ym. (2008) tutkivat tämän oletuksen oikeellisuuden lisäksi myös muiden tekijöiden, kuten oppilaan perheen tulojen ja koulutuksen, koulun ja koulupiirin suuruuden, luokkakoon, asukastiheyden, kilpailevien yksityiskoulujen ja sijaisopettajien, vaikutusta API-luvun kasvuun.

Driscoll ym. (2008) näyttävät, että lukion, jonka API-luku on 100 pistettä alempi kuin toisen samankaltaisen lukion, voi odottaa parantavan API-lukuaan noin 23 pistettä enemmän. Toisin sanoen, jokainen olemassa oleva API-piste vähentää odotettua pisteiden kasvua 0,23 pisteellä, mikä viittaa siihen, että paremmat koulut eivät saa yhtä suuria API-lukujen parannuksia kuin heikommät koulut. Tulos osittain osoittaa myös, että paremmat koulut eivät pysty manipuloidaan API-pisteitä valmistelemalla oppilaitaan kokeiden odotettuja sisältöjä varten, tai ainakaan tämä manipulointi ei ole riittävän tehokasta kumotakseen vähitellen alenevien parannusten vaikutuksen.

Lisäksi kaikkien koulutustasojen koulut, joilla on samankaltaiset API-luvut, mutta joiden oppilaat ovat varakkaista perheistä, saavuttavat suurempia parannuksia. Lukioiden, joiden oppilaiden perheiden mediaanitulo on 10 000 dollaria korkeampi kuin muuten samankaltaisen lukion, voi odottaa parantavan API-lukuaan noin 4 pistettä muita enemmän. Kaikilla koulutustasoilla yksityiskoulujen oppilaiden osuuden vaikutus API-luvun kasvuun on negatiivinen, mikä on vastoin odotuksia, sillä voisi olettaa, että kilpailu yksityiskoulujen kanssa motivoisi julkisia kouluja parantamaan API-lukuaan. Asukastiheys puolestaan vaikuttaa lukioiden API-luvun kasvuun positiivisesti: mikäli kaupunkialueilla vanhemmat eivät työssäkäynniltään ehdi osallistua lastensa koulunkäyntiin, koetulokset voivat motivoida ohjelmauudistuksiin, joita ei muuten tapahtuisi.

Driscollin ym. (2008) tutkimus näyttää, että kun standardoidut kokeet otetaan koulujen paremmuusjärjestyksen perusteeksi, koulujen pisteiden kasvun vaihtelu heijastaa sekä koulujen että niiden oppilaiden ominaisuuksia. Jos palkkioita ei mukauteta vastaamaan näitä eroavaisuuksia, tavoitteiden saavuttamisen vaikeus voi lannistaa opettajat ja koulut ja saada ne luopumaan parempien oppimistulosten tavoittelusta. Nykyisin Kalifornian koulujen ja niiden oppilaiden ominaisuudet otetaan ilmeisesti jollain tavalla huomioon, kun kouluja vertaillaan ryhmissä, joiden kouluilla on ”samankaltaisia haasteita”. Seuraava esimerkki Chilen palkitsemisohjelmasta kuitenkin osoittaa, ettei hienostuneinkaan ryhmittely aina luo kaikille kouluille yhtäläisiä kannusteita tavoitella ohjelman palkkioita.

5.2 Chile

Myös Latinalaisessa Amerikassa on sovellettu erilaisia koulujen ja opettajien palkitsemisohjelmia. Koulujen palkitsemisohjelmista on esimerkkejä muun muassa Boliviasta ja El Salvadorista, joissa ohjelmien tarkoituksena on ollut kannustaa koulujen opettajia ja muuta henkilökuntaa työskentelemään yhdessä ja parantamaan oppilaille tarjotun koulutuksen laatua. Esimerkiksi El Salvadorissa palkkio jaetaan opettajille niissä kouluissa, jotka saavuttavat opetusministeriön asettamat oppimistulostavoitteet.

Kaikista tunnetuin palkitsemisohjelma löytyy kuitenkin Chilestä, jossa kansallinen koulujen arviointijärjestelmä (Sistema Nacional de Evaluación de Desempeño de los Establecimientos Educacionales Subvencionados, SNED) on ollut käytössä vuodesta 1996 lähtien. Ohjelma on suunnattu kaikille peruskouluille ja lukioille, jotka saavat rahoitusta valtiolta, olivat ne sitten kunnallisia tai yksityisiä. SNEDin tavoitteena on parantaa opettajien motivaatiota ja tarjota yhteisöille tietoa koulujen laadusta vertailemalla keskenään sellaisia kouluja, jotka palvelevat sosioekonomisesti samankaltaisia lapsia (Mizala ja Romaguera, 2004).

Rau ja Contreras (2011) sekä Mizala ja Romaguera (2004) esittelevät artikkeleissaan SNEDin keskeisimmät piirteet. Parhaat koulut valitaan kahden vuoden välein, ja koulukohtaisen palkkion suuruus määräytyy koulun oppilaiden määrän mukaan. Kaiken kaikkiaan palkkioita jaetaan niin monelle koululle, että palkitut koulut kattavat 25 prosenttia tietyin alueen oppilaista. 90 prosenttia palkkiosta tulee jakaa opettajille heidän työtuntiansa suhteessa. Koulu allokoi loput 10 prosenttia niille opettajille, joiden osuus oppimistulosten parantamisessa on ollut merkittävä. Palkkio maksetaan neljännesvuosittain. 1990-luvun lopussa ja 2000-luvun alussa opettajien saama palkkio edusti noin 5–7 prosenttia heidän vuosittaisesta palkastaan. Koska osuus oli näin pieni, SNEDin vaikutukset opettajien käyttäytymiseen arveltiin rajallisiksi. Vuosina 2004–2006 oppilaskohtaista palkkiota nostettiin niin, että palkittujen koulujen opettajat saivat noin kaksinkertaisen palkkion edellisvuosiin verrattuna (Mizala ja Romaguera, 2005). Samalla kuitenkin palkittavien koulujen ja opettajien määrää nostettiin määrittelemällä, että kouluille, jotka kattavat 35 prosenttia jonkin alueen oppilasluvusta, jaetaan palkkio. Ylempään 25 prosentin kategoriaan kuuluvat koulut saivat täyden palkkiosumman, kun taas seuraavan 10 prosentin kategoriaan kuuluvat koulut saivat 60 prosenttia täydestä palkkiosummasta.

Ennen kuin koulut asetetaan paremmuusjärjestykseen, ne ryhmitellään ensin alueen, sitten tarkemman sijainnin (kaupunki/maaseutu) mukaan. Peruskoulut ja lukiot erotetaan omiksi ryhmikseen. Tämän jälkeen koulut ryhmitellään vielä neljään alaryhmään oppilaiden sosioekonomisten ominaisuuksien perusteella. Näitä ovat oppilaan perheen tulot, vanhempien koulutustaso sekä koulun ns. haavoittuvuusindeksi, joka kuvaa sen palveleman populaation köyhyyden aiheuttamaa haavoittuvuutta. Kouluja vertaillaan ja asetetaan paremmuusjärjestykseen

näiden sosioekonomisesti samankaltaisten ryhmien sisällä, jotta vertailu olisi tasapuolinen.

Kaikille kouluille määritellään vuosittain SNED-indeksit, joiden perusteella koulut asetetaan paremmuusjärjestykseen. SNED-indeksi koostuu kuudesta komponentista, joiden sisältöä Mizala ja Romaguera (2004) sekä Rau ja Contreras (2011) ovat selostaneet. Ensimmäinen komponentti on tehokkuus (37 prosenttia indeksistä), joka mittaa koulun oppimistuloksia suhteessa koulun palvelemaan populaatioon. Oppimistuloksia mitataan 4. – 8. – ja 10. – luokkalaisten keskimääräisillä koetuloksilla ns. SIMCE -kokeissa matematiikassa ja äidinkiessä. Toinen komponentti on kehitys (28 prosenttia indeksistä), joka mittaa SIMCE-kokeiden tulosten parannuksia. SNED siis yhdistää indeksissään sekä oppimistulosten tason että niiden parannusten seuraamisen. Kolmas komponentti on aloitteellisuus (6 prosenttia indeksistä), jota mitataan koulujen koulutusprojekteilla, opetustyöpajoilla ja muilla aktiviteeteilla, jotka kuvaavat koulun kykyä omaksua koulutuksellisia innovaatioita ja ulkopuolista opetuksen ohjausta. Neljäs komponentti on työolojen kehitys (2 prosenttia indeksistä), joka perustuu koulun opetusministeriön tarkastusjärjestelmän luokitukseen. Viides komponentti on mahdollisuuksien tasa-arvoisuus (22 prosenttia indeksistä), joka muodostuu koulujen luokkia kertaavien oppilaiden ja pudokkaiden osuudesta sekä siitä, kuinka koulut kohtelevat oppilaitaan. Koulut eivät esimerkiksi saa syrjiä oppilaita erottamalla heitä koulusta luokan reputtamisen perusteella, eivätkä saa käyttää kurinpidollisia toimenpiteitä muista kuin oppilaiden käytökseen liittyvistä syistä. Kuudes komponentti on opettajien ja vanhempien integraatio (5 prosenttia indeksistä), joka perustuu koulutason kyselyihin aktiviteeteista, joilla koulut kannustavat integraatioon ja informaation jakamiseen sekä kyselyihin, joilla selvitetään vanhempien mielipiteitä koulujen laadusta.

Vaikka SNED-indeksin koostumus on monipuolinen ja melko hienostunut, se lienee kuitenkin hallinnollisesti melko raskas. Mikäli vastaavanlainen indeksi rakennettaisiin Suomen lukioiden arviointia varten, kaikkea em. informaatiota ei luultavastikaan olisi saatavilla, vaan se tulisi kerätä nimenomaan indeksia varten, mistä aiheutuisi mahdollisesti merkittäviäkin kustannuksia. Kerätyn informaation laatu voisi kuitenkin paikoitellen olla melko kyseenalaista (esim. kyselyistä saatu informaatio), mikä heikentäisi koko indeksin tarkoituksenmukaisuutta ja luotettavuutta. Lisäksi Chile on kehittyvä maa, jossa on suuret sosiaaliset erot ja tuloerot. Suomi on yksi maailman tasa-arvoisimpia maita, joten näin hienojakoinen luokittelu Suomessa tuskin on tarpeen. Koska Chilen järjestelmällä pyritään korjaamaan hyvin erilaisia ongelmia kuin mitä Suomessa on, se ei ole sellaisenaan sovellettavissa Suomeen.

Carnoy ym. (2007) kritisoivat SNED-ohjelmaa siitä, että se kannustaa kouluja keskittymään oppimistulosten parantamiseen vain niillä luokkatasoilla, jotka osallistuvat SIMCE-kokeisiin. Sen sijaan, että samojen oppilaiden koetuloksia vertailtaisiin pitkällä aikavälillä, tietynä palkitsemisvuotena tarkastellaan tietyn

luokkatason, esim. 4.-luokkalaisten, saamia koetuloksia ja verrataan niitä edellisen palkitsemisvuoden 4.-luokkalaisten saamiin koetuloksiin. Tästä johtuen ei ole varmaa, että SNED lisää jonkin tietyn kohortin, tai tietyn 4.-luokkalaisten, oppimista. SNED ei ohjelmana tarjoa tähän kannusteita ja taidot, joita vaaditaan koetulosten parantamiseksi tietyssä kokeessa, eivät ole samat kuin ne taidot, joita vaaditaan oppimisen parantamiseen.

Carnoy ym. (2007) tutkivatkin, parantavatko SNEDin mukaisesti jaetut palkkiot lainkaan koetuloksia kohorttien sisällä ja toteavat, että palkituissa kouluissa eivät keskimääräiset kohorttien sisäiset koetulokset parantuneet enempää kuin ei-palkituissa kouluissa. Tulos herättää kysymyksiä siitä, mitä SNEDillä oikeastaan halutaan palkita, millaiseen toimintaan se kannustaa kouluja sekä voidaanko näillä kannustimilla saavuttaa toivottuja koulutuksellisia tavoitteita. Carnoy ym. (2007) tulosten mukaan näyttääkin siltä, että saadakseen palkkioita useammin, koulun olisi järkevintä keskittyä parantamaan 4.-, 8.-, ja 10.-luokkalaisten koetuloksia joka toinen vuosi, eikä auttamaan oppilaita kehittymään näiden luokkatasojen välillä.

Rau ja Contreras (2011) puolestaan tutkivat, oliko SNEDin käyttöönotolla vaikutuksia oppilaiden koetuloksiin vertailemalla ohjelmaan osallistuvien koulujen keskimääräisiä koetuloksia ennen ja jälkeen ohjelman käyttöönottoa vuonna 1996. Heidän mukaansa koetulokset matematiikassa ja äidinkielessä paranivat tilastollisesti merkittävästi vuosien 1995 ja 1997 välillä. He näyttävät myös, että ohjelman vaikutus näkyy pääasiassa kouluissa, joiden palkkion saamisen todennäköisyys on suurempi kuin 0,6 tai 0,65, mikä viittaa siihen, että osa kouluista olisi varmoja häviäjiä. Tämän tuloksen puolesta puhuu myös se, että lähes puolet ohjelmaan osallistuvista kouluista ei kertaakaan voittanut palkkiota vuosina 1996–2007. Rau ja Contreras (2011) toteavatkin, että ohjelma näyttäisi luovan kannusteita vain murto-osalle kouluista. Mizala ja Romaguera (2004) puolestaan huomauttavat, että SNED tarjoaa vahvimmat kannustimet opettajille niissä kouluissa, joiden SNED-indeksi on lähellä sitä indeksilukua, joka erottaa voittajakoulut häviäjäkouluista.

Edellä esitellyt tutkimukset osoittavat, että edes varsin pitkälle mietitty ja monta eri ongelmallista aspektia huomioon ottava palkitsemisohjelma ei välttämättä tarjoa opettajille ja kouluille oikeanlaisia kannusteita. Vaikka SNED-ohjelma otti huomioon koulujen oppilaiden sosioekonomisen taustan ja vertaili vain samankaltaisia kouluja keskenään, niin koulujen muodostamien ryhmien sisällä näyttäisi silti olevan eroja siinä, kuinka hyvin ohjelma onnistuu kannustamaan kouluja parantamaan oppilaidensa koetuloksia. Vaikka ryhmiä olisi kuinka monta – SNEDin tapauksessa 109 vuosina 2002 ja 2003 – on mahdollista, että tietyt koulut ovat jatkuvasti ryhmänsä kärjessä, toiset taas häntäpäässä, ja että kannustinten vaikutukset näille kouluille ovat hyvin erilaiset ja erisuuruiset. Ongelmallista on myös se, että vaikka ohjelmassa on arvonlisää eli koetuloksien parantumista mittaava komponentti, sen laskentatapa ei kerro mitään yksittäisen

oppilaan kehittymisestä eikä myöskään kannusta opettajia panostamaan siihen. Chilen palkitsemisohjelman tulosten perusteella voisikin sanoa, että hallinnon ja ehkä kannusteiden kannalta voisi olla järkevää karsia mahdollisesti liian yksityiskohtaiseksi tehtyä ohjelmaa.

6 Lukioiden rahoitus- ja arviointijärjestelmiä Euroopassa

6.1 Iso-Britannia

Euroopassa ei juuri ole käytössä lukioiden tuloksellisuuteen perustuvaa rahoitusta, vaan rahoitus määräytyy useimmiten lukioiden oppilasmäärän mukaan. Tässä luvussa on ensiksi esitelty Isossa-Britanniassa käytössä olevat lukioiden arviointi- ja rahoitusmallit, sillä vaikka siellä ei suoranaista palkitsemisohjelmaa olekaan käytössä, lukioiden arviointi ja osittain rahoituskin perustuvat oppilaiden koulumenestykseen tai sen kehitykseen. Myös muiden Euroopan maiden lukioiden arviointi- ja rahoituskäytäntöjä esitellään, ja keskustellaan siitä, millaiset mahdollisuudet mailla on arviointikäytäntöjään kehittää.

6.1.1 Lukioiden arviointi

Iso-Britannia on siinä mielessä poikkeus Euroopan maista, että siellä kouluja, koulutusasteesta riippumatta, arvioidaan erityisellä arvonnisää mittaavalla mittarilla. Ison-Britannian lukioita (Key Stage 5) on perinteisesti arvioitu tavallisella arvonnisäämittarilla, mutta vuosina 2008–2010 käytössä oli ns. kontekstuaalinen arvonnisäämittari, joka vertasi oppilailtaan samankaltaisia kouluja. Opetusministeriö (DfE, 2010) päätti kuitenkin luopua tästä mittarista, koska sitä pidettiin vaikeasti ymmärrettävänä, eikä sitä pidetty sen parempana menestyksen ennustajana kuin pelkkiin koetuloksiin perustuvia mittareita. Kontekstuaalinen arvonnisäämittari asetti myös erilaiset odotukset eri oppilasryhmille heidän taustansa perusteella, mitä pidettiin jopa moraalisesti vääränä. Päättäjät tulivatkin siihen tulokseen, että koulut tulisi asettaa paremmuusjärjestykseen sen perusteella, kuinka paljon ne parantavat kunkin oppilaan oppimistuloksia huolimatta heidän taustastaan.

Vuonna 2008 opetusministeriössä (DCSF, 2008) oltiin kuitenkin toista mieltä. Tällöin todettiin, että lukioita vertaillessa tulee tunnistaa oppilaiden eri lähtökohdat ja se, että koulut palvelevat taustoiltaan hyvinkin erilaisia oppilaita. Kontekstuaalista arvonnisäämittaria pidettiin oikeudenmukaisena menetelmänä mitattaessa koulun tehokkuutta ja hyvänä pohjana koulujen välisille vertailuille. Sillä saatuja vertailuja kehoitettiin kuitenkin tarkastelemaan rinnakkain taustatekijöistä riippumattoman koetulosaineiston kanssa. Opetusministeriön (DCSF, 2008) malli perustui tiettyssä kohortissa saavutettuihin koetuloksiin: mallissa laskettiin tiettyyn kohorttiin kuuluvien, samankaltaisten oppilaiden saama kansallinen koetuloskeskiarvo, eli tilastollinen ”ennuste”, ja verrattiin näiden oppilaiden koetuloksia ennusteeseen. Oppilaat, jotka saivat ennustetta paremmat koetulokset ”ansaitsivat” koulullensa positiivisia arvonnisäpisteitä, kun taas ennustetta huonommin menestyneet oppilaat hankkivat koulullensa negatiivisia pisteitä. Näiden pisteiden keskiarvo muodosti lukion arvonnisäpisteet.

Lukioiden kontekstuaalinen arvonlisämittari erosi kuitenkin muilla koulutus-tasoilla käytössä olleista arvonlisämittareista siinä, ettei se ottanut huomioon oppilaiden etnistä tai sosioekonomista taustaa, sillä opetusministeriöllä ei ollut saatavilla kyseistä tilastoaineistoa (DCSF, 2008). Se otti huomioon kuitenkin muita tekijöitä, kuten suoritettun tutkinnon vaativuuden ja aiotun opinto-ohjelman laajuuden. Arvonlisämittari mittasi oppilaiden menestystä sekä oppimisen laadun että määrän näkökulmasta ja tuotti korkeammat ennusteet oppilaille, jotka pyrkivät suorittamaan laajemmat opinto-ohjelmat. Oppilaita vertailtiin kuitenkin vain toisiin oppilaisiin, jotka suorittivat yhtä laajoja opinto-ohjelmia. Kunkin oppilaan pisteet määräytyivät erilaisista oppilaan yläasteella (Key Stage 4) saavuttamista arvosanoista, oppilaan A-levels -kokeisiin tai vastaaviin tutkintoihin tähtäävien opintojen määrästä, oppilaan sukupuolesta, oppilaan kohortin keskimääräisestä aikaisemmasta opintomenestyksestä ja näiden tekijöiden interaktioista. Oppilaan kontribuutio koulunsa kontekstuaalisiin arvonlisäpisteisiin saatiin, kun hänen todellisista pisteistään vähennettiin ennustetut pisteet.

Vuonna 2011 kontekstuaalisesta arvonlisämittarista kuitenkin luovuttiin kaikilla luokkatasoilla, jotta kouluilla olisi kannustimet panostaa sekä heikompiin oppilaisiin että sellaisiin oppilaisiin, joiden opintomenestys on lähellä jotain tavoitearvoa (DfE, 2010). Alemmilla luokkatasoilla (Key Stage 2 – Key Stage 4) otettiin käyttöön tavallinen arvonlisämittari, joka perustuu oppilaiden kahdeksaan parhaaseen tulokseen ja myöntää ”bonuksia” erityisen hyvistä saavutuksista englannin kielessä ja matematiikassa. Lukioissa kaikille julkinen kontekstuaalinen arvonlisämittari korvataan tulevina vuosina. Koulut ja opetusviranomaiset voivat kuitenkin seurata oppilaiden edistystä ns. oppilaiden saavutusten seuranta – työkalulla (Learner Achievement Tracker, LAT), joka soveltaa perinteisempää arvonlisämittaria oppilaiden kehityksen seurannassa (YPLA, 2010). LAT-työkalun mittaama arvonlisä näyttää oppilaan edistymisen suhteessa keskimääräiseen, samankaltaisten oppilaiden edistykseen samoissa tutkinnoissa ja oppiaineissa ja ottaa huomioon oppilaiden koulumenestyksen yläasteella. Työkalun tarkoituksena on ennustaa oppilaiden saavutuksia ja niiden perusteella määrittää tavoitearvosanoja tai läpäisyasteita, joihin oppilaat ja koulut voivat pyrkiä. Työkalun avulla voidaan myös arvioida, menestyvätkö jonkin tietyn koulun oppilaat yhtä hyvin, huonommin vai paremmin kuin muut oppilaat, jotka suorittavat samaa tutkintoa tai opiskelevat samoja aineita ja joilla on samankaltainen aiempi koulumenestys.

6.1.2 Lukioiden rahoitus

Isossa-Britanniassa on käytössä ns. kansallinen rahoituskaava (national funding formula), jonka mukaisesti kaikkia Ison-Britannian lukioita (6th form / college) rahoitetaan. Kunkin lukion saama vuosittainen rahoitus määräytyy neljän komponentin mukaan: koulussa opetetut tunnit, koulun ominaisuudet (alue, opetettujen oppilaiden tausta, opetetut kouluaineet ja oppilaiden menestys), vuosittain määräytyvä rahoitusaste per koulussa opetetut tunnit sekä lisätuki sellaisille luki-

oille, jotka joutuvat erityisesti panostamaan oppilaiden tukiopetukseen. Kertomalla komponentit keskenään saadaan yksittäisen koulun vuosittainen rahoitus. (YPLA, 2011a)

YPLA (2011b) määrittelee, että yksittäisen koulun ominaisuuksiin kuuluva oppilaiden ”menestysaste” saadaan jakamalla heidän vuoden aikana suorittamien kurssien lukumäärä suoritettaviksi aiottujen kurssien lukumäärällä. Suoritusta mitataan ainoastaan asteikolla hyväksytyt/hylätyt, eli oppilaiden koetuloksia ei oteta huomioon laskelmissa. Otetaan esimerkki lukioista, jossa on 100 oppilasta. Kaikki heistä opiskelevat vuoden aikana neljää oppiainetta, jolloin heidän suoritettavaksi aiottujen oppimistavoitteiden lukumäärä on $100 \cdot 4 = 400$. Mikäli vuoden aikana koulussa kertyi 300 hyväksytyä suoritusta, koulun menestysaste tälle vuodelle on $300/400 = 0,75$ eli 75 prosenttia. Koulujen rahoituskavassa käytettävä kerroin saadaan, kun tämä menestysaste jaetaan kahdella ja siihen lisätään 0,5. Tämä kerroin on suurempi kuin menestysaste ja lasketaan siksi, että opetuksesta koituu kustannuksia sekä siksi, että koulutuksella on itseisarvo, vaikkei oppilas läpäisisikään vuosikurssiaan. Vaikka oppilaiden koetuloksia ei otetakaan huomioon laskelmissa, kertoimen mukana olo koulujen rahoituskavassa kannustaa kouluja auttamaan oppilaidensa vuosikurssista läpäisyä.

6.2 Manner-Eurooppa

6.2.1 Lukioiden arviointi

Eurydice verkosto tuottaa tietoa EU:n 33 jäsenmaan koulutusjärjestelmistä. Useimmissa Euroopan Unionin jäsenmaissa kouluja, koulutusasteesta riippumatta, arvioi jokin ulkopuolinen instituutio (EACEA, 2012). Lisäksi koulujen itsearviointi on pakollista tai vahvasti suositeltavaa kaikkialla paitsi Belgian Valloniassa ja Irlannissa. Ulkoisista arvioinneista on usein vastuussa maan koulutusvirasto tai -ministeriö, mutta myös paikallisemmat tai keskusjohdosta itsenäiset tahot saattavat suorittaa arviointeja. 21:ssä Euroopan maassa koulujen ohella myös rehtorit tai muu johtava henkilöstö arvioi yksittäisiä opettajia. Koulujen arviointien on tarkoitus seurata tai parantaa koulujen toimintaa ja oppilaiden oppimistuloksia, eikä lopputuloksena syntyvät raportit yleensä sisällä arvioita yksittäisten opettajien toiminnasta.

Ulkopuolisten suorittamat koulujen arvioinnit ovat Euroopassa melko lailla standardisoidut, ja lukuvuonna 2010–2011 14 Euroopan maassa keskusjohtotason viranomaiset arvioivat kouluja standardoitujen kriteerien pohjalta (EACEA, 2012). Arviointikriteerit perustuvat kahteen komponenttiin: arvioitavan alueen johonkin mitattavaan ominaisuuteen sekä vaadittuun standardiin. Eri koulujen toimintoja koskevien kriteerien muoto voi kuitenkin vaihdella ja kaikkia kouluja ei välttämättä arvioida samassa arviointikehikossa: standardoituja kriteerejä voidaan käyttää esimerkiksi arvioinnin ensimmäisessä vaiheessa, jossa tunnistetaan riskinalaiset koulut. Mahdollinen toinen vaihe voi olla täysin sopeutettu tällaisten

koulujen tarpeisiin. Tämänkaltainen arviointijärjestelmä on käytössä Alankomaissa.

Kaikilla Euroopan mailla ei ole käytössään tarkkaa arviointikriteereiden listaa, vaan keskusjohtotason arvioijat määrittävät arvioitavat teemat kansallisesta lainsäädännön ja koulutuksellisten tavoitteiden perusteella (EACEA, 2012). Jotkut arvioinnit keskittyvätkin vain tiettyihin teemoihin, kuten säännösten tai koulujen kehityssuunnitelmien noudattamiseen. Tyypillisesti arvioidaan myös opettamisen ja oppimisen laatua, oppilaiden koulumenestystä sekä koulujen johtoa. Mikäli paikalliset tahot vastaavat koulujen arvioinnista, arviointikriteerit johdetaan usein lainsäädännöstä ja koulutustavoitteista. Keskusjohto saattaa kuitenkin joissakin tapauksissa määrittää paikallistason arvioinnille standardit, kuten Tanskassa. Myös Puolassa alueelliset arviointitahot ovat velvoitettuja käyttämään tiettyjä standardoituja kriteereitä.

Useimmissa Euroopan maissa perusasteen ja toisen asteen koulujen arvioinneissa otetaan huomioon oppilaiden koulumenestys (EACEA, 2012). Poikkeuksia ovat Viro, Itävalta, Slovenia ja Slovakia, joissa ulkopuolinen arviointi koskee pääasiassa koulujen toimintoja ja/tai niiden säännöstenmukaisuutta. Saksassa ja Espanjassa on erilaisia käytäntöjä hallintoalueesta riippuen. Tyypillisimmin arvioinneissa seurataan oppilaiden tuloksia kansallisesti standardoiduissa kokeissa, mutta muitakin indikaattoreita, kuten oppilaiden arvioita opettajistaan ja oppilaiden etenemistä koulutusjärjestelmässä, saatetaan seurata.

Oppilaiden koetulosten seurannassa arvioijat voivat yleensä tulkita tuloksia ilman, että vertaavat niitä mihinkään ylemmällä tasolla määriteltyihin standardeihin. Poikkeuksia tähänkin käytäntöön kuitenkin on: Unkarissa, Portugalissa ja Isossa-Britanniassa (Englannissa) on määritelty oppimistulosten taso, joka koulujen tulee saavuttaa välttääkseen erinäiset ulkopuolisten interventiot. Unkarissa kansallisissa kokeissa huonosti menestyvien koulujen tulee valmistella parempaan kehitykseen tähtäävä toimintasuunnitelma. Portugalissa heikkojen koulujen tulee määrittää aikataulu, jonka mukaan ne käynnistävät erilaisia korjaavia toimenpiteitä, mukaan lukien alisuoriutuville oppilaille annettavan tukiopetuksen. Isossa-Britanniassa koulut, joissa tietty osuus oppilaista ei saavuta asetettuja oppimistavoitteita, ovat oikeutettuja lisätukeen ja -seurantaan.

Lähes kaikki Euroopan maat seuraavat koulutusjärjestelmiään kansallisella tasolla vertaamalla järjestelmän aikaansaamia tuloksia asetettuihin tavoitteisiin ja standardeihin. Seuranta perustuu tyypillisimmin oppilaiden tuloksiin kansallisissa kokeissa sekä edellä esiteltyihin koulujen arviointeihin. Ainoa maa, joka ei käytä kumpaakaan näistä menetelmistä, on Itävalta, mutta tilanne on todennäköisesti muuttumassa sielläkin vuodenvaihteessa 2012–2013. Oppilaiden koetulosten kansallinen seuranta tarkoittaa tiettyjen ikäryhmien oppilaiden keskimääräisten koetulosten seuranta. Samoja tuloksia voidaan käyttää muun muassa tutkintojen

myöntämisen perusteena, koulujen toiminnan seurannassa ja oppilaiden tarpeiden määrittämisessä.

Kaikissa Euroopan maissa lukion hyväksytysti suorittaneet oppilaat saavat todistuksen, joka on yleensä vähimmäisvaatimuksena korkea-asteen koulutukseen pääsemiselle (EACEA, 2012). Kuudessa maassa – Espanjassa, Unkarissa, Puolassa, Slovakiassa, Ruotsissa ja Turkissa – todistus myönnetään viimeisen tai viimeisten lukiovuosien aikaisen jatkuvan arvioinnin perusteella. Unkarissa, Puolassa ja Slovakiassa on kuitenkin mahdollista saada todistus myös loppukokeen perusteella.

Lähes kaikissa Euroopan maissa todistuksen saaminen perustuu siis loppukokeeseen, joka usein perustuu sekä kouluissa suoritettaviin arviointeihin että ulkopuolisten tekemiin arviointeihin (EACEA, 2012). Siinä missä Belgiassa, Islannissa ja Liechtensteinissa arviointi suoritetaan vain kouluissa, Irlannissa, Ranskassa, Suomessa, Maltalla, Romaniassa, Sloveniassa ja Kroatiassa ulkopuolinen taho hoitaa arvioinnin. Useimmissa maissa loppukoe koostuu kirjallisesta ja suullisesta osasta – poikkeuksia ovat Bulgaria, Kreikka, Kypros, Liettua, Portugali, Suomi ja Kroatia, joissa on vain kirjallinen koe. Useimmiten jokin koulun ulkopuolinen taho on laatinut loppukokeen. Maiden välillä on eroja siinä, kuka päättää oppilaiden saaman lopullisen arvosanan. Arvosanan voi antaa koulun opettajat, perustuen joko heidän omiin kriteereihinsä tai ulkopuolelta määräytyneisiin kriteereihin, tai ulkopuolinen arviointitaho. Jälkimmäinen tapa on käytössä Irlannissa, Ranskassa, Liettuassa, Luxemburgissa, Maltalla, Romaniassa, Sloveniassa, Isossa-Britanniassa ja Kroatiassa.

Euroopan maiden kouluissa on laajalti käytössä kansallisia kokeita, joilla on erilaisia tarkoituksia (EACEA, 2012). Ensiksi, kokeet jotka summaavat oppilaiden saavutukset jonkin tietyn koulutusasteen päätteeksi voivat merkittävällä tavalla vaikuttaa oppilaiden jatkokoulutukseen. Tällaiset kokeet ovat yleensä kaikille koulujen oppilaille pakollisia. Toiseksi, kokeiden tarkoitus voi olla seurata ja arvioida kouluja ja koulutusjärjestelmää yleisesti. Kolmanneksi, kokeet voivat tukea oppimisprosessia tunnistamalla oppilaiden erityiset oppimistarpeet. Ensimmäiseen ryhmään kuuluvia kansallisia kokeita järjestetään useimpien Euroopan maiden lukioissa, joko lukion lopussa tai useampana lukiovuonna, ja ne ovat ainakin merkitykseltään verrattavissa Suomen ylioppilaskokeisiin. Toiseen ryhmään kuuluvissa maissa voi olla käytössä loppukoe, mutta se ei joko ole kansallinen tai se ei pyri määrittämään oppilaiden jatkokoulutuskelpoisuutta.²⁷

Koulujen arvonlisää, koulutusasteesta riippumatta, ei juuri OECD-maissa – mukaan lukien niihin kuulumattomissa Euroopan maissa – mallinneta tai mitata. Pienempiä alueellisia ja pilottihankkeita on kehitelty useammassa maissa ja laa-

²⁷ Tähän ryhmään kuuluvat Belgia, Tanska, Saksa, Kreikka, Espanja, Liettua, Luxemburg, Unkari, Itävalta, Islanti ja Liechtenstein

jemmin arvonlisää on mitattu Yhdysvalloissa ja Isossa-Britanniassa (OECD, 2008). Vuonna 2005 OECD:n koulutusosasto, Norjan edustajan toimesta, käynnisti hankkeen tutkimaan koulutason arvonlisämittareita OECD-maissa. Norjan hallitus alkoikin rahoittaa vuonna 2006 alkanutta projektia, johon osallistui 13 maata: Australia, Belgia, Tshekki, Tanska, Ranska, Alankomaat, Norja, Puola, Portugali, Slovenia, Espanja, Ruotsi ja Iso-Britannia. Projektin lopputuloksena syntyi OECD:n raportti *Measuring Improvements in Learning Outcomes – Best Practices to Assess the Value-Added of Schools* (2008), joka perustuu asiantuntijaedustajien tapaamisissa esittämiin raportteihin, ja jotkin edustajista laativat myös yksityiskohtaisempia osioita raporttiin.

OECD:n raportti (2008) esittelee olemassa olevia oppilaiden arviointeja, joita projektin osallistujamaissa voitaisiin hyödyntää arvonlisämallien ja -mittareiden lähtökohtana. Projektin osallistuvista maista Tshekin tasavallassa, Tanskassa, Ranskassa, Norjassa, Puolassa, Portugalissa, Sloveniassa ja Ruotsissa on käytössä lukioiden arviointeja, joille mahdollinen arvonlisämittari voisi perustua. Arvioinnit kuitenkin eroavat paljolti siinä suhteessa, kuinka laajan valikoiman oppiaineita ne kattavat. Toisen asteen koulutuksessa useimmat arvioinneista ovat Suomen ylioppilaskokeita vastaavia, jolloin koetuloksia kaikista testattavissa oppiaineista voidaan arvonlisämallista riippuen käyttää arvonlisän mittaamisessa. Tällaisia arviointeja on käytössä esimerkiksi Tshekin tasavallassa (*Maturita*), Ranskassa (*Baccalaureate*) ja Puolassa (*Matura*). Arvonlisän mittaamiseksi tarvitaan kuitenkin arviointi aiemmasta koulumenestyksestä, jonka tulee olla vertailtavissa käytettävään arviointiin. Monista osallistujamaista tällainen vertailukelpoinen aiempi arviointi puuttuu, mutta esimerkkejä yhdenmukaisista arvioinneista löytyy kuitenkin Isosta-Britanniasta, Tanskasta ja Belgian Flanderista.

OECD:n raportti (2008) esittelee myös muita kansallisia tietoaineistoja, joita osallistujamaat voisivat käyttää arvonlisämalleissaan. Käytännössä tämä tarkoittaisi sitä, että oppilaiden erilaiset sosioekonomiset taustat voitaisiin ottaa huomioon arvonlisän mallintamisessa. Useilla osallistujamailla on informaatiota esimerkiksi oppilaiden demografisista ominaisuuksista, mahdollisesta maahanmuuttotaustasta ja oppimisvaikeuksista, perheiden rakenteesta, koulutuksesta ja taloudellisista resursseista sekä oppilaiden ja perheiden käyttämistä sosiaali- ja terveyspalveluksista. Ei ole kuitenkaan täysin selvää, onko oppilaiden taustan huomioon ottaminen relevanttia arvonlisämalleissa, jotka jo hyödyntävät jotakin mittaustulosta oppilaiden aiemmasta koulumenestyksestä. OECD:n (2008) mukaan taustamuuttujien sisällyttäminen voi olla merkittävää kuitenkin silloin, kun oppilaita arvioidaan harvakseltaan: pitemmällä aikavälillä oppilaan sosioekonominen asema voi korreloida vahvasti oppilaan edistymisen kanssa arviointien välisellä ajanjaksolla. Taustamuuttujien sisällyttäminen voi myös lisätä koulutustahojen luottamusta arvonlisämalliin ja sen käytettävyyteen.

6.2.2 Lukioden rahoitus

Euroopassa koulutus kaiken kaikkiaan rahoitetaan pääasiassa julkisista varoista (EACEAM, 2012). Julkisilla varoilla tarkoitetaan kaikkea valtiolta, kunnilta tms. saatavaa rahoitusta. Yksityinen rahoitus puolestaan sisältää tyypillisesti kotitalouksien (oppilaiden tai vanhempien) maksamat lukukausimaksut sekä maksut yrityksiltä ja voittoa tavoittelemattomilta yhteisöiltä.

Euroopan komission alaisen Eurybasen kokoamien maaraporttien mukaan lukio-koulutuksen rahoitus on lähes yksinomaan valtion ja/tai kuntien vastuulla. Pääasiassa vapaaehtoisuuteen perustuvaa yksityistä rahoitusta perheiltä, yrityksiltä tai muilta organisaatioilta kerätään vain Isossa-Britanniassa, Portugalissa, Italiassa, Unkarissa ja Luxemburgissa. Kaikista Eurybasen maaraporteista ei käy ilmi lukioden rahoituksen perusteita, mutta useimmissa maissa rahoitus määräytyy pääasiassa oppilasmäärän mukaan. Joissakin maissa, kuten Isossa-Britanniassa, Belgian Flanderissa ja Luxemburgissa lukioden rahoituksessa otetaan huomioon joitakin oppilaiden sosioekonomiseen taustaan liittyviä tekijöitä. Tietävästi missään Euroopan maassa lukioden rahoitus ei perustu oppilaiden oppimistuloksiin. Kuriositeettina voi kuitenkin mainita Romanian, jossa koulujen rahoitus perustuu osittain ”opetuksen laatuun”.

Levačić (2008) esittelee artikkelissaan Euroopassa käytössä olevia koulujen rahoitusjärjestelmiä. Euroopassa kouluja, koulutustasosta riippumatta, rahoitetaan pääasiassa joko jonkin systemaattisen säännön tai ”hallinnollisen harkinnanvaraisuuden” pohjalta. Systemaattisia sääntöjä on kahdenlaisia. Ensimmäinen on ns. muutosääntö, joka muuntaa kullakin luokkatasolla opiskelevien oppilaiden määrän tarvittavien opettajien määräksi, kun luokkakokoa koskevat säännöt sekä lukujärjestyksessä määritellyt työtunnit on ennalta määrätty. Toista sääntöä edustavat erilaiset matemaattiset rahoituskaavat, jollainen on käytössä esimerkiksi Isossa-Britanniassa. Nämä kaavat sisältävät useita muuttujia (esimerkiksi oppilaiden lukumäärän kullakin luokkatasolla, koulun alueellisen sijainnin, sekä koulun oppilaiden köyhyyteen ja oppimistarpeisiin liittyviä indikaattoreita), joihin jokaiseen liittyy jokin rahamäärä. Rahoituskaavasta määräytyy kullekin koululle vuosittainen budjettitulo. Hallinnollisella harkinnanvaraisuudella puolestaan tarkoitetaan opetusviranomaisia, jotka eivät käytä systemaattisia sääntöjä rahoituspäätöksissään vaan määrittävät koulujen tarpeet jokaisessa koulussa tehtävien yksilöllisten arviointien perusteella. Tämä rahoitusmenetelmä sisältää rahoituksen, joka perustuu koulujen edellisvuotisiin kuluihin. Edellä mainittuja menetelmiä voidaan käyttää myös rinnakkain esimerkiksi määrittämään rahoituksen eri osioita.

Levačić (2008) ja Fazekas (2012) kuvailevat rahoituskaavojen rakennetta. Tyypillisesti kaavat voivat koostua neljästä komponentista, jotka sisältävät yhden tai useampia kaavassa huomioitavia muuttujia. Ensimmäisessä komponentissa määritellään oppilaskohtainen allokatio, jonka suuruus riippuu usein oppilaan luok-

katasosta tai iästä ja sen on tarkoitus kuvata eri luokkatasojen erilaisia opetusvaatimuksia. Toisessa komponentissa määritellään lisäresursseja tietyille lukujärjestyksen oppiaineille, kuten musiikille, kielille tai liikunnalle. Lisäresursseja voidaan tarvita lisäkurssien järjestämiseen, kalliimpien oppimateriaalien hankkimiseen, korkeampiin erityisopettajien palkkoihin tms. Kolmas komponentti ottaa huomioon oppilaiden erilaiset opetustarpeet: koulut tarvitsevat lisäresursseja esimerkiksi oppilaiden, joilla on oppimisvaikeuksia tai joilla on heikompi sosioekonominen tausta, opettamiseksi. Lisäresursseilla näille oppilaille voidaan tarjota tukiopetusta, erityisiä oppimateriaaleja ja pienempiä opetusryhmiä. Neljäs komponentti heijastaa koulujen erilaisia kustannuksia, joita aiheuttavat koulun koko, koulun yhteisön suhteellinen eristyneisyys, koulutilojen ominaisuudet sekä paikallinen hintataso.

Rahoituskaavat voivat sisältää joko yhden tai useamman em. komponenteista ja niiden saamat painokertoimet heijastavat koulutuspolitiikan tehokkuus- ja tasa-arvotavoitteita (Levačić, 2008). Fazekas (2012) kuitenkin huomauttaa, että koulujen rahoituskaavat ovat aina keskittyneet lähes yksinomaan koulutuksellisiin panoksiin tuotoksien sijaan. Joitakin poikkeuksiakin on, kuten edellä esitelty Ison-Britannian rahoituskaava tai Tanskan ns. taximeter stelmä. Silti, vaikka oppilaiden oppimistulokset otettaisiin kaavoissa jollain tavalla huomioon, järjestelmien ensisijainen tavoite on tunnistaa koulujen lisäresurssien tarve eikä niinkään palkita oppimistuloksista tai koulun tuottamasta arvonnalisästä. Fazekasin (2012) mukaan pääsyy rahoituskaavoista tuotosta mittaavien muuttujien puuttumiselle on yksityiskohtaisen ja luotettavan tietoaineiston puuttuminen.

Levačićin (2008) mukaan 2000-luvun alussa 14 Euroopan maan – Belgian, Tshekin tasavallan, Tanskan, Viron, Suomen, Unkarin, Latvian, Alankomaiden, Puolan, Portugalin, Espanjan, Ruotsin, Islannin ja Ison-Britannian – kouluissa oli käytössä rahoituskaava, joko kansallisesti tai vain tiettyjen opetusviranomaisten soveltamana. Rahoituskaava oli näiden maiden lisäksi käytössä myös Venäjällä. Levačić (2008) tarkastelee näistä maista seitsemää tarkemmin ja jakaa niiden käyttämät rahoituskaavat em. neljään komponenttiin. Hän tarkastelee kaavoja niiden kattavuuden, tasa-arvon ja tehokkuuden näkökulmasta. Levačić (2008) keskustelee myös siitä, voiko rahoituskaavan käyttö sellaisenaan parantaa oppilaiden oppimistuloksia. Näin voisi tapahtua, mikäli koulut osaisivat käyttää annettua rahoitussumman aikaisempaa tehokkaammin. Hänen mukaansa rahoituskaavan tuottamalla uudelleenallokaatiolla on sitä suurempi vaikutus oppimistuloksiin, mitä suurempi sen osuus koulujen budjeteista on ja mitä suurempi vaikutus annetulla oppimisresursseihin käytetyllä rahasummalla on oppimistuloksiin.

Vaikutuksia oppimistuloksiin on Levačićin (2008) mukaan kuitenkin vaikea havaita, sillä rahoituskaavat ovat tyypillisesti käytössä maanlaajuisesti, eli vertailu- ja rahoituskaavaa käyttävien ja ei-käyttävien koulujen välillä ei voida tehdä. Lisäksi rahoituskaavojen käyttöönnoton yhteydessä on usein otettu käyttöön muita

oppilaiden oppimistuloksien parantamiseen tähtääviä toimenpiteitä, joten pelkän rahoituskaavan käyttöönoton vaikutuksia on vaikea erottaa muiden käytäntöjen vaikutuksista. Oikein suunniteltu rahoituskaava voi kuitenkin edistää opetusresurssien tehokkaampaa käyttöä sekä oppilaiden välistä oikeudenmukaisuutta.

7 Suomi

7.1 Suomen lukioiden rahoituksen nykytila

Suomessa lukiokoulutuksen rahoittavat valtio ja kunnat yhdessä. Opetus- ja kulttuuriministeriö myöntää koulutuksen järjestäjälle luvan lukiokoulutuksen järjestämiseen ja päättää valtionosuuksista. Keskimäärin noin 42 prosenttia rahoituksesta tulee valtiolta.

Lukion valtiolta saama rahoitus määräytyy opiskelijaa kohden määrätyn, järjestäjäkohtaisen yksikköhinnan ja lukiokoulutuksen opiskelijamäärän tulosta (Opetushallitus, 2012). Kunnille ja kuntayhtymille määrättäviä yksikköhintoja porrastetaan koulutuksen järjestäjän lukiokoulutuksen alle 18-vuotiaina aloittaneiden lukio-opiskelijoiden kokonaismäärän perusteella seuraavasti. Yksikköhinta vastaa keskimääräistä yksikköhintaa, silloin kun lukion opiskelijoiden lukumäärä on vähintään 200. Mikäli opiskelijoiden määrä on alle 200 opiskelijaa ja vähintään 60 opiskelijaa, yksikköhintaa korotetaan kaavan $0,4 \times (200 - \text{lukiossa opiskelevien määrä})$ mukaisesti. Jos opiskelijoiden määrä on alle 60 opiskelijaa, yksikköhintaa korotetaan lisäksi luvulla, joka saadaan kaavasta $2,1 \times (60 - \text{lukioissa opiskelevien määrä})$. Yksikköhintaa korotetaan kuitenkin enintään luvulla 106, ts. maksimikorotus saavutetaan 40 oppilaalla. Lisäksi yksikköhintaa voidaan korottaa, jos koulutus on määrätty järjestettäväksi sisäoppilaitoksessa, koululle annetun erityisen tehtävän perusteella tai opetuskielen (suomi ja ruotsi) perusteella. Tässä suhteessa suomalainen rahoitusjärjestelmä on linjassa muissa Euroopan maissa käytettävien järjestelmien kanssa.

Rahoitusjärjestelmä ja laskentakaavojen painotukset heijastavat arvovalintoja, jotka ovat usein ääneen lausumattomia. Niiden avulla myös pyritään ohjaamaan toimijoita haluttuun suuntaan. Suomen kaavasta voidaan ”lukea”, että keskushallinto haluaa tukea pieniä lukioita, eli ylläpitää harvaan asuttujen alueiden kouluverkkoa. Myös erityistehtävän saaneet koulut, esimerkiksi musiikki- tai liikuntapainotteiset lukiot, voivat saada valtionosuuskorotusta. Valtio siis tukee lukiolaisten kulttuuri- ja liikuntaharrastusta. Samoin ruotsinkieliset koulut ja uskonnollisten yhteisöjen ylläpitämät koulut saavat korkeampaa valtionosuutta. Näitä arvoja nykyinen järjestelmä pitää tärkeinä.

Samalla tavalla opettajien palkan määräytyminen kuvaa, miten opettajien ammattaitoa arvostetaan. Suomessa lukion opettajan peruspalkka on määrätty alan työehtosopimuksessa, joka tehdään kunnalliselle ja yksityiselle puolelle erikseen. Lisäksi peruspalkka on jaettu kahteen kalleusluokkaan koulun maantieteellisen sijainnin suhteen. Sen ilmeisenä tavoitteena on saada hyviä opettajia pienempien asutuskeskusten kouluihin. Peruspalkan lisäksi opettajan palkkaan vaikuttavat tehtävän vaativuuden mukainen korotus, työkokemukseen perustuva lisä (nk. ikälisä), henkilökohtainen lisä sekä ylityötunneista ja erityistehtävistä maksetta-

vat lisät ja korvaukset. Kukin työnantaja pystyy työn vaativuuden ja opettajan henkilökohtaisen arvioinnin perusteella sovittamaan kunkin opettajan palkan työssä suoriutumisen mukaiseksi.

Käytännössä suomalaisten opettajien palkkausjärjestelmä jo sisältää tulospalkkauksen elementtejä. Tiedossamme ei ole, miten aktiivisesti näitä kahta komponenttia käytetään opettajien palkitsemisessa. Jos Suomessa siirryttäisiin lukioiden tuloksellisuusrahoitukseen, koulutuksen järjestäjillä olisi nykyisessä palkkausjärjestelmässä mahdollisuus siirtää hyvästä tuloksesta saatu palkkio koulun suoriutumiseen eniten vaikuttaneille opettajille.

7.2 Suomen ammattikoulujen tuloksellisuusrahoitus

Kuten lukioita, Suomen ammatillista peruskoulutusta rahoittavat sekä valtio että kunnat. Vuoden 2002 alusta lukien ammattikoulujen rahoitusta on asteittain muutettu siten, että rahoitus määräytyy nykyisin osittain ammattikoulujen tuloksellisuuden perusteella, sillä aiemman rahoitus- ja ohjausjärjestelmän, eli opiskelijamääriin perustuvan laskennallisen valtionosuusrahoituksen, ei koettu kannustavan kouluja riittävästi koulutuksen tuloksellisuuden ja laadun kehittämiseen (Virtanen, 2006). Tulorahoituksella on kaksi päätavoitetta: Ensiksi se pyrkii edistämään ammatillisen peruskoulutuksen tuloksellisuutta ja sille asetettavien tavoitteiden saavuttamista. Toiseksi, sen on määrä tuottaa kehittämis- ja ohjausinformaatiota sekä koulutuksen järjestäjien että opetushallinnon käyttöön.

Varsinainen rahoitusta koskeva lakimuutos astui voimaan vuoden 2006 alussa, minkä jälkeen tulorahoitus on ollut osa perusrahoitusta. Vuoden 2009 asetuksessa määrätään, että ammatillisen peruskoulutuksen yksikköhintaa porrastetaan tulosindeksin avulla. Porrastus tehdään kustannusneutraalisti siten, että ensin kaikilta koulutuksen järjestäjiltä vähennetään niiden yksikköhinnasta kolme prosenttia. Näin saatu rahasumma jaetaan parhaiden koulutuksen järjestäjien kesken siten, että paras viidennes koulutuksen järjestäjistä saa tulorahoitusta, mutta muilla tulosindeksin vaikutus yksikköhintaan on nolla. Jos vielä tuloksellisuuden perusteella tehtävä porrastus ylittää 10 prosenttia, ylittävästä osuudesta otetaan huomioon vain kolmannes. Hyvä tulos ei siis siirry koulutuksen järjestäjän saamaan rahoitukseen täysimääräisenä. Koska rahoitus ohjataan koulutuksen järjestäjälle eikä kouluihin, voi käydä niin, että se ei koskaan mene perille hyvän tuloksen tehneeseen kouluun sen opettajista puhumattakaan.

Arviomme mukaan vuonna 2012 voimassa olleiden sääntöjen perusteella 30 koulutuksen järjestäjää 140:stä oli oikeutettu saamaan tuloksellisuusrahoitusta. Korotukset näyttävät vaihdelleen nollan ja 13 prosentin välillä, kun keskimääräinen korotus on laskelmamme perusteella 5,7 prosenttia. Kun kaikilta koulutuksen järjestäjiltä yksikköhintaa alennettiin aluksi kolme prosenttia, ei siihen päälle tuleva vajaan kuuden prosentin korotus tunnu kovin kummoiselta. Jos yksikkö-

hinta ilman korotusta olisi keskimäärin 10 000 euroa, kolmen prosentin vähenyksen ja kuuden prosentin korotuksen jälkeen se olisi 10 253 euroa. Jos koulutuksen järjestäjällä olisi koulutettavanaan vaikkapa 700 oppilasta, hyvä tulos toisille 177 000 euron tuloksellisuuslisän. Koulutuksen järjestäjällä olisi tämä summa aiempaa enemmän käytettävissään laadun parantamiseen, opettajien palkittamiseen tai muuhun tarpeelliseksi katsomaansa toimintaan.

7.2.1 Mittariston osa-alueet

Virtanen (2006) on ammatillisen peruskoulutuksen tulosrahoitusmittariston arvioinnissaan esitellyt ja arvioinut aiemmin käytetyn mittariston pääpiirteitä. Sittemmin mittaristoa on muun muassa Virtasen selvitystyön jälkeen muutettu ja yksinkertaistettu. Vuodesta 2012 alkaen käyttöön otettu uusittu mittaristo on esitelty Opetushallituksen (2011) oppaassa.

Tulosrahoituksen pohjana toimivan mittariston arviointikohteet heijastavat kolmea osa-aluetta tai näkökulmaa: vaikuttavuutta, opettajien kelpoisuutta ja henkilöstön kehittämistä. Mittarit on pyritty muodostamaan siten, että ne hyödyntävät mahdollisimman paljon nykyisiä aineistoja, mikä pakottaa tekemään kompromisseja mittaamisessa. Esimerkiksi arvioinnin kohteena oleva aikajakso riippuu mittarista.

Vaikuttavuuden näkökulmaa, joka edustaa 90 prosenttia koko mittaristosta, heijastavat työllistymis-, jatko-opiskelu-, keskeyttämis- ja läpäisymittarit, jotka perustuvat Tilastokeskuksen koulutus- ja työssäkäyntitilastojen henkilö pohjaisiin rekisteriaineistoihin. Koska koulutuksen järjestäjät poikkeavat toisistaan muun muassa sijainniltaan, koulutusalarakenteeltaan ja opiskelijaominaisuuksiltaan, niin nämä taustatekijät otetaan huomioon mallintamalla opiskelijoiden sijoittumisen todennäköisyyttä tulosryhmiin lineaarisella regressiomallilla. Vuoden 2012 arvioinnissa vaikuttavuusmittarin tiedot koskevat vuosia 2006–2009. Koulutuksen läpäisseistä työllistyneet saavat suurimman painon 10, kun jatko-opiskelijat saavat painon 6 ja ryhmä ei-työlliset eikä jatko-opiskelijat saavat painon 3. Opinnot keskeyttäneistä työssä olevat tai muualla opiskelevat saavat painon 1 ja näiden luokkien ulkopuolella olevat saavat painon 0.

Henkilöstönäkökulmaa kuvaavat kelpoisuusmittari (7 prosenttia) ja kehittämismittari (3 prosenttia). Kelpoisuusmittari kuvaa, kuinka suuri osuus koko opetushenkilökunnasta on muodollisesti pätevää, eli täyttää opetustoimen henkilöstön kelpoisuudesta annetun asetuksen vaatimukset. Kehittämismittari seuraa koulutuksen järjestäjien henkilöstön kehittämismenojen suhdetta kaikkiin henkilöstömenoihin. Sen tiedot vuonna 2012 saadaan vuoden 2010 kustannustiedoista. Kelpoisuustiedot kerätään koulutuksen järjestäjiltä ja nekin vastaavat vuoden 2010 tilannetta.

Yhdistämällä edellä esiteltyjen mittareiden arvot saadaan tulosrahoituksen myöntämisen perusteena käytettävä tulosindeksi. Mittareiden arvojen standardoimisen jälkeen saadut arvot yhdistetään kertomalla kukin mittariarvo ensin sille asetetulla painoarvolla, ja laskemalla saadut tulot yhteen. Koska indeksien summa ei välttämättä noudata normaali jakaumaa, saatu mittari standardoidaan vielä painojen neliöiden summan neliöjuurella.

Mittariston muokkaaminen muun muassa Virtasen (2006) tekemän arvion perusteella on korjannut siinä ilmenneitä puutteita. Vaikuttavuuden suuri paino mittarissa tuntuu perustellulta, sillä sen avulla koulutuksen järjestäjille on mahdollista luoda kannuste muun muassa aloituspaikkojen kohdentamiseen aloille ja alueille, joilla työvoimalle on kysyntää. Työllistyneiden osuus vaikuttavuusmittarissa on 50 prosenttia, jatko-opiskelijoiden 30 prosenttia ja muiden 20 prosenttia. Koulutuksesta ja työelämästä kokonaan pudonneiden paino on mittaristossa 0, eli käytännössä heidät poistetaan tuloslaskelmasta. Koska käytetyt luokat ovat toisensa poissulkevia, yhden ryhmän pudottaminen mittarista on perusteltua. Koulutuksen järjestäjillä, joilla keskeyttäneiden osuus on suuri täytyy muiden luokkien summan olla pieni, eli keskeyttäneet ovat tuloslaskelmassa epäsuorasti mukana.

Vaikuttavuusmittariin liittyvä tärkein puute on se, että esimerkiksi vuonna 2012 sen tiedot kootaan vuosina 2006–2009 valmistuneista ja heidän työmarkkinatilanteestaan vuoden 2009 viimeiseltä *viikolta*. Pidemmän tarkastelujakson käyttämisen etuna on, että usein työllistyminen vie aikaa ja useamman kohortin tarkastelu vähentää vuosittaista vaihtelua. Mittarin varjopuoli kuitenkin on, että se pääsääntöisesti muuttuu hitaasti ja pitkäköllä viiveellä. Kannusteiden näkökulmasta pitkä viive ja useiden kohorttien tarkastelu aiheuttaa sen, että vaikka koulutuksen järjestäjä korjaisikin toimintaansa parempaan suuntaan, näkyvät tulokset tässä mittarissa vasta vuosien päästä. Samalla kerran palkittu koulutuksen järjestäjä voi ennakoida helposti, että se tulee todennäköisesti palkituksi seuraavanakin vuonna, onhan sen vaikuttavuusmittarista käytännössä vain yksi havainto korvattu uudella ja muuten samojen henkilöiden työllisyystilanne vuoden 2010 lopulla selvitetty. Mittari muuttuu äkillisesti vain, jos jotakin aluetta kohtaa odottamaton taloudellinen shokki, jonka seurauksena iso osa koulutuksen järjestäjän entisistä oppilaista joutuu yllättäen työttömäksi.²⁸ Negatiivisen shokin vaikutus näkyy mittarissa niin kauan, että työttömäksi joutuneet koulusta jo vuosia sitten valmistuneet pääsevät uusin töihin tai hakeutuvat jatko-opintoihin.

Vastaava kritiikki voidaan yleistää koskemaan myös henkilöstönäkökulman mittareita: vaikka aineisto on tuoreempi ja vain yhdeltä vuodelta, muutokset henkilöstössä eivät ole järin nopeita. Näiden ollen tänä vuonna palkittava koulutuksen

²⁸ Esim. Salon tai Metson massiiviset irtisanomiset tulevat luultavasti näkyviin paikallisen ammattikoulun tuloksissa muutaman vuoden kuluttua, kun sen viime vuosina valmistuneet nuoret eivät ehkä enää olekaan tämän vuoden lopussa töissä. Mahdolliset seutukuntadummyt eivät tätä vaikutusta välttämättä korjaa.

järjestäjä voi turvallisimmielinen odottaa palkintoa ensi vuonnakin vaikka ei muuttaisi toimintaansa. Parannuksia tästä vuodesta alkaen tekevä koulutuksen järjestäjä puolestaan voi odottaa palkintoa vasta noin viiden vuoden päästä, jolloin vaikuttavuusmittarin tiedot päivittyvät 2010-luvulle. On mahdollista, että palkinto siintää liian kaukana tulevaisuudessa, jolloin sen kannustevaikutus vähenee.

Kolmas mittaristoon liittyvä ongelma on osin tekninen. Vaikuttavuusmittarin arvo lasketaan estimoimalla kullekin koulutuksen järjestäjälle nk. kiinteiden vaikutusten regressiomalli, jossa valmistuneiden sijoittumista työelämään on ennustettu muun muassa häneen ja kouluun liittyvillä taustatekijöillä Virtasen (2006) esimerkin mukaan. Näin koulutuksen järjestäjän tuottama arvonlisäys voidaan erottaa siitä johtumattomista tekijöistä. Vaikka menetelmä on aiemmin esitellyn mallintamisfilosofian mukainen ja sen muuttajat on valittu huolella, mallin riski on, että koulutuksen järjestäjiä palkitaan silloinkin, kun järjestäjien välillä ei ole eroja. Jos järjestäjien väliset erot ovat hyvin pienet, tai ne eivät ole tilastollisesti merkitseviä, päädytään palkitsemaan koulutuksen järjestäjiä, jotka eivät ole lainkaan muita parempia. Koska rahoitusta laskettaessa kaikilta koulutuksen järjestäjiltä otetaan ensiksi kolme prosenttia niiden yksikköhinnasta, ja näin saatu summa jaetaan uudelleen ”voittajille”, järjestelmä voi päätyä rankaisemaan ja palkitsemaan koulutuksen järjestäjiä satunnaisesti.

Sen enempää Opetushallitus (2011) kuin Virtanenkaan (2006) eivät raportoi estimoidun mallin tilastollisia tunnuslukuja, joten mallin sopivuutta koulutuksen vaikuttavuuden tarkasteluun ulkopuolinen ei voi arvioida. On mahdollista, että käytetty malli ei selitä oppilaiden sijoittumista erityisen hyvin tai tuottaa harhaisen tuloksen arvioidulle koulutuksen järjestäjän vaikutukselle. Läpinäkyvyyden puute on vastoin hyvän järjestelmän tunnuspiirteitä. Lisäksi, kuten myöhemmin esitettävästä lukioesimerkistä käy ilmi, koulujen väliset erot voivat olla hyvin pieniä aivan yksinkertaisessakin arvonlisäysmallissa. Tämä tarkoittaa sitä, että koulutuksen järjestäjät voi asettaa luotettavasti paremmuusjärjestykseen mahdollisesti vain aivan jakauman ääripäissä. Virtanen (2006) hahmottaa osin paremmuusjärjestyksen satunnaisuutta raportin liitteessä 7, jossa esitetään koulun estimoidun vaikutuksen luottamusvälit. Koska luottamusvälit ovat tyypillisesti suuret, ei voida pois sulkea sitä mahdollisuutta, että toiseksi paras koulutuksen järjestäjä olisi ehkä paras tai vasta kolmanneksi paras. Erityisesti ei voida pois sulkea sitä mahdollisuutta, että koulutuksen järjestäjä, joka palkitaan viimeisenä, ei olekaan parempi kuin se koulutuksen järjestäjä, joka ensimmäisenä jää ilman palkintoa. Toisin kuin Opetushallitus (2011) oppaassaan ilmoittaa, mittaria ei *lasketa* vaan se *estimoidaan* ja estimointituloksiin liittyy aina virheen mahdollisuus.

Tuloksellisuusrahoituksen tärkein ominaisuus liittyy valtion koulutuksen järjestäjille myöntämän rahoituksen uudelleenjakoon. Tuloksellisuusrahoituksen avulla valtio voi kohdentaa rahoitustaan enemmän niille, jotka tekevät ”oikeaksi katsottuja” asioita ja hyväksi arvioitua tulosta. Valtion koulutukselle asettamat tavoit-

teet ovat selvästi ääneen lausutut ja koulutuksen järjestäjiä, jotka näihin tavoitteisiin parhaiten pääsevät palkitaan. Näin on luotu ohjausjärjestelmä, joka porkkanan avulla ohjaa koulutuksen järjestäjiä itse sopeuttamaan toimintaansa haluttuun suuntaan. Aika näyttää, saadaanko tällä halutut muutokset aikaan vai toteutuuko edellä hahmotellut riskit.

7.2.2 Mittariston soveltuvuus lukioihin

Selvin ero ammatillisen koulutuksen tulosrahoitusmittariston ja aiemmin esiteltyjen arviointimittaristojen välillä on se, ettei tämä mittaristo sisällä oppilaiden koetuloksia kuvaavaa komponenttia. Tämä johtunee osin ammatillisen koulutuksen käytännönläheisestä luonteesta: opiskelijoita arvioidaan perinteisten kokeiden lisäksi ns. ammattiosaamisen näytöillä, joilla opiskelija osoittaa käytännössä, missä määrin hän on saavuttanut tulevissa työtehtävissään vaaditun ammattitaidon. Ammatilliset perustutkinnot voidaan suorittaa myös näyttötutkintona tai opiskelijoiden opetussuunnitelmiin perustuvana koulutuksena.²⁹ Vaikka ammatilliset perustutkinnot arvioidaankin arvosanoin, suorituskäytännöt poikkeavat selvästi lukioiden standardoiduista ylioppilaskokeista. Opiskelijoiden menestyksen mittaaminen työllistymisellä ja jatkokoulutuksella kuvanneekin ammatillisen koulutuksen tavoitteita paremmin kuin pelkin arvosanoin mitattava koulumenestys. Huomionarvoista on myös, että työllistymistä painotetaan mittaristossa selvästi enemmän kuin jatko-opintoja, mikä viittaa siihen, että ammatillisen koulutuksen ensisijaisena tavoitteena on saada opiskelijat työelämään heti opintojen päätyttyä.

Mahdollisessa lukioiden arviointimittaristossa vastaavanlaisten komponenttien painotukset varmasti kääntyisivät päällelleen: lukioiden tarjoama yleissivistävä koulutus valmistaa opiskelijat ennen kaikkea jatko-opintoihin, eikä juuri anna ammatillisia valmiuksia menestyä työelämässä. Voisikin ajatella, että lukioiden mittaristoon sisältyisi työllistymis- ja jatkokoulutusmittarin sijaan esimerkiksi ”yliopistomittari” ja ”ammattikorkeakoulumittari”, jotka mittaisivat yliopistoihin ja ammattikorkeakouluihin siirtyneiden osuutta ylioppilaiksi valmistuneista opiskelijoista. Lisäksi lukioiden mittariston osana olisi selvästi luontevampaa käyttää standardoitujen ylioppilaskokeiden tuloksia tai niistä rakennettua arvonlisämittaria, kuten muissa maissa sovelletuissa palkitsemisohjelmissa on tehty.

Ammatillisen koulutuksen mittaristossa on käytetty moniin amerikkalaisiin mittaristoihin sisältyvää pudokkaiden eli tutkintonsa keskeyttäneiden osuutta. Mittari onkin ammatillisen koulutuksen arvioinnissa varsin relevantti, sillä Tilastokeskuksen mukaan ammatillisen koulutuksen keskeyttäneiden osuus kasvoi lukuvuonna 2009–2010 9,1 prosenttiin edellisen lukuvuoden yli 8 prosentista. Lukiokoulutuksessa vastaava luku lukuvuonna 2009–2010 oli vain 4 prosenttia, eikä se ollut kasvanut edellisestä lukuvuodesta. Kun em. luvuista vähennetään

²⁹ http://www.minedu.fi/OPM/Koulutus/amatillinen_koulutus/opiskelu_ja_tutkinnot/

koulutussektoria vaihtaneiden osuus, ammatillisen koulutuksen kokonaan keskeyttäneiden osuus putoaa 8,1 prosenttiin lukuvuonna 2009–2010, kun taas lukiokoulutuksessa vastaava luku on vain 2,0 prosenttia.³⁰ Näiden tilastojen perusteella lukioiden arviointimittaristoon ei ehkä olisi tarvetta lisätä lukion keskeyttämistä kuvaavaa komponenttia. Sen sijaan voisi pohtia, olisiko ylioppilastutkinnon hylätyksi suorittaneiden osuutta syytä ottaa huomioon lukioiden mittaristossa: ylioppilastutkintolautakunnan mukaan 5,8 prosenttia ylioppilaskokelaista hylättiin kevään 2010 ylioppilastutkinnossa. Lisäksi maakuntien hylkäysprosenttien välillä on selviä eroja: pienimmillään hylkäysprosentti keväällä 2010 oli 2,3 prosenttia Ahvenanmaalla, kun taas suurimman hylkäysprosentin Lapissa hylättyjen osuus oli 8,0 prosenttia. Syksyisin hylkäysprosentit ovat kaikkialla selvästi suuremmat, mutta pienten kokelasmäärien vuoksi satunnaisuudella on varsin suuri vaikutus hylättyjen osuuksiin.³¹

Henkilöstön kehittymistä kuvaavat mittarit, erityisesti kelpoisuusmittari, puoltavat paikkaansa ammatillisen koulutuksen tulosrahoitusmittaristossa, sillä 1.8.2011 voimaan astuneet ammatillisen koulutuksen opettajien kelpoisuusvaatimukset väljentyivät, jotta ne sopisivat paremmin työelämän muutoksia ja mahdollistaisivat alojen parhaiden asiantuntijoiden rekrytoimisen. Vaikka vaatimus pedagogisista opinnoista ja vähintään kolmen vuoden käytännön työkokemuksesta säilyivät muutoksessa entisellään, soveltuvaa korkeakoulututkintoa koskevasta vaatimuksesta on mahdollista poiketa, jos opetustehtävän edellyttämää soveltuvaa korkeakoulututkintoa ei ole olemassa tai jos ”opetustehtävä edellyttää erityisen vahvaa tai erikoistunutta käytännön ammattitaitoa”. Lisäksi riittää, että opettaja suorittaa pedagogiset opintonsa kolmen vuoden kuluessa palvelussuhteen alkamisesta. (Valtioneuvosto, 2010)

Lukioissa aineenopettajien kelpoisuusvaatimuksiin kuuluvat ylempi korkeakoulututkinto, vähintään 60 opintopisteen laajuiset opettajan pedagogiset opinnot ja vähintään 120 opintopisteen aineenopettajan koulutukseen kuuluvat opetettavan aineen opinnot. Henkilöä, joka näitä vaatimuksia ei täytä, ei voida palkata yli kuuden kuukauden mittaiseen palvelussuhteeseen ilman, että ensin on pyritty löytämään kelpoisuusvaatimukset täyttävää henkilöä tehtävään. (Valtioneuvosto, 1998) Vaatimukset ovat siis lukiokoulutuksessa selvästi tiukemmat kuin ammatillisessa koulutuksessa. Luultavasti osittain tästä johtuen lukiokoulutuksen ja ammatillisen koulutuksen välillä on melkoinen ero muodollisesti pätevien opettajien osuudessa opetushenkilöstöstä, mikä käy ilmi Opettajat Suomessa 2010 -selvityksestä: lukioissa 93,1 prosenttia rehtoreista ja kaikista opettajista oli muodollisesti pätevää toimimaan hoitamassaan tehtävässä mutta ammatillisessa koulutuksessa vastaava osuus oli 73 prosenttia. Ammatillisen koulutuksen opettajien kelpoisuuden vähäisyys johtui erityisesti pedagogisten opintojen puuttumisesta (Kumpulainen, 2010).

³⁰ http://www.stat.fi/til/kkesk/2010/kkesk_2010_2012-03-20_tie_001_fi.html

³¹ http://www.ylioppilastutkinto.fi/Tilastoja/tilastojulkaisu2010/Ylioppilastutkinto_2010.pdf

Niinpä lukioiden tulosmittaristoon ei välttämättä tarvitsisi sisällyttää opettajien kelpoisuutta tarkastelevaa mittaria, kun taas ammatillisessa koulutuksessa pätevien opettajien osuuden seuraaminen on hyödyllinen koulutuksen laadun mittari. Mutta, kuten Virtanenkin (2006) huomauttaa, opettajan ja opetuksen laatua ei voida taata pelkästään opettajien kelpoisuutta seuraamalla, vaan myös heidän muut ominaisuutensa ja osaamisensa olisi syytä ottaa huomioon kattavammassa mittarissa. Opettajien kelpoisuus on kuitenkin helposti mitattavissa, toisin kuin esimerkiksi opettajien motivoituneisuus tai innostavuus, joiden selvittämiseksi tarvittaisiin esimerkiksi kyselytutkimus.

Kehittämismittarin pyrkimys kuvata henkilöstön kehittämistoiminnan laajuutta on sinänsä arvokas, joten teoriassa mittari on perusteltavissa osaksi tulosrahoitusmittaristoa niin ammatillisessa kuin lukiokoulutuksessa. Mittarin luotettavuuteen liittyvät ongelmat sekä sen puutteellinen kyky kuvata kehittämistoiminnan määrää ja laatua kuitenkin osoittavat, ettei mittari ole käytännössä riittävän osuva ja kattava kuvaamaan henkilöstön kehittämistoimintaa.

7.3 Lukioiden palkitsemiseen soveltuvan aineiston läpikäyminen

Olemme tähän osaan tehneet yksinkertaisia koelaskelmia siitä, mistä lukioiden väliset vaikutuserot muodostuvat ja miltä lukioiden paremmuusjärjestys voisi näyttää. Aineistona on käytetty yksilötason tietoja vuosien 2005–2007 ylioppilastutkinnon tuloksista sekä vuosien 1996–2004 yhteisvalintarekisterin tietoja. Tavoitteena on tehdä mahdollisimman yksinkertaisia arvonlisäykseen perustuvia vertailuja, joissa oppilaiden taustasta on kontrolloitu ainoastaan lähtötaso, peruskoulun päättötodistuksen lukuaineiden keskiarvolla mitaten, mutta ei perhe-taustaa, sukupuolta, äidinkieltä, koulun resursseja tai mitään muuta ylioppilaskirjoitusten tuloksiin mahdollisesti vaikuttavaa seikkaa. Vertailun tavoite on selvittää, kuinka paljon lukioiden väliltä on mahdollista löytää eroja. Ovatko koulujen väliset erot niin merkittävät, että laadun parannuksista palkitsevalle tuloksellisuusrahoitukselle on tarvetta?

Kuusela (2003) on tehnyt osin vastaavaa analyysiä, mutta hänellä ei ole ollut käytössään oppilaskohtaisia tietoja, mikä hävittää aineistosta tärkeimmät vaihtelun lähteet: koulujen sisäisen ja opiskelijoiden välisen vaihtelun. Toisaalta, koska aineistomme on yksilötasolla, näiden ohella ei ole mielekäästä käyttää koulukohtaisia keskiarvotietoja, kuten jatko-opintotietoja tai tietoa keskeyttäneistä. Koulujen mahdollisista tuotoksista tulee tässä tarkastelussa huomioon otetuksi ainoastaan ylioppilaskirjoitusten tulokset, joiden pitäisi olla merkittävä osa mahdollista tuloksellisuusmittaria.

Aineisto sisältää 102912 havaintoa vuosien 2005–2007 kevään ja syksyn ylioppilaskirjoituksiin osallistuneesta kokelaasta yhteensä 475 eri koulusta. Kouluissa on mukana aikuislukiot, iltalukiot, kansanopistot ja kaikki koulut, joissa voi suorittaa ylioppilastutkinnon. Tavoite on pitää aineistossa mahdollisimman hetero-

geeni joukko oppilaitoksia. Aineistoa on rajattu niin, että jos jollakin kirjoituskeralla ylioppilastutkinnon suorittaneita kokelaita on koulussa ollut vain yksi, kyseinen havainto on pudotettu aineistosta pois.³² Otoksen kokoon tehty rajausta ei vaikuta. Kuitenkin erityisesti syksyn kirjoituksissa kaikki koulut eivät ole edustettuina vaan syksyltä 2005 aineistosta putoaa 15 koulua, syksyltä 2006 kahdeksan koulua, syksyltä 2007 26 koulua sekä keväältä 2006 yksi koulu. Kaiken kaikkiaan tehty rajausta koskee 41 oppilaitosta, joista suurin osa pudotetaan pois vain kerran. Tiukempiakin koulun kokoon liittyviä rajauksia olisi perustellusti voitu tehdä, mutta halusimme pitää mukana mahdollisimman paljon informaatiota. Lisäksi tarkastelussa ovat vain ne kokelaat, joille yhteisvalinta-aineistosta on hakutieto, ja siten peruskoulun päättötodistuksen lukuaineiden keskiarvotieto joltakin kirjoituksista edeltävien vuosien yhteishausta. Lopullisessa aineistossa on 101660 ylioppilaskokelasta 475:stä lukiokoulutusta tarjoavasta oppilaitoksesta.

7.3.1 Koulun vaikutus oppimistuloksiin

Analyysi rakentuu seuraavasti. Aineiston avulla on ensiksi estimoitu nk. naiivi malli, jossa kouluja verrataan toisiinsa pelkkien *yksilökohtaisten* oppimistulosten perusteella. Sen jälkeen on estimoitu *yksilökohtainen* arvolisämalli, jossa ylioppilaskirjoitusten puoltoääniä selitetään peruskoulun päättötodistuksen lukuaineiden keskiarvolla. Mittareissa on puutteita, mutta jäljempänä esitetyt tulokset osoittavat, että ne ovat riittävät tarkastelumme kannalta.

Arvonlisäykseen perustuvat tarkastelut on tehty kahdesta näkökulmasta: oppilaskohtaisesta ja koulukohtaisesta näkökulmasta. Koulukohtainen näkökulma tarkoittaa vertailua, jossa kouluja, joilla on *keskenään samankaltainen oppilaskoukko*, verrataan toisiinsa. Tällainen vertailu on luvussa 2.2.3 esitelty kontekstuaalinen arvonlisävertailu, jossa oppilaiden samankaltaisuutta arvioidaan oppilaiden taustaominaisuuksien eli sosiodemografisten ominaisuuksien perusteella. Meidän analyysissämme oppilaiden samankaltaisuutta on arvioitu peruskoulun päättötodistuksen keskiarvon perusteella. Oppilaan näkökulmasta tehty vertailu perustuu vastauksen etsimiseen seuraavaan kysymykseen: Jos tiedetään oppilaan peruskoulun päättötodistuksen lukuaineiden keskiarvo (jäljempänä pelkkä keskiarvo), niin kuinka hyvin oppilaan valitsema lukio suoriutuu verrattuna muihin lukioihin, joihin kyseisellä päättötodistuksella olisi voinut (periaatteessa) päästä? Eli olisiko kyseinen nuori voinut kirjoittaa *paremmin* jossakin toisessa lukiossa. Kontekstuaalinen vertailu lienee kiinnostavaa palkitsemisen näkökulmasta, mutta oppilaskohtainen vertailu on kiinnostavaa koulua valitsevalle nuorelle.

³² Tutkinto on aineistossa kirjattu suoritetuksi silloin, kun kokelas on suorittanut viimeisen hyväksytyyn kokeen, riippumatta siitä, onko hän jo sitä ennen saanut tutkinnon. Valittu tapa helpotti aineiston käsittelyä, joskin se ajoittaa suorituksia väärin ajankohtiin. Koska valitun määritelmän vaikutuksen pitäisi kohdentua kaikkiin kouluihin samalla tavalla ja se koskee vain pientä osaa aineistostamme, valitun kirjaamistavan ei pitäisi vaikuttaa saatuihin tuloksiin.

Yksinkertaisessa paneeliregressiossa puoltoäänten keskiarvoa selitetään pelkästään vakiolla ja aikadummyillä kiinteiden vaikutusten mallissa.³³ Tässä mallissa koulujen vaikutusta arvioidaan pelkästään oppimistulosten perusteella ilman, että oppilaiden aiempaa osaamista on kontrolloitu millään lailla. Kiinteät vaikutukset estimoidaan koulukohtaisina ja ne sisältävät kaiken ajan suhteen vakioisen koulukohtaisen vaihtelun. Tähän termiin sisältyvät esimerkiksi koulun ”laatu”, johon näin yksinkertaisessa mallissa sisältyvät oppilaan kaikki taustatiedot, vertaisryhmätekijä, opettajien ammattitaito ja muut käytetyt resurssit, jotka ovat koululle kutakuinkin vakiot ja joita emme kontrolloi. Mallin keskivirheet on laskettu siten, että niissä on otettu huomioon oppilaiden klusteroituminen eri kouluihin. Aikadummyjä ei yksinkertaisuuden vuoksi raportoida, mutta ne ovat kaikissa mallispesifikaatioissa tilastollisesti merkitseviä. Tämän nk. naiivin mallin keskeiset tulokset on esitetty Taulukossa 1 sen alimmilla riveillä.

Koulujen välisten vaikuttavuuserojen hahmottamisen kannalta keskeinen indikaattori on mallin *rho*, joka suhteuttaa mallin virhetermin ja koulujen kiinteiden vaikutusten varianssin toisiinsa.³⁴ Rho on sitä suurempi, mitä suurempi koulujen välinen (laadun) vaihtelu ja mitä pienempää on koulujen sisäinen vaihtelu. Suuri rho on merkki suurista koulujen välisistä eroista. Vaikka naiivissa mallissa ei oppilaiden suoriutumista ylioppilastutkintokokeessa selitetä millään muuttujalla, sen rho on sängen pieni. Vain 12 prosenttia varianssista johtuu koulujen välisistä eroista ja loppu aiheutuu koulujen sisäisistä eroista eli oppilaiden heterogeenisyydestä. Mallin selitysaste niin koulujen väliselle kuin sisäiselle vaihtelulle on erittäin matala, R^2 on alle 0,10. Suurin osa oppimistulosten vaihtelusta selittyy mallin ulkopuolisilla tekijöillä. Saatu tulos koulujen välisistä oppimistulosten eroista ovat samansuuntaiset kuin muun muassa Kane ja Staiger (2002) raportoimat tulokset.

Seuraavaksi on estimoitu perusmalli, jossa kunkin ylioppilaan kirjoitusten puoltoäänten keskiarvoa selitetään peruskoulun päättötodistuksensa lukuaineiden keskiarvolla, eli oppilaan aiempi osaaminen on otettu huomioon arvonlisäysmallin edellyttämällä tavalla. Tämä yksinkertainen malli selittää hyvin koulujen välistä vaihtelua, sillä ryhmien välinen R^2 on 0,65. Koulujen sisäinenkin vaihtelu selittyy kohtalaisesti, kun otetaan huomioon, että mallissa ei ole lainkaan oppilaan sosiodemografisia kontrollimuuttujia. Mallin virhetermin, joka sisältää kiinteän vaikutuksen ja satunnaisen komponentin, varianssista vain 8 prosenttia johtuu koulujen *välisistä* eroista. Toisin sanoen, koulujen ”laadun” sisältävän vaihtelun osuus kaikesta vaihtelusta on perusmallissa vain 8 prosenttia. Vaikka malli ei ota huomioon koulujen resurssieroja, niin siitä huolimatta koulujen välillä näyttäisi olevan vain vähäisiä eroja. Koulujen väliset vaikuttavuuserot ovat

³³ Aikadummit kontrolloivat eri kirjoituskertoihin liittyvää vaihtelua.

³⁴ Juuri tämä tilastollinen tunnusluku Virtasen (2006) ja Opetushallituksen (2011) pitäisi vähintään raportoida. Virtanen (2006) raportoi estimoitujen kiinteiden vaikutusten luottamusvälit, mikä myös kertoo vertailun epävarmuudesta.

hyvin pienet ja kun kontrolloidaan oppilaan aiempi osaaminen, niin koulujen välisistä eroista häviää noin kolmannes.

Taulukko 1 Yhdistetyn naiivin mallin ja perusmallin tulokset $N_1=101660$ ja $N_2=475$

Ylioppilaskirjoitusten puoltoäänien keskiarvo						
PERUSMALLI						
Peruskoulun päättötodistus		Vakio		R ²		
Regressio-kerroin	Keski- virhe	Regressio-kerroin	Keski- virhe	ryhmien sisäinen	ryhmien välinen	rho
0,00895	0,00009	-3,1458	0,07455	0,429	0,652	0,079
NAIIVI MALLI						
		4,3273	0,00653	0,032	0,067	0,121

Tulosten käytettävyyden kannalta on tärkeää arvioida kouluvaikutuksen luottamusvälejä. Lukiodien estimoidulle vaikutukselle lasketut luottamusvälit ovat aineistoissamme hyvin kapeat. Koulun vaikutuksissa havaittavat erot ovat kuitenkin niin pienet, että useimmat koulun vaikutuksen suhteen samalle tasolle asettuvat koulut eivät poikkea toisistaan tilastollisesti merkitsevästi. Näyttää siltä, että koulut poikkeavat merkitsevästi toisistaan vain jakauman ääripäissä, joissa 8-9 ”huonointa” koulua ja 6-7 ”parasta” koulua erottuvat kaikista muista kouluista. Näiden väliin jäävät noin 460 koulua eivät välttämättä poikkea tilastollisesti merkitsevästi vastaavalle tasolle arvioiduista kouluista. Käytännössä koulujen väliset erot ovat niin pienet, että koulujen asettaminen niiden estimoitujen vaikutusten suhteen paremmuusjärjestykseen voi osoittautua hankalaksi.³⁵

Perusmalli toimii sängen hyvin, vaikka siitä puuttuvat lähes kaikki tavallisesti selittävinä tekijöinä käytetyt muuttujat (vanhempien koulutus tai ammatti, käytetyt resurssi jne.).³⁶ Peruskoulun päättötodistus ennakoii hyvin ylioppilaskirjoitusten tuloksia: yhden arvosanan parannus päättötodistuksessa, ts. 100 pisteen parannus keskiarvossa ennakoii noin 0,9 arvosanan parannusta ylioppilaskokeen puoltoäänien keskiarvosanassa. Malli siis ennakoii, että peruskoulunsa kahdeksikon arvosanalla päättänyt oppilas kirjoittaa keskimäärin 4,5 puoltoäänien arvoisesti, mikä vastaisi esimerkiksi kolmea magna ja kolmea cum laudea.

³⁵ Vaikka käytetty aineisto poikkeaa huomattavasti Kuuselan (2003) aineistosta, tulokset ovat hyvin samansuuntaisia Kuuselan raportointien tulosten kanssa. Siinä missä Kuuselalla äidin koulutus selitti noin puolet ylioppilaskirjoitusten tulosten variaatiosta, meillä peruskoulun päättötodistus selittää noin kolmannes variaatiosta. Luultavasti vanhempien koulutustaso ja lapsen peruskoulun päättötodistuksen keskiarvo ovat korreloituneet keskenään ja sisältävät osin samaa informaatiota.

³⁶ Jos malliin lisätään oppilaan sukupuoli, äidinkieli ja koulun koko, niin koulun kokoa lukuun ottamatta nämä muuttujat ovat joissakin spesifikaatioissa tilastollisesti merkitseviä, mutta mallin selitysvoima ei juuri parane, eivätkä nämä juuri vaikuta koulujen välisiin eroihin. Lisäksi tulosten vakautta on tarkasteltu estimoidulla mallilla pelkästään päivälukioille tai kontrolloimalla omistussuhteita, mutta tulokset eivät muuttuneet lainkaan.

Seuraavaksi tarkastelemme koulujen välisiä eroja yksittäisen oppilaan näkökulmasta. Olemme jakaneet ylioppilaat ryhmiin peruskoulun päättötodistuksensa lukuaineiden keskiarvon mukaan. Jaottelu perustuu siihen, että huomattava osa kouluista valitsee oppilaansa päättötodistusten mukaan. Jos oppilas voisi valita täysin vapaasti haluamansa lukion, voisi hyvillä arvosanoilla peruskoulunsa Sonkajärvellä päättänyt oppilas tulla valituksi johonkin pääkaupunkiseudun ”hyvään” lukioon. Ryhmiin jakaminen on tehty klusterointimenetelmällä minimoimalla kunkin ryhmän varianssi, eli rakentamalla mahdollisimman homogeeninen luokittelu. Ryhmien lukumääräksi asetettiin ennalta seitsemän, mutta myös viittä, kuutta ja kahdeksaa ryhmää kokeiltiin tehdyssä herkkyyksianalyysissä.³⁷ Hypotesimme on, että peruskoulun päättötodistuksen ja ylioppilaskirjoitusten tulosten välinen suhde voi olla epälineaarinen. Lisäksi eri koulut voivat tuottaa erilaisen arvonlisäyksen taustaltaan erilaisille oppilaille.

Taulukossa 2 on esitetty ryhmien tilastollisia tunnuslukuja. Ensimmäiseen ryhmään kuuluvat ne ylioppilaat, joiden peruskoulun päättötodistuksen lukuaineiden keskiarvo oli 5,0:n ja 7,3:n välillä. Tähän ryhmään kuului yli kymmentuhatta ylioppilasta 456 koulussa, eli lähes kaikista Suomen lukiokoulutusta antavista oppilaitoksista löytyy oppilaita, joiden keskiarvo on ollut alle 7,30. Tämän ryhmän keskiarvo on 7,0 ja keskihajonta 0,32, eli matalat keskiarvot ovat harvinaisia. Ryhmän sisäinen vaihtelu on luokittelusta huolimatta edelleen suuri. Muut ryhmät ovat huomattavasti homogeenisempia ja arvosanalokan pituus niissä on noin puoli arvosanaa ja keskihajonta on siksi vähäistä.

Taulukko 2 Ryhmittelyn tulokset samankaltaisille oppilasryhmille

Ryhmä	N ₁	N ₂	Minimi	Maksimi	Keskiarvo	Keskihajonta
1	10873	456	5,00	7,30	6,99	0,32
2	17830	465	7,33	7,83	7,62	0,15
3	17781	473	7,86	8,23	8,05	0,11
4	18015	473	8,25	8,60	8,43	0,11
5	14733	471	8,63	8,92	8,79	0,09
6	14203	469	9,00	9,38	9,14	0,12
7	8185	459	9,40	10,00	9,65	0,18

Seuraavaksi perusmalli estimoidaan uudelleen peruskoulun päättötodistuksen suhteen homogeenisille oppilasryhmille. Emme raportoi kaikkia mallin tuloksia, vaan keskitymme erityisesti mallidiagnostiikkaan eli R^2 :n ja rho:n. Estimointitulokset kullekin ryhmälle on raportoitu Taulukossa 3.

³⁷ Ryhmittely on jossain määrin add hoc, koska sille ei löydy objektiivisia perusteita. Seitsemään ryhmään päädyttiin, koska saadut tulokset osoittavat, että joidenkin joskaan ei kaikkien, ryhmien välillä parametriestimaateissa on eroja. Käytännön kannalta parametriestimaattien erot eivät ole erityisen merkitykselliset.

Peruskoulun päättötodistuksen lukuaineiden keskiarvo ennustaa tilastollisesti merkitsevästi kaikissa ryhmissä ylioppilaskirjoitusten tuloksia, joskin vaikutus näyttäisi kasvavan keskiarvoluokittain ja on suurin parhaiden oppilaiden ryhmässä. Parhaiden oppilaiden päättötodistuksen vaikutus ylioppilaskirjoitusten puoltoääniin on noin viisinkertainen heikoimpiin verrattuna, ja kerrointen ero on tilastollisesti merkitsevä. Peruskoulun päättötodistuksen vaikutus ylioppilaskirjoitusten tuloksiin on siten heterogeeninen erityisesti ääripäissä: ryhmä 1 poikkeaa kaikista muista, ja samoin ryhmät 6 ja 7, vaikka ovatkin keskenään samanlaiset.

Taulukko 3 Estimointitulokset ryhmittäin samankaltaisille oppilasryhmille

Ryhmä	Ylioppilaskirjoitusten puoltoäänten keskiarvo						
	Peruskoulun päättötodistus		Vakio		R ²		rho
	Regressio-kerroin	Keskivirhe	Regressio-kerroin	Keskivirhe	ryhmien sisäinen vaihtelu	ryhmien välinen vaihtelu	
1	0,00255	0,00033	1,4895	0,2305	0,027	0,046	0,163
2	0,00653	0,00036	-1,3597	0,2758	0,034	0,031	0,139
3	0,00774	0,00054	-2,3012	0,4337	0,029	0,049	0,135
4	0,00808	0,00057	-2,5327	0,4775	0,030	0,031	0,146
5	0,00871	0,00062	-2,9412	0,5487	0,029	0,025	0,156
6	0,01396	0,00054	-7,6339	0,4980	0,065	0,026	0,155
7	0,01423	0,00040	-8,0096	0,3854	0,154	0,076	0,227

Toinen huomion arvoinen seikka on, että R²:n arvoilla arvioiden malli selittää kohtalaisesti vain parhaiden oppilaiden ryhmän sisäistä tai välistä vaihtelua. Muille ryhmille yksikertainen malli on huono. On hieman hämmentävä tulos, että paloitettu malli selittää aineiston vaihtelua huonommin kuin yllä esitelty koko aineistolle estimoitu malli. Ehkä löydöstä selittää rho:n arvojen ”suuruus”. Kun koko aineistolle estimoidun mallin rho oli vain 0,08, se on nyt pienimmilläänkin lähes kaksinkertainen. Tämä johtuu siitä, että estimoimalla malli homogeenisissä ryhmissä olemme keinotekoisesti vähentäneet koulujen sisäistä vaihtelua. Siksi koulujen välisen vaihtelun osuus on suurempaa, joten vertailu perusmallin rho:hon ei ole järkevää. Sen sijaan rho:n ryhmien välinen vertailu on kiinnostavampaa, ja on mielenkiintoista, että koulujen välinen vaihtelu on erityisen suuri parhaille oppilaille.

Mitä nämä tulokset tarkoittavat? Tulkintaa voidaan hakea tarkastelemalla parhaiden oppilaiden ryhmää. Oppilas, jonka peruskoulun päättötodistuksen lukuaineiden keskiarvo on vaikkapa 9,5 voi löytyä mistä tahansa koulusta, sillä Suomessa oli vuosina 2005–2007 vain 16 lukiokoulutusta tarjoavaa koulua, jossa ei ollut yhtään tällaista oppilasta. Niissä 459 koulussa, jossa heitä oli, on merkittäviä eroja. Hyvälle oppilaille on merkitystä, mihin kouluun hän menee, sillä 23 prosenttia havaitusta vaihtelusta aiheutuu koulujen välisistä eroista. Hyvä oppilas

mahdollisesti hyötyy hyvästä opetuksesta tai siitä, että toiset koulun oppilaat ovat hyviä. Vähiten merkitystä koululla on ryhmien 2–4 oppilaille, eli 7,33–8,43 keskiarvon oppilaille.

Vaihtoehtoinen tulkinta saadulle tulokselle on, että osa hyvistä oppilaista ei itse asiassa olekaan niin hyviä, kuin heidän yläasteen opettajansa ovat arvioineet. Tällöin saadut tulokset olisivat merkki arvosanainflaatiosta eli siitä, että opettajat antaisivat oppilailleen parempia arvosanoja kuin he aidosti ansaitsisivatkaan. Vaihtoehtoisesti opettajien voi olla, varsinkin pienissä kouluissa, joissa oppilaita on vähän, vaikea arvioida oppilaiden hyvyyttä jakauman molemmissa päissä; kaikki kymppin oppilaat eivät ole hyviä kuten eivät ole kaikki vitosen oppilaatkaan huonoja.³⁸

Näyttää siltä, että oppilaalle on merkitystä millainen koulu on tai ainakin sillä, keitä koulun muut oppilaat ovat, jolloin on kiinnostavaa estimoida kontekstuaalinen arvonnäkökulma. Jaamme koulut kolmeen luokkaan sen vuosina 2005–2007 ylioppilaksi kirjoittaneiden oppilaiden peruskoulujen keskiarvojen keskiarvon mukaan. Toisin sanoen, koulun oppilaiden peruskoulun päättötodistusten keskiarvoista on laskettu keskiarvo kullekin vuodelle, joista on vielä laskettu keskiarvo. Tavoitteena on ottaa huomioon pienten koulujen hakijajoukon vaihtelu, joka voisi johtaa siihen, että koulu olisi eri vuosina eri ryhmissä. Toisaalta, koska tarkastellaan usean vuoden keskiarvoa, ryhmiä ei tarvitse rakentaa niin montaa kuin edellä, sillä eri vuosien aineiston yhdistäminen ja keskiarvon ottaminen hävittää koulusta riippumatonta, satunnaista vaihtelua. Näin ollen aineiston pilkkominen useampaan osaan ei ole perusteltua.

Taulukossa 4 on esitelty ryhmittelyn tulokset ja ryhmien tunnuslukuja. Momentit viittaavat koulukohtaisiin usean vuoden keskiarvoihin ja ryhmän aidot minimi- ja maksimit on raportoitu suluissa. Yksittäisten oppilaiden keskiarvot vaihtelevat kaikissa ryhmissä lähes koko skaalalla 5,0 ja 10,0 välillä (raportoitu suluissa) ja koulujen kaikkien oppilaiden keskiarvot vaihtelevat 7,3:n ja 9,5:n välillä. Heikoimman suomenkielistä lukiokoulutusta tarjoavan koulun oppilaiden keskiarvojen keskiarvo on ollut 7,3 ja parhaan koulun 9,5. Vain ryhmästä 3 kaikkein heikoimmat oppilaat puuttuvat kokonaan ja keskiarvojen keskihajonta on hivenen muita pienempi, mutta sielläkin matalin keskiarvo on 5,8, joten oppilaiden heterogeenisuus on kaikissa ryhmissä huomattavaa.

³⁸ Kun oppilaita on vähän, myös aivan huippuja tai heikkoja oppilaita on todennäköisesti hyvin vähän. Siksi oppilaiden todellisen, valtakunnallisen tason, arvioiminen on vaikeaa ilman valtakunnallista koetta.

Taulukko 4 Ryhmittelyn tulokset samankaltaisia oppilaita palveleville kouluille

Ryhmä	N ₁	N ₂	Minimi	Maksimi	Keskiarvo	Keskihajonta
1	17226	102	7,27 (5,00)	8,07 (10,00)	7,84	0,158
2	66491	322	8,07 (5,00)	8,57 (10,00)	8,30	0,117
3	17943	51	8,57 (5,82)	9,52 (10,00)	8,85	0,213

Suomessa on noin 102 lukiokoulutusta antavaa oppilaitosta, joiden vuosien 2005–2007 ylioppilaiden peruskoulun päättötodistusten keskiarvojen keskiarvo on ollut alle 8,07. Suosituimmat 51 koulua puolestaan valitsevat oppilaita, joiden keskiarvojen keskiarvo on vähintään 8,57. Vakioinnin jälkeenkin tässä ryhmässä on muita ryhmiä enemmän vaihtelua, joten aineiston voisi halutessaan pilkkoa useampaan osaan. Mutta jos aineisto pilkotaan viiteen ryhmään, se näyttää hävitävän lähes kaiken koulujen välisen vaihtelun, jolloin minkäänlaisia ”liigoja” ei kannata muodostaa.³⁹

Tässä tavoitteena on verrata keskenään samanlaisten koulujen tuottamaa arvonlisäystä potentiaalisesti taustoiltaan hyvinkin erilaisille oppilaille. Turnajaiskirjallisuudessa on osoitettu, että kannusteet toimivat parhaiten silloin, kun keskenään samankaltaiset kilpailevat tasaveroisesti palkinnosta. Yllä tehty ryhmittely siten simuloi yhtä tapaa jakaa koulut ryhmiin ja asettaa ne ryhmien sisällä paremmuusjärjestykseen. Taulukossa 5 on esitetty estimointitulokset ryhmittäin.

Taulukko 5 Estimointitulokset ryhmittäin samankaltaisia oppilaita palveleville kouluille

Ryhmä	Ylioppilaskirjoitusten puoltoäänien keskiarvo						
	Peruskoulun päättötodistus		Vakio		R ²		rho
	Regressio-kerroin	Keski- virhe	Regressio-kerroin	Keski- virhe	ryhmien sisäinen vaihtelu	ryhmien välinen vaihtelu	
1	0,0071	0,0002	-1,707	0,1862	0,326	0,366	0,069
2	0,0092	0,0001	-3,397	0,0493	0,468	0,183	0,075
3	0,0104	0,0002	-4,323	0,1437	0,390	0,650	0,082

Aiemmin estimoitu yhdistetty perusmalli osoittaa, että yksinkertainen arvonlisäysmalli selittää suuren osan koulujen välisestä vaihtelusta. Siksi aineiston pilkkomisesta ei ole merkittävää hyötyä. Siitä huolimatta ryhmitellyssä mallissa Wald-testin mukaan päättötodistuksen regressiokertoimet poikkeavat kaikissa

³⁹ Jos aineisto pilkotaan viiteen osaan ja tarkastellaan uutta parhaiden koulujen ryhmää lähemmin, havaitaan että nämä 21 koulua sijaitsevat Suomen suurimmissa kaupungeissa: Helsingissä, Espoossa, Oulussa, Jyväskylässä, Tampereella, Turussa ja Kuopiossa. Koulujen nimet ovat myös tutut mm. MTV3:n vuosittain tekemän kouluvertailun listan kärjestä.

ryhmissä tilastollisesti merkitsevästi toisistaan, samoin kuin vakiotkin. Ryhmitteilyn jälkeen yksinkertainen arvonlisäysmalli selittää hyvin eri tavoin ryhmien sisäistä ja välistä vaihtelua eri ryhmien kouluille: parhaat oppilaat valitsemissa kouluissa malli selittää huomattavan osan ryhmien välisistä eroista, eli oppilaiden peruskoulun päättötodistus selittää valtaosan tämän ryhmän koulujen havaituista eroista. Toisessa ryhmässä malli selittää kohtalaisesti ryhmien sisäisiä eroja: keskikastin kouluissa oppilaiden peruskoulun päättötodistusten erot selittävät eroja heidän ylioppilaskirjoitusmenestyksessään, mutta eivät niinkään koulujen välisiä eroja. Koulujen väliset erot johtuvat siis muista seikoista.

Vaikka aineiston pilkkomisesta näyttäisi Taulukon 5 tulosten mukaan olevan hyötyä, tärkeämpää on selvittää, johtaisiko yhdistetyn perusmallin käyttö vääriin johtopäätöksiin, jos kouluja palkittaisiin tai rangaistaisiin sen perustella. Sen arvioimiseksi molemmista malleista voidaan laskea kiinteiden vaikutusten termi kullekin koululle ja tuottaa niiden perusteella kaksi ”ranking”-listaa ja verrata listoja toisiinsa. Oletetaan, että kustakin ryhmästä neljä parasta koulua saisi palkinnon, ja yhdistetyssä mallissa parhaat 12 koulua palkittaisiin. Tällöin yhdistetyn mallin parhaasta 12 koulusta ryhmittelyssä parhaina tulisi palkituksi 11 koulua. Koulu, joka kolmen liigan tapauksessa tulisi palkituksi, mutta yhdistetyn mallin mukaan jäisi ilman palkintoa, on yhdistetyn mallin listalla sijalla 14. Käytännössä koulujen väliset erot ovat niin pienet, että muutaman sijan muutos ei voi olla tilastollisesti merkitsevä. Parhaat koulut erottuvat kuitenkin joukosta arvioitiinpa niitä kokonaisuutena tai ryhminä. Parhaan 12 koulun listasta havaitaan, että käytännössä yhdistetty malli ja ryhmitelty malli, eli kontekstuaalinen arvonlisäysmalli, tuottavat täsmälleen saman tuloksen. Tämä puhuu yhdistetyn mallin käyttämisen puolesta, sillä aineiston pilkkomisesta ei näyttäisi saatavan suurtaakaan hyötyä.

Jos haluttaisiin rangaista heikoimpia kouluja, ne olisi jo määritelmällisesti vaikeampi tunnistaa. Kuuluuko heikkoja kouluja löytyä kaikista kolmesta ryhmästä, vai olisivatko ne yhdistetyn aineiston heikoimmat koulut?

Yhdistetyn mallin tuottamasta 12 heikoimman koulun listalla 11 löytyy *ryhmän 2* viimeisiltä sijoilta ja yksi *ryhmän 1* toiseksi viimeiseltä sijalta. Opiskelemaan valittujen oppilaiden keskiarvosanojen perusteella heikoimmat koulut löytyvätkin ”keskikastin” kouluista, mihin mallin heikkous ryhmien välisen vaihtelun selittäjänä viittaa. Jos jokaisesta ryhmästä valittaisiin neljä koulua, joita rangaistaisiin niiden ”huonoudesta”, paras niistä löytyisi yhdistetyn mallin listalta sijalta 373 (475 koulusta). Vaikka muut 11 koulua sijoittuvatkin yhdistetyn mallin listalla 420:n heikommalle puolelle, niin silti näistä vain 5 koulua tulisi kuulumaan yhdistetyn mallin listan mukaan rangaistujen joukkoon. Samalla tavalla kuin listan huipulla myös sen pohjalla erot ovat pienet, mutta ainakin osa kouluista voisi joutua väärin perustein toimenpiteiden kohteeksi. Liigan käyttäminen ei näin ollen näytä perustellulta, sillä koulujen ryhmittelyllä ei näytetä saavuttavan merkitsevää etua. Ryhmittelyn tärkein lisäarvo lienee, että sen avulla havaitaan, että

”heikoimmat” koulut mahdollisesti sijaitsevat keskikastin koulujen ryhmässä: ne saavat varsin kohtuullisia oppilaita peruskouluista, mutta osa kouluista ei saa ”leivottua” heistä yhtä hyviä ylioppilaita kuin saman ryhmän parhaat koulut.⁴⁰

Hallitusohjelman mukaan lukioden tuloksellisuusrahoituksen tulisi olla sellainen, ”joka palkitsee koulutuksen laadusta ja laadun parantamisesta”, mikä voidaan myös tulkita arvonnalisäyksen muutokseksi. Laadun parantamisen käytännön arviointi tarkoittaisi sitä, että yllä laskettu perusmalli estimoitaisiin kullekin vuodelle erikseen ja jokaiselle koululle laskettaisiin kiinteät vaikutukset kunakin vuonna. Jos lisäksi oletetaan, että kiinteiden vaikutusten termi sisältää ainoastaan laatuinformaatiota, niin silloin kiinteiden vaikutusten termi mittaa koulun laatua. Koska ylioppilaskirjoitusten arvosanat kuitenkin noudattavat standardoitua normaalijakaumaa, niin absoluuttisia parannuksia ei havaita edes arvonnalisäyksessä, vaan ainoastaan suhteelliset muutokset. Asettamalla koulut järjestykseen kunkin vuoden kiinteän vaikutuksen mukaan saadaan joka vuosi uusi paremmuusjärjestys, jota voidaan vertailla eri vuosina.

Taulukossa 6 on esitetty koulujen sijoitusten korrelaatio eri vuosina koko aineistossa sekä parhaille kolmellekymmenelle ja heikoimmille kolmellekymmenelle koululle erikseen laskettuna. Koko aineistossa eri vuosien sijoitusten välinen korrelaatio on kohtalainen, yli 0,60. Se johtaa ajattelemaan, että koulun sijoitus pysyy suhteellisen vakaana vuodesta toiseen. Mutta kun katsotaan jakauman ääripäitä, kouluja, jotka sijoittuvat joko huipulle tai pohjalle havaitaan, että korrelaatio on heikkoa. Kuvissa 1 ja 2 hahmotellaan tarkemmin koulun sijoituksen vaihtelua vuosien 2005 ja 2007 välillä.

Taulukko 6 Koulujen eri vuosien sijoitusten järjestyskorrelaatiot

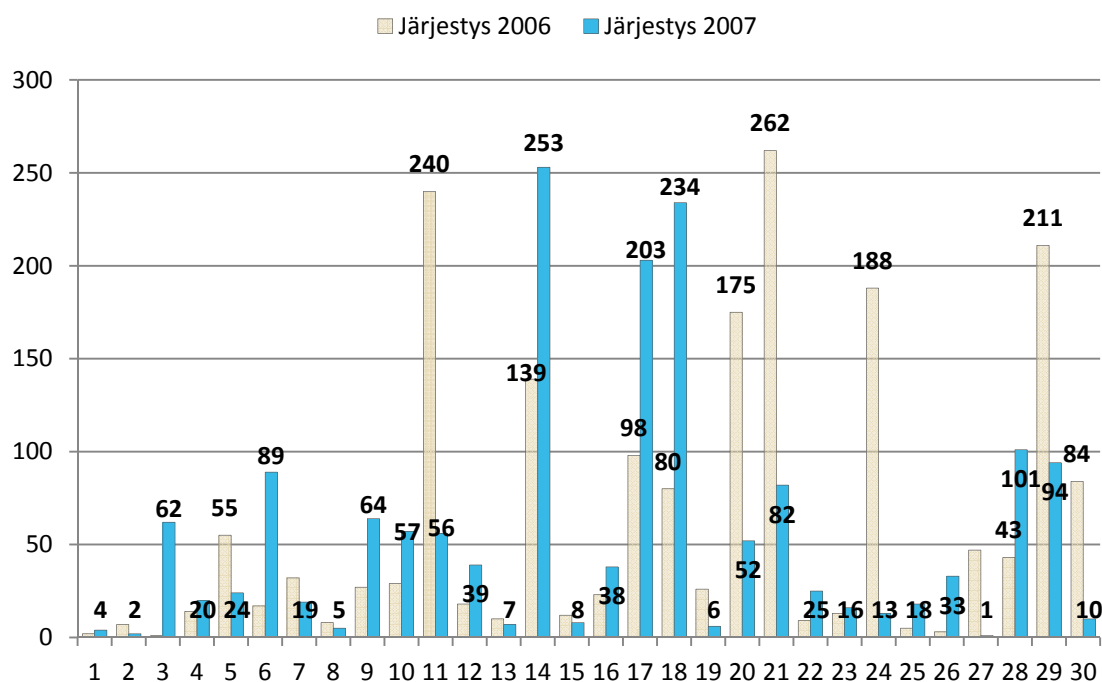
Korrelaatio	2005 ja 2006	2006 ja 2007	2005 ja 2007
Koko aineisto	0,62	0,61	0,60
Parhaat 30	0,33	0,35	0,08
Heikoimmat 30	0,27	0,43	0,17

Kuvassa 1 on esitetty vuoden 2005 parhaiden 30 koulun sijoitus vuosina 2006 ja 2007. Aineistosta on poistettu aivan pienimmät koulut, joista on tuona aikana valmistunut vähemmän kuin kymmenen ylioppilasta. Se vähentää jonkin verran koulusta riippumatonta satunnaista vaihtelua, mutta edelleen koulun estimoitu ”laatu” voi heijastaa oppilaspopulaation muutoksia eri vuosina. Vaihtelu vuoden 2005 parhaiden koulujen joukossa on hyvin voimakas: vuonna 2005 kolmenkymmenen parhaan joukkoon sijoittuva koulu voi vuotta myöhemmin löytyä sijalta 240 ja seuraavan vuonna sijalta 56.

⁴⁰ Teimme vastaavan tarkastelun ammatillisen peruskoulutuksen tulosindeksillä vuosille 2011 ja 2012. Palkituista 30 koulusta 18 oli samoja kahtena peräkkäisenä vuonna ja heikoimmista 25 koulusta samoja oli 14.

Kuvassa 2 puolestaan on tarkasteltu vuonna 2005 kolmenkymmenen heikoimman koulun joukkoa vuosina 2006 ja 2007. Heikoimpien koulujen joukossa vaihtelu ei ole yhtä dramaattista kuin huipulla, vaan valtaosa kouluista sijoittuu vertailun pohjalle vuodesta toiseen. Heikot koulut ovat ilmeisesti muita niin paljon heikompia, että niiden sijoitus ei vaihtelee yhtä rajusti oppilasaineksen mukaan. Vaihtoehtoisesti mallistamme puuttui jotakin näiden koulujen toimintaa oleellisesti selittävää informaatiota, vaikkapa koulun resursseista tai toimintaympäristöstä, joka heijastuu estimoidussa laadussa. Toisaalta muutamat koulut pysyvät myös parantamaan sijoitustaan selvästi: esimerkiksi vuonna 2005 sijalle 394 sijoittunut koulu sijoittuu seuraavana vuonna sijalle 244 ja nousee peräti sijalle 21 vuotta myöhemmin.

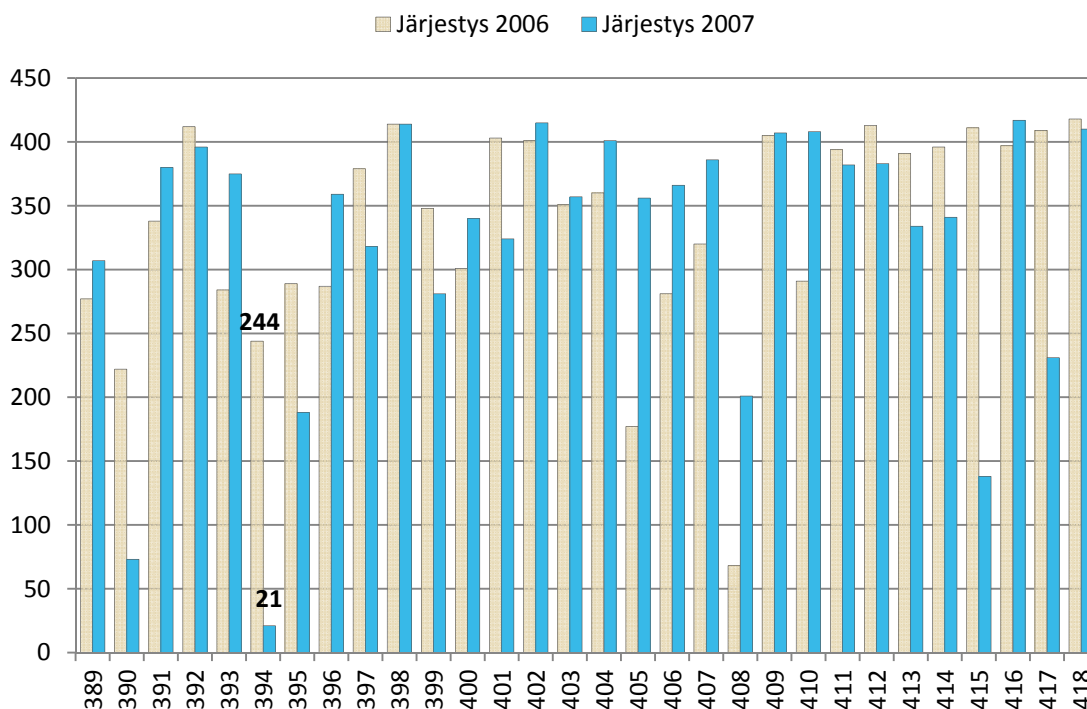
Kuva 1 Koulun estimoidun vaikutuksen vaihtelu. Kolmekymmentä vuoden 2005 parasta koulua vuosina 2006 ja 2007



Molemmissa kuvissa havaittu vaihtelu näyttää kertovan eniten pienten koulujen oppilasaineksen vaihtelusta, jota oppilaan peruskoulun päättötodistuksen keskiarvolla ei pystytä kontrolloimaan. Kun tarkastellaan eniten sijoitustaan muuttaneita kouluja, niin niissä on ollut alle 30 ylioppilaskokelasta kunakin vuonna. Tuloksellisuusrahoitusta on vaikea laatia niin, että pienten koulujen oppilasaineksesta johtuva variaatiota pystyttäisiin kokonaan kontrolloimaan, jolloin rahoitukseen välttämättä liittyy väärää satunnaisuutta. Pieniä lukioita ei voitane myöskään jättää rahoituksen ulkopuolelle, sillä huomattava osa, noin kolmannes,

Suomen lukioista on sellaisia, joissa ylioppilaskokelaita oli kunakin vuonna vähemmän kuin 30.⁴¹

Kuva 2 Koulun estimoidun vaikutuksen vaihtelu. Kolmekymmentä vuoden 2005 heikointa koulua vuosina 2006 ja 2007



Koulun koon vaikutusta ylioppilaskirjoitusten tuloksiin voidaan testata suoraan ainakin kahdella tavalla. Malliin voidaan sisällyttää koulun kokoa mittaava muuttuja, joka on kunakin vuonna ylioppilaskirjoituksiin osallistuneiden lukumäärä. Toinen tapa on jakaa koulut ”pieniin” ja ”suuriin” jonkun ennalta määrätyn rajan mukaan ja jakaa aineisto osiin sen mukaan. Molemmat tuottavat karkeasti arvioiden saman tuloksen: pienissä kouluissa kirjoitetaan keskimäärin suuria kouluja heikommin.⁴² Tämä siitä huolimatta, että nykyinen rahoitusjärjestelmä tukee pieniä kouluja muita kouluja korkeamman valtionosuuden kautta.

7.3.2 Rajauksia tuloksiin ja niiden käytettävyyteen

Arvioimisen kannalta laadun parantamiseen liittyy eräitä teknisiä ongelmia. Tärkein niistä on, että absoluuttisia muutoksia ei havaita, joten palkitseminen perus-

⁴¹ Aiemmat tutkimukset osoittavat, että pienten koulujen kokema satunnainen vaihtelu arvonlisäyksessä heittää kouluja satunnaisesti vuoroin huipulle vuoroin pohjalle (Kane ja Staiger, 2002a ja 2002b).

⁴² Lisäämällä ensimmäiseen mallin kokelaiden lukumäärän neliö, voidaan testata kasvavatko kirjoitusten tulokset opiskelijamäärän mukana ilman rajaa, vai löytyykö joku oppilaiden määrän optimiarvo. Näyttäisi siltä, että suuruuden ekonomialla on rajansa, mutta estimointitulokset ovat tältä osin hyvin epävarmoja. Koulun optimikoko voisi asettua jonnekin 60 ja 140 väliin per oppilasikäluokka, mikä tarkoittaisi noin 180 – 420 oppilaan koulukokoa.

tuu oletukseen, että parannuksia ylipäätään olisi mahdollista saada aikaan. Ilman kansainvälistä vertailua, on epäselvää missä määrin Suomen järjestelmässä on tehottomuutta. Ainoa tulos, joka tässä tutkimuksessa edes hivenen viittaa tähän mahdollisuuteen on oppilasklusteroinnin perusteella saadut tulokset, joiden mukaan koulut tuottavat erilaista laatua lähtökohdiltaan ”hyville” ja ”heikoille” oppilaille. Koska aineistomme oli erittäin suppea ja kontrollimuuttujia vähän, saatu tulos voi johtua mallin spesifikaation puutteista. Tuloksista ei ole siksi syytä tehdä johtopäätöksiä ennen kuin tiedämme enemmän havaituista eroista. Jos tulos pysyy voimassa laajemmalla kontrollimuuttujilla tehdyn analyysin jälkeen, voisi olla paikallaan pohtia, onko kannusteilla mahdollista vaikuttaa eri oppilasryhmien kouluista saamaan lisäarvoon.

Toinen varaus liittyy peruskoulun päättötodistusten arvosanatietojen käyttöön arvonlisäyksen mittaamisessa. Se perustuu oletukseen, että päättötodistuksen arvosanat heijastavat oppilaiden todellista osaamista, vaikka on jokin verran näyttöä siitä, että näin ei ole (mm. Alatupa, 2007). Pahimmillaan virheellinen lähtötason mittaustulos tuo harhaa mallin parametriestimaatteihin (Ladd ja Walsh, 2002). Kun oppilaskohtaisen ryhmittelyn malleihin otetaan selittäväksi tekijäksi sukupuoli, havaitaan, että ensimmäisessä, toisessa ja seitsemännessä ryhmässä sukupuolella on tilastollisesti merkitsevä vaikutus ylioppilaskirjoitusten tuloksiin. Parhaat ja heikoimmat tytöt menestyvät heikommin, kuin mitä heidän peruskoulun päättötodistuksen arvosanansa antaisi odottaa.⁴³ Vastaavasti arvosanajakauman keskivaiheilla, ryhmissä neljä ja viisi, tytöt menestyvät kirjoituksissa paremmin, kuin mitä peruskoulun päättötodistus antaisi odottaa. Peruskoulun päättötodistuksen arvostelu ei siten välttämättä takaa hyvää lähtökohtaa lukion laadun tai arvonlisän mittaamiselle.⁴⁴ Peruskoulussa suoritettavat valtakunnalliset kokeet voisivat antaa lisäinformaatiota oppilaiden todellisista kyvyistä. Ainakin ne antaisivat yhden välineen oppilaan lopullisen arvosanan ja kansallisen kokeen suoritusten vertailuun.

⁴³ Vaikutus on suurin heikoimmille tytöille, joilla sukupuoli vastaa lähes 0,3 keskiarvon muutosta, Esi-merkiksi se, että kokelas on tyttö vastaa samaa kuin, jos päättötodistuksen keskiarvo laskisi 7,3:sta 7,0:aan,

⁴⁴ Sekin on ongelma, jos oppilaat valikoituvat peruskoulun päättötodistuksensa arvosanojen perusteella aivan väärille aloille.

8 Johtopäätöksiä ja suosituksia

Kannustejärjestelmät ja niihin liittyvä tutkimus on ollut erityisen suosittua Yhdysvalloissa. Yhdessä toimintaympäristössä aikaansaadut tulokset eivät kuitenkaan aina ole suoraan sovellettavissa muihin tilanteisiin, joten erilaisissa toimintaympäristöissä saatuja tuloksia on hyvä tarkastella varauksella. Suomen ja Yhdysvaltojen koulutusjärjestelmän tärkein ero liittyy regulaatioon; siinä missä Yhdysvalloissa harjoitetaan laadunvalvontaa oppilastulosten arviointien kautta, Suomessa on luotettu korkeatasoiseksi arvioituun opettajakoulutukseen ja opettajien ja koulujen johtajien valistuneisuuteen, joiden on oletettu johtavan toivotuun lopputulokseen. Hieman kärjistäen, Suomessa vastuu koulutuksen laadusta on siis yksittäisellä opettajalla ja koulun rehtorilla, jotka toimivat annetun budjetin ja tuntikehyksen puitteissa.

Kansainvälisessä tarkastelussa julkisten varojen näin vapaa käyttö ilman tulosten ja säästöjen noudattamisen valvontaa on poikkeuksellista. Siksi oppilaan oikeuksien toteutumisen ja hyvän hallinnon näkökulmasta keskushallinnon pohtima tuloksellisuusrahoitus, johon liittyy laadunarviointi, voi hyvinkin olla askel parempaan suuntaan. Toisaalta, jos opettajakoulutukseen pyrkivien valinta tai aineopettajien opintojen läpäisykriteerit on laadittu oikein, tämä ennakkokarsinta voi hyvinkin riittää ja muu sääntely lisää vain byrokratiaa ja tuottaa turhia kustannuksia. Järjestelmien ero on jossain määrin filosofinen; kun monissa maissa opettajia ei seulota läheskään yhtä tiukasti kuin Suomessa, koulujen arviointia ja valvontaa on tarpeen tehdä enemmän. Koska Suomessa seula tunnetusti on tiukka, on jälkikäteisarviointille muita maita vähemmän tarvetta.

Tätä raporttia tehdessämme eteemme ei ole tullut tukimusta, jonka perusteella Suomen lukiokoulutuksessa ilmenisi puutteita, joita ”laadusta” palkitsevalla ohjelmalla voitaisiin korjata. Saatujen tulosten perusteella näyttää siltä, että lukiovaikutukset ja niissä tapahtuvat muutokset, ovat monen koulun kohdalla satunnaisia ja johtuvat koulusta riippumattomista tekijöistä. Lisäksi tuotosmittarin suhteellisesta luonteesta johtuen edes laatuarvioinnin kärkeen sijoittuminen ei takaa korkeata laatua tai absoluuttista laadun parannusta. Määritelmällisesti jonkun on aina oltava paras, vaikka absoluuttisesti laatu kaikilla olisi heikkoa tai heikkenisi vuosien myötä. Silti laadun parantamisesta palkitseminen voisi olla mielekästä, jos kansainväliset vertailut tai kotimainen arviointityö osoittaisivat, että suomalainen lukiokoulutus on heikkolaatuista.

Saatujen tulosten perusteella ainakaan 2000-luvun puolivälissä lukioiden vaikutuksissa oppimistuloksiin ei keskimäärin ollut suuria eroja. Aivan yksinkertaisin mahdollinen arvonlisäysmalli johtaa tulokseen, että kouluja voi olla vaikea panna luotettavasti paremmuusjärjestykseen niiden vaikuttavuuden suhteen. Eri asia on, onko tästä mahdollisesta satunnaisuudesta haittaa, vai onko tärkeämpää rakentaa kouluille kannuste tavoitella hyvää laatua? Ilman ääneen lausuttuja tavoitteita tai

osoitettua puutetta, tähän kysymykseen ei voida ottaa kantaa. Siksi päätöksentekijöiden tulisi lausua uuden järjestelmän tavoitteet ääneen. Erityisesti pitäisi kertoa, mitä *Suomessa* havaittua ongelmaa tuloksellisuusrahoituksella halutaan korjata. Juuri selkeä tavoitteenasettelu ja ohjelman motivointi kertomalla mahdollisista koulutuksessa ilmenneistä puutteista on ensiarvoista, jotta ohjelma olisi vaikuttava ja sitä alettaisiin kouluissa toteuttaa.

Käytännön ja koetun oikeudenmukaisuuden kannalta voi olla ongelmallista, jos palkitseminen on satunnaista, mutta sen väitetään perustuvan estimoituun ”laatuun” tai koulun vaikutukseen. Jos palkinto kuitenkin saa koulut kiinnittämään huomiota laaduksi arvioituun seikkoihin, voi palkitseminen, vaikka se sisältäisi jonkin verran satunnaisuutta, saada koulut panostamaan laatuun. Siitä huolimatta palkitseminen tai ainakin koulun vaikutuksen mittaaminen vuosittain ei ole mielekästä, koska erityisesti pienissä kouluissa satunnaisten tekijöiden vaihtelu on tuolloin liian suurta. Myös byrokratian vähentämiseksi arviointia voisi olla syytä tehdä harvemmallä syklillä, esimerkiksi kahden, kolmen vuoden välein. Toisaalta, jos arviointia tehdään harvoin, voivat palkinnot siintää liian kaukana, jotta ne motivoisivat kouluja muutokseen. Opetushallituksessa on hiljattain otettu käyttöön vaikuttavuusmittari ammatillisessa koulutuksessa, jonka käytännön sovelluksista saamme lisää tietoa muutaman seuraavan vuoden aikana. Uskomme, että päättäjien kannattaa odottaa arviointituloksia ammatillisen koulutuksen alalta ennen, kuin vaikuttavuusmittari otetaan käyttöön lukiokoulutuksessa.

Toisin kuin ammatillisessa koulutuksessa tai monissa muissa maissa, suomalainen ylioppilastutkinto itse asiassa on jo kannustejärjestelmä, nk. high stake -test, jossa menestymisestä seuraa palkkio oppilaalle. Siksi se kannustaa ainakin akateemisesti suuntautuneita nuoria tuomaan esille todelliset kykynsä ja luultavasti myös opettajilla on jo olemassa *sosiaaliset* kannusteet opettaa oppilaitaan mahdollisimman hyvin. Siksi opettajan palkan tai koulun rahoituksen sitominen oppilaiden ylioppilaskoemenestykseen ei välttämättä muuta opettajien opetuskäytäntöjä. Lisäksi on vaikea hahmottaa, miten koulujen palkitseminen oppilaiden pääsystä jatko-opintoihin voisi vaikuttaa opettajien kannusteisiin, sillä siinä kouluja palkitaan osin samasta työstä kuin mitä tarvitaan hyvään ylioppilastutkintomenestykseen.

Jos jatko-opintoihin pääsyä voidaan tehostaa parantuneella oppilaanohjauksella, voisi olla järkevämpää kohdentaa resursseja suoraan siihen. Samoin, jos koulu voi vaikuttaa keskeyttämissiin lisäämällä oppilashuoltoa tai oppilaanohjausta, loisi rahoituksen sitominen näihin indikaattoreihin kouluille kannusteen lisätä näitä toimenpiteitä ja ohjata niihin lisää resursseja. Ensin täytyy kuitenkin varmistaa, missä määrin keskeyttämiset johtuvat koulusta eivätkä oppilaasta tai koulun toimintaympäristöstä. Jos keskeyttämiset johtuvat koulusta riippumattomista syistä, ei palkitsemisen sitomisella vähäisiin keskeyttämissiin ole toivottua vaikutusta eikä se ole edes toivottavaa.

Muiden maiden järjestelmien tarkastelu toi esiin, että Unkarissa kansallisissa kokeissa huonosti menestyvien koulujen tulee valmistella kehitykseen tähtäävä toimintasuunnitelma. Portugalissa heikkojen koulujen tulee määrittää aikataulu, jonka mukaan ne käynnistävät erilaisia korjaavia toimenpiteitä, mukaan lukien alisuoriutuville oppilaille annettava tukiopetusta. Isossa-Britanniassa koulut, joissa tietty osuus oppilaista ei saavuta asetettuja oppimistavoitteita, ovat oikeutettuja lisätukeen ja -seurantaan. Ehkä Suomessakin voitaisiin harkita heikkojen koulujen tehostettua seurantaa, sillä näissä malleissa koulut pakotetaan pohtimaan laatua ja oppilaiden palvelemista eksplisiittisesti eikä jonkun mahdollisesti epämääräiseltä vaikuttavan mittarin välityksellä. Usein toimenpiteiden kohdistaminen suoraan ongelmakohtaan on tehokkaampaa kuin toimenpiteiden kohdistaminen koko ryhmälle. Jos lukioiden oppilaista voidaan tunnistaa osajoukko, joka hyötyisi kohdennetuista toimenpiteistä, resursseja kannattaa suunnata sinne. Tässä tutkimuksessa emme pysty tällaista oppilasryhmää seulomaan, mutta asian selvittämiseen kannattanee varata aikaa ja resursseja.

Koko koulutusjärjestelmän kannalta olisi parasta, jos suomalaiset viranomaiset olisivat valmiita kokeelliseen tutkimukseen samaan tapaan, kuin luvuissa 4 ja 5 esitettiin. Tällöin saataisiin tutkittua tietoa päätöksenteon tueksi eivätkä päätökset perustuisi arvauksiin. Pahin mahdollinen virhe on ajaa läpi suuri muutos, jonka toteuttamisesta koituu kustannuksia, mutta jonka odotettavista hyödyistä ei ole riittävästi tietoa.

Lähteet

- Alatupa, S. – Karppinen, K. – Keltikangas-Järvinen, L. – Savioja, H. (2007):
Koulu, syrjäytyminen ja sosiaalinen pääoma – Löytyykö huono-osaisuuden
syy koulusta vai oppilaasta? Sitran raportteja 75.
- Amrein-Beardsley, A. (2008): Methodological concerns about the education
value-added assessment system. *Educational Researcher* 37 (2): 65–75.
- Angrist, J. – Lavy V. (2009): The effects of high stakes high school achievement
awards: Evidence from randomized trial. *American Economic Review* 99 (4):
1384–1414.
- Ashraf, N. – Bandiera, O. – Jack, K. (2012): No margin, no mission? A field
experiment on incentives for pro-social tasks. *CEPR Discussion Paper* No.
8834.
- Ballou, D. – Sanders, W. – Wright, P. (2004): Controlling for student
background in value-added assessment of teachers. *Journal of Educational
and Behavioral Statistics* 29 (1): 37–65.
- Bebchuk, L.A. – Cohen, A. – Spamann, H. (2010a): The wages of failure:
Executive compensation at Bear Stearns and Lehman 2000–2008. *Yale Journal
of Regulation* 27: 257–282.
- Bebchuk, L. – Grinstein, Y. – Peyer, U. (2010b): Lucky CEOs and lucky
directors. *Journal of Finance* 65 (6): 2363–2410.
- Bowles, S. – Gintis, H. (2002): Schooling in capitalist america revisited.
Sociology of Education 75 (1): 1-18.
- Burgess, S. – Croxson, B. – Gregg, P. – Propper, C. (2001): The intricacies of the
relationship between pay and performance for teachers: do teachers respond to
performance related pay schemes. *CMPO Working Paper*, Bristol.
- Carnoy, M. – Brodziak, I. – Molina, A. – Socías, M. (2007): The limitations of
teacher pay incentive programs based on inter-cohort comparisons: the case of
Chile’s SNED. *Education Finance and Policy* 2 (3): 189–227.
- Che, Y. – Yoo, S. (2001): Optimal incentives for teams. *American Economic
Review* 91 (3): 525–541.
- Cullen, J.B. – Jacob, B.A. – Levitt, S. (2006): The effect of school choice on
participants: Evidence from randomized lotteries. *Econometrica* 74: 1191–
1230.
- Department for Children, Schools and Families DCSF (2008): *Technical guide to
post-16 contextual value added 2008 model*.

- Department for Education DfE (2010): *The importance of teaching – the schools white paper 2010*.
- Driscoll, D. – Halcoussis, D. – Svorny, S. (2008): Gains in standardized test scores: evidence of diminishing returns to achievement. *Economics of Education Review* 27 (2): 211–220.
- Education, Audiovisual and Culture Executive Agency EACEA (2012): *Key data on education in Europe 2012*. Eurydice: Brysseli.
- Ehrlich, I. (1975): The Deterrent Effect of Capital Punishment: A question of life and death. *American Economic Review* 65 (3), pp. 397–417.
- Eurybase (useita vuosia): *National education system descriptions*.
http://eacea.ec.europa.eu/education/eurydice/eurybase_en.php#description
- Eurydice (2012): Key data on Education in Europe 2012. European Commission, Eurydice.
- Fazekas, M. (2012): School funding formulas – Review of main characteristics and impacts. *OECD Education Working Paper* No. 74.
- Fryer, R.G. (2011): Financial incentives and student achievement: Evidence from randomized trials. *The Quarterly Journal of Economics* 126 (2):1755–1798.
- Fryer, R.G. (2011): Teacher incentives and student achievement: Evidence from New York City public schools. *NBER Working Paper* 16850.
- Fryer, R.G. – Levitt, S.D. – List, J. – Sadoff, S. (2012): Enhancing the efficacy of teacher incentives through loss aversion: A field experiment. *NBER Working Paper* 18237.
- Gelber, A.M. – Isen, A. (2011): Children’s schooling and parents’ investment in children: Evidence from the head start impact study. *NBER Working Paper* 17704.
- Gersback, H. – Keil, M. (2005): Productivity improvements in public organizations. *The Economic Journal* 225: 671–688.
- Green, J. – Stokey, N. (1983): A comparison of tournaments and contracts. *Journal of Political Economy* 91 (3): 349–364.
- Hægeland, T. – Kirkebøen, L. (2008): School performance and value-added indicators – what is the importance of controlling for socioeconomic background? A background report for the oecd project on the development of value-added models in education systems.
- Hamilton, L. (2005): Lessons from performance measurement in education. kirjassa *High-performing government: structure, leadership, incentives*, R. Klitgaard ja P.C. Light (toim.), RAND Corporation: Santa Monica, CA.

- Hanushek, E.A. (2002): Publicly provided education. Kirjassa *Handbook of public economics* Vol. 4, Auerbach, A.J. – Feldstein, M. (toim.), Elsevier B.V.: Amsterdam.
- Hanushek, E.A. (2003): The failure of input-based schooling policies. *The Economic Journal* 113: F64–F98.
- Hanushek, E.A. (2010): Education production functions: Developed country evidence. *International Encyclopedia of Education 2*: 407–411.
- Hautamäki, J. – Säkkinen, T. – Tenhunen, M.-L. – Ursin, J. – Vuorinen, J. – Kamppi, P. – Knubb-Manninen, G. (2012): Lukion tuottamat jatkokoulutusvalmiudet korkeakoulutuksen näkökulmasta. *Koulutuksen arviointineuvoston julkaisuja* 59.
- Heneman, H. – Milanowski, A. – Kimball, S. (2007): Teacher performance pay: synthesis of plans, research, and guidelines for practice. *CPRE Policy Briefs RB-46*.
- Holmström, B. (1979): Moral hazard and observability. *The Bell Journal of Economics*, 10 (1): 74–91.
- Holmström, B. (1982): Moral hazard in teams. *The Bell Journal of Economics*, 13 (2): 324–340.
- Holmström, B. – Milgrom, P. (1991): Multitask principal-agent analyses: incentive contracts, asset ownership and job design. *Journal of Law, Economics and Organization* 7 (special issue): 24–52.
- Jacob, B. A. (2005): Accountability, incentives and behavior: Evidence from school reform in Chicago. *Journal of Public Economics* 89 (5–6): 761–796.
- Jacob, B. A. – Lefgren, L. (2008): Can principals identify effective teachers? evidence on subjective performance evaluation in education. *Journal of Labor Economics* 26 (1): 101–136.
- Jacob, B. A. – Levitt, S.D. (2003): Rotten apples: An investigation of the prevalence and predictors of teacher cheating. *Quarterly Journal of Economics* 118 (3): 843–877.
- Jacob, B. A. – Ludwig, J. (2009): Improving educational outcomes for poor children. Kirjassa *Changing Poverty, Changing Policies*, Cancian, M. – Danziger, S. (toim.), Russell Sage Foundation: New York.
- Kahneman, D. – Tversky, A. (2000): *Choices, values and frames*. Cambridge University Press: Cambridge.
- Kandel, E. – Lazear, E. (1992): Peer pressure and partnerships. *The Journal of Political Economy* 100 (4): 801–817.

- Kane, T. – Staiger, D. (2001): Improving school accountability measures. *NBER Working Paper* No. 8156. National Bureau of Economic Research: Cambridge, MA.
- Katz, L. – Levitt, S. – Shustorovich, E. (2003): Prison conditions, capital punishment, and deterrence. *American Law and Economics Review* 5: 318–343.
- Kelley, C. (2002): Financial incentives in state accountability systems: performance pay for teachers. Center on reinventing public education and the Washington state academic achievement and accountability commission, Seattle, WA.
- Kelley, C. – Heneman, H. – Milanowski, A. (2000): School-based performance award programs, teacher motivation, and school performance: Findings from a study of three programs. *CPRE Research Report Series RR-44*.
- Kumpulainen, T. (toim.) (2011): *Opettajat Suomessa 2010*. Koulutuksen seurantaraportti 2011:6, Opetushallitus. Tammerprint: Tampere.
- Kuusela, J. (2003): Lukioiden tuloksiin vaikuttavista tekijöistä. Opetushallitus, Moniste 13/2003.
- Ladd, H. (1999): The Dallas school accountability and incentive program: an evaluation of its impacts on student outcomes. *Economics of Education Review* 18 (1): 1–16.
- Ladd, H. – Walsh, R. (2002): Implementing value-added measures of school effectiveness: Getting the incentives right. *Economics of Education Review* 21 (1): 1–17.
- Lavy, V. (2009): Performance pay and teachers' effort, productivity and grading ethics. *American Economic Review* 99 (5): 1979–2011.
- Lavy, V. (2002): Evaluating the effect of teachers' group performance incentives on pupil achievements. *Journal of Political Economy* 110 (6): 1286–1317.
- Lavy, V. – Paserman, M.D. – Schlosser, A. (2012): Inside the black box of ability peer effects: Evidence from variation in the proportion of low achievers in the classroom. *The Economic Journal* 122: 208–237.
- Lazear, E. (2000): Performance pay and productivity. *American Economic Review* 90 (5): 1346–1361.
- Lazear, E. (1998): *Personnel economics for managers*. Wiley: New York.
- Lazear, E. (1996): Performance pay and productivity. *NBER Working Paper* No. 5672.
- Lazear, E. (1995): *Personnel economics*. MIT Press: Cambridge, MA.

- Lazear, E. – Rosen, S. (1981): Rank-order tournaments as optimum labor contracts. *Journal of Political Economy* 89 (5): 841–864.
- Levačić, R. (2008): Funding schools by formula. In *Governance and performance of education systems*, Soguel, N. ja Jaccard, P. (eds.), 205–235. Springer: Alankomaat.
- Lewis, J. – Springer, M. (2008): Performance incentives in texas: why schools chose not to participate. *Working Paper #2008–18*. National center on performance incentives, Nashville, TN.
- Levitt, S.D. – List, J.A. – Neckermann, S. – Sadoff, S. (2012): The behavioralist goes to school: Leveraging behavioral economics to improve educational performance. *NBER Working Paper* 18165.
- Locke, E. (1968): Toward a theory of task motivation and incentives. *Organizational behavior and human performance* 3 (2): 157–189.
- Locke, E. – Latham, G. – Smith, K. (1990): *A theory of goal setting and task performance*. Prentice-Hall: Upper Saddle River, NJ.
- Locke, E. – Saari, L. – Shaw, K. – Latham, G. (1981): Goal setting and task performance: 1968-1980. *Psychological Bulletin* 90 (1): 125–152.
- McCaffrey, D. – Lockwood, J. – Koretz, D. – Hamilton, L. (2003): *Evaluating value-added models for teacher accountability*. The RAND Corporation: Santa Monica, CA.
- McCaffrey, D. – Han, B. – Lockwood, J. (2009): Incentive system design and measurement. *kirjassa performance incentives: their growing impact on american k-12 education*, M.G. Springer (ed.), Brookings institution press: Washington D.C.
- Milgrom, P. – Roberts, J. (1992): *Economics, organization & management*. Prentice hall: Englewood cliffs, CA.
- Mizala, A. – Romaguera, P. (2004): School and teacher performance incentives: the latin american experience. *International journal of educational development* 24 (6): 739–754.
- Mizala, A. – Romaguera, P. (2005): Teachers' salary structure and incentives in Chile. *Kirjassa Incentives to improve teaching: lessons from Latin America*, Vegas, E. (ed.), The World bank: Washington D.C.
- Mizala, A. – Romaguera, P. – Urquiola, M. (2007): Socioeconomic status or noise? Tradeoffs in the generation of school quality information, *Journal of development economics*, 84(1): 61–75.

- Murnane, R. – Cohen, D. (1986): Merit pay and the evaluation problem: why most merit pay plans fail and few survive. *Harvard education review* 56 (1): 1–17.
- Nalebuff, B. – Stiglitz, J. (1983): Prizes and incentives: towards a general theory of compensation and competition. *The Bell journal of economics* 14 (1): 21–43.
- Neal, D. (2009): Designing incentive systems for schools. kirjassa *performance incentives: their growing impact on american k-12 education*, Springer, M.G. (ed.), Brookings institution press: Washington D.C.
- OECD (2011): *Education at a glance 2011: OECD Indicators*.
- OECD (2009): *Evaluating and rewarding the quality of teachers. International practices*. OECD PUBLICATIONS: Pariisi.
- OECD (2008): *Measuring improvements in learning outcomes – best practices to assess the value-added of schools*. OECD Publications: Pariisi.
- OECD (2007): *Education at a glance*. OECD Publications: Pariisi.
- O’Keeffe, M. – Viscusi, W. – Zeckhauser, R. (1984): Economic contests: comparative reward schemes. *Journal of labor economics* 2 (1): 27–56.
- Opetushallitus (2011): Ammatillisen peruskoulutuksen tulosrahoitus. Tulomittariseloste vuodelle 2012. Oppaat ja käsikirjat 2011:13.
- Opetushallitus (2012): Opetus- ja kulttuuritoimen rahoitus. Yksikköhintojen ja rahoituksen määräytyminen vuonna 2012. *Oppaat ja käsikirjat* 2012:2.
- Ostrom, E. (1996): Crossing the great divide: Coproduction, synergy and development. *World Development* 24 (6): 1073-1087.
- Podgursky, M. – Springer, M. (2007): Teacher performance pay: A Review. *Journal of policy analysis and management* 26 (4): 909–949.
- Pursiainen, H. – Pääkkönen, J. – Seppälä, T. (2011): Julkisten palvelujen tuottavuusseurannan kehittäminen, valtioneuvoston kanslian raporttisarja 14/2011.
- Pfeffer, J. (1995): *Competitive advantage through people: unleashing the power of the work force*. Harvard business school press: Boston, MA.
- Rau, T. – Contreras, D. (2011): Tournament incentives for teachers: the case of chile. University of Chile, Department of economics, Chile.
- Raudenbush, S. (2004): *Schooling, statistics and poverty: can we measure school improvement?* Princeton, NJ: Educational Testing Service.
- Richardson, R. (1999): Performance related pay in schools: an assessment of the green papers. London School of Economics and Political Science, Lontoo.

- Rivkin, S. – Hanushek, E. – Kain, J. (2005): Teachers, schools, and academic achievement. *Econometrica* 73 (2): 417–458.
- Rouse, C.E. – Hannaway, J. – Goldhaber, D. – Figlio, D. (2007): Feeling the Florida heat? How low-performing schools respond to voucher and accountability pressure. *NBER Working Paper* 13681.
- Smith, S. – Mickelson, R. (2000): All that glitters is not gold: school reform in Charlotte-Mecklenburg. *Educational Evaluation and Policy Analysis* 22 (2): 101–127.
- Tekwe, C.D. – Carter, R.L. – Ma, C.-X. – Algina, J. – Lucas, M.E. – Roth, J. – Ariet, M. – Fisher, T. – Besnick, M.B. (2004): An empirical comparison of statistical models for value-added assessment of school performance. *Journal of Educational and Behavioral Statistics* 29 (1): 11–36.
- Todd, P. – Wolpin, K. (2003): On the specification and estimation of the production function for cognitive achievement. *Economic Journal* 113 (485): F3–F33.
- Valtioneuvosto (2010): *Valtioneuvoston asetus opetustoimen henkilöstön kelpoisuusvaatimuksista annetun asetuksen muuttamisesta.*
- Valtioneuvosto (1998): *Valtioneuvoston asetus opetustoimen henkilöstön kelpoisuusvaatimuksista 14.12.1998/986.*
- Valtioneuvoston kanslia (2011): Neuvottelutulos hallitusohjelmasta, 17.6.2011.
- Valtiovarainministeriö (2012): Valtion talousarvioesitys.
- Virtanen, H. – Maliranta, M. (2006): *Tulosrahoitusmittariston arviointi – Palkkiomuotoinen tulosrahoitus ammatillisessa peruskoulutuksessa.* ETLA/Taloustieto Oy: Helsinki.
- Webster, W. – Mendro, R. – Bearden, D. – Bembry, K. – Jordan, H. (1997): Rewarding effective schools – theory and practice in an outstanding schools awards program. Annual Meeting of the American Educational Research Association, Chicago.
- Willis, T. – Koch, K. – Lampe, G. – Young, R. – Kellor, E. – Odden, A. (1999): A case study of the state of Kentucky’s school-based performance award program. CPRE: Madison, WI.
- Young People’s Learning Agency YPLA (2011a): *A field guide to 2011/2012 sixth form funding allocations.* YPLA: Coventry, Iso-Britannia.
- Young People’s Learning Agency YPLA (2011b): *A field guide to sixth form success rates.* YPLA: Coventry, Iso-Britannia.

Young People's Learning Agency YPLA (2010): *Value added for 16-18 learners: learner achievement tracker handbook*. YPLA: Coventry, Iso-Britannia.

VATT VALMISTELURAPORTIT -SARJASSA ILMESTYNEITÄ
PUBLISHED VATT POLICY REPORTS

1. Janne Tukiainen – Ville Mälkönen: Jätekuljetuksen sopimusmallien yritysvaikutukset. Helsinki 2010.
2. Antti Moisio – Heikki A. Loikkanen – Lasse Oulasvirta: Public services at the local level – The Finnish way. Helsinki 2010.
3. Juha Halonen – Timo Rauhanen – Erkki Siivonen: Julkisten organisaatioiden veroetudet. Helsinki 2010.
4. Tuuli Ylinen: Työkyvyttömyyseläkkeet ja työnantajan omavastuuperiaate TEL:sta TyEL:iin. Helsinki 2010.
5. Outi Kröger – Timo Rauhanen (toim.): Verotuet Suomessa 2009. Helsinki 2010.
6. Jenni Pääkkönen: Koulutuksen markkinoilla – arvioita ammattikorkeakoulujen tehokkuuseroista ja niiden syistä. Helsinki 2010.
7. Antti Moisio: Metropolialueen hallintomallit taloustutkimuksen näkökulmasta. Helsinki 2011.
8. Heikki Pursiainen – Markku Siikanen – Janne Tukiainen – Ville Mälkönen: Asfalttikartellin kunnille aiheuttamien vahinkojen arviointi – Loppuraportti. Helsinki 2011.
9. Tanja Saxell – Janne Tukiainen – Markku Siikanen: Asfalttikartellin valtiolle aiheuttamien vahinkojen arviointi – Loppuraportti. Helsinki 2011.
10. Antti Moisio: Finanssipolitiikan säännöt ja kuntatalous. Helsinki 2011.
11. Iida Häkkinen Skans: Kotitalouspalveluiden verovähennykset Suomessa ja Ruotsissa. Helsinki 2011.
12. Janne Niemi – Juha Honkatukia: Implications of EU emission trading in the Nordic countries. Helsinki 2011.
13. Tiihonen Jari – Kangasharju Aki – Mäkelä Pia – Saari Juho – Österberg Esa : Päivittäistavarakaupoissa myytävien juomien alkoholipitoisuuden alentamisen vaikutus alkoholin kulutukseen ja kuolemiin. Helsinki 2011.
14. Pursiainen Heikki: Alueellistamisen taloudelliset vaikutukset. Helsinki 2013.
15. Honkatukia Juha – Tamminen Saara: Puolustusvoimauudistuksen aluetaloudelliset vaikutukset. Helsinki 2013.