



OPETUSHALLITUS
UTBILDNINGSTYRELSEN

Najat Ouakrim-Soivio

TOIMIVATKO PÄÄTTÖARVIOINNIN KRITEERIT?

Oppilaiden saamat arvosanat ja Opetushallituksen oppimistulosten seuranta-arviointi koulujen välisten osaamiserojen mittareina

Najat Ouakrim-Soivio

TOIMIVATKO PÄÄTTÖARVIOINNIN KRITEERIT?

Oppilaiden saamat arvosanat ja Opetushallituksen oppimistulosten seuranta-arviointi koulujen välisten osaamiserojen mittareina



Esitetään Helsingin yliopiston käyttäytymistieteellisen tiedekunnan suostumuksella julkisesti tarkastettavaksi Helsingin yliopiston päärakennuksen auditoriossa XII (Fabianinkatu 33) perjantaina 8.11.2013 klo 12.

Ohjaajat

Professori Jukka Rantala
Helsingin yliopisto

Esitarkastajat

Professori Antero Malin
Jyväskylän yliopisto

Erikoistutkija emeritus Jorma Kuusela

Dosentti Kari Törmäkangas
Jyväskylän yliopisto

Dosentti Jan Löfström
Helsingin yliopisto

Kustos

Professori Jukka Rantala
Helsingin yliopisto

Vastaväittäjä

Professori Arja Virta
Turun yliopisto

© Opetushallitus ja tekijä

Raportit ja selvitykset 2013:9

Taitto: Sirpa Ropponen

ISBN 978-952-13-5579-0 (nid.)

ISBN 978-952-13-5580-6 (pdf)

ISSN-L 1798-8918

ISSN 1798-8918 (painettu)

ISSN 1798-8926 (verkkójulkaisu)

Juvenes Print – Suomen Yliopistopaino Oy, Tampere 2013

SISÄLLYS

TIIVISTELMÄ	5
ABSTRACT	7
ESIPUHE	11
1 JOHDANTO	13
1.1 Tutkimuksen rakenne	15
1.2 Oppimistulosten arviointijärjestelmä Suomessa.....	18
1.2.1 Perusopetuksen opetussuunnitelman perusteet osana ohjausjärjestelmää	20
1.2.2 Oppilaan arviointi ja arvosanat.....	24
1.2.3 Seuranta-arvioinnin lähtökohdat	25
1.2.4 Yhteiskunnallisten aineiden oppimistulosten arviointi osana perusopetuksen ohjausjärjestelmää	27
1.3 Aiempia tutkimuksia.....	29
1.3.1 Yhteiskunnallisten aineiden kansainväliset tutkimukset CIVED ja ICCS	31
1.3.2 Oppimistulosten seuranta-arviointeihin ja arvosanan antamiseen liittyviä tutkimuksia	33
1.4 Tutkimuskysymykset.....	38
2 ARVOSANOJEN JA KANSALLISTEN ARVIOINTIEN TARKASTELUA.....	41
2.1 Suhteellinen arviointi ranskalaisessa sentralisoidussa järjestelmässä .	50
2.2 Suhteellinen arviointi suomalaisessa desentralisoidussa järjestelmässä	52
2.3 Kriteereihin perustuva desentralisoitu järjestelmä Ruotsissa	56
2.4 Standardeihin ja suhteelliseen arviointiin perustuva osittain sentralisoitu, osittain desentralisoitu järjestelmä, tapausesimerkkinä Yhdysvallat	64
3 OPETUSSUUNNITELMA- JA KRITEERIPERUSTEINEN OPPILAAN ARVIOINTI SUOMESSA	73
3.1 Oppilaan arviointi perusopetuslaissa ja -asetuksessa	76
3.2 Oppilaan arviointi Perusopetuksen opetussuunnitelman perusteissa (2004).....	79
3.2.1 Arviointi opintojen aikana.....	81
3.2.2 Päätösarviointi	83
3.3 Yhteiskuntaopin ainekohtainen arviointi	84
3.3.1 Yhteiskuntaopin opetuksen keskeiset painopisteet, resurssit ja tavoitteet	89
3.3.2 Yhteiskuntaopin taidot ja tiedot	94
3.3.3 Oppilaan arviointi yhteiskuntaopissa.....	100
3.3.4 Perusteiden antama tuki yhteiskuntaopin arvioinnissa ja arvosanan antamisessa	103

4 VALIDIUS-KÄSITE JA SEN TEORIATAUSTAA	111
4.1 Erilaisia näkökulmia validiuden tarkasteluun	112
4.2 Modernin validiusteorian muotoutuminen	114
4.2.1 Käsitevalidius validiustarkastelun näkökulmana	116
4.2.2 Modernin validiuskäsitteen haasteita ja kritiikkiä	125
4.3 Arviointi ja validius	128
4.3.1 Testin validius	129
4.3.2 Luokkatyöskentelyn arvioinnin validius	135
4.3.3 Arviointiin liittyvä epävalidius	137
4.4 Toulminin argumentaatiomalli ja validiusargumentointi	139
5 TUTKIMUKSEN LUOTETTAVUUSTARKASTELU	145
5.1 Tutkimusaineiston validius	145
5.1.1 Ulkoinen validius.....	145
5.1.2 Sisäinen validius	147
5.2 Tutkimuksen reliaabelius	153
6 TUTKIMUKSEN TULOKSET: KOULUJEN VÄLISET EROT ARVOSANOIHIN SUHTEUTETUISSA SEURANTA-ARVIOINNIN RATKAISUOSUUKSISSA.....	157
6.1 Oppilaiden arvosanat ja seuranta-arvioinnin ratkaisuosuudet	158
6.1.1 Oppilaiden arvosanoista	159
6.1.2 Oppilaiden ratkaisuosuuksista ja arvosanoista	160
6.1.3 Oppilaiden ratkaisuosuuksista ja arvosanoista eri taustamuuttujien mukaan luokiteltuna	166
6.2 Seuranta-arvioinnissa menestyminen sekä arvosanat yhteiskuntaopissa ja historiassa	169
6.2.1 Oppilastasolla	169
6.2.2 Koulujen sisällä.....	173
6.2.3 Koulujen välillä	174
6.3 Koulujen väliset erot keskimääräisissä ratkaisuosuuksissa ja arvosanojen keskiarvoissa	181
6.3.1 Koulujen välisten yhteiskuntaopin ratkaisuosuuksien ja arvosanaerojen laskeminen	181
6.3.2 Odotusarvojen laskeminen	182
6.4 Arvosanat ovat validi osaamisen mittari?.....	188
6.4.1 Väitteen, arvosanat ovat validi mittari, koetteleminen.....	190
6.4.2 Väitteen, arvosanat eivät ole validi mittari, koetteleminen.....	202
7 KESKEISET TULOKSET	205
8 POHDINTA	213
8.1 Tulosten arviointia.....	213
8.2 Tutkimuksen arviointia ja jatkotutkimusaiheita.....	220
8.3 Tutkimustulosten yhteiskunnallinen merkitys.....	223
LÄHTEET	227
LIITTEET: TILASTOLIITE	247

Tässä tutkimuksessa on analysoitu oppilaan arvioinnissa ilmenneitä koulujen välisiä eroja suhteessa oppilaiden arvosanoihin ja Opetushallituksen oppimistulosten seuranta-arvioinnissa osoittamaan osaamiseen. Tutkimuksessa tarkasteltiin yhtäältä opettajien antamien arvosanojen luotettavuutta osaamisen mittarina ja toisaalta keväällä 2011 Opetushallituksen toteuttaman seuranta-arvioinnin kykyä mitata yhteiskunnallisille aineille perusopetuksen opetussuunnitelman perusteissa määritetyjä osaamisen tavoitteita. Oppilaiden perusopetuksen päättövaiheen arvosanoja ja seuranta-arvioinnissa osoitettua osaamista yhteiskuntaopissa verrattiin oppilaiden saamiin historian päättöarvosanoihin ja oppimistuloksiin.

Tutkimuksen taustalla oli tieto siitä, että 2000-luvun alusta lähtien äidinkielen ja matematiikan seuranta-arvioinneissa on toistuvasti ilmennyt koulujen välisiä eroja keskimääräisten arvosanojen ja seuranta-arvioinnissa osoitetun keskimääräisen osaamisen suhteen. Samalla osaamisen tasolla on siten saattanut saada eri arvosanan. Yhteiskuntaopin ja historian oppimistulosten arviointi osoitti, että myös yhteiskunnallisissa aineissa oli systemaattisia eroja seuranta-arvioinnissa osoitetun osaamisen ja koulujen antamien arvosanojen välillä.

Tässä tutkimuksessa on pyritty vastaamaan seuraaviin kysymyksiin:

1. Miten oppimistulosten seuranta-arviointi ja arvosanat vastaavat toisiaan?
2. Esiintyykö kahden arvioinnin välillä mahdollisia systemaattisia eroja ja jos esiintyy, miten ne ilmenevät?
3. Voidaanko seuranta-arvioinnin perusteella tehdä johtopäätöksiä arvosanojen validiudesta, vai jääkö kahden mittauksen validiusongelma ratkaisematta?
4. Voidaanko kahden oppiaineen samanaikaisella arvioinnilla ratkaista kahden mittauksen eli arvosanojen ja seuranta-arvioinnin validiutta koskeva ongelma?

Tutkimuksen lopuksi on pohdittu minkälaisia johtopäätöksiä voi aineiston perusteella tehdä oppilaiden arvosanoista, arvosanojen antamista ohjaavista normeista ja arvosanojen antamisen käytänteistä. Tutkimuksen otos käsitti 3 613 suomenkielisten ja 545 ruotsinkielisten koulujen oppilasta eli yhteensä 4 158 oppilasta.

Tutkimuksen teoriaosassa on verrattu suomalaista koulutusjärjestelmää ranskalaiseen, ruotsalaiseen ja yhdysvaltalaiseen koulutusjärjestelmään. Suomessa oppilaiden päättöarvosanat perustuvat valtakunnallisesti määriteltyihin päättöarvioinnin kriteereihin ja opettajan asema arvioijana on itsenäinen. Oppimistulosten arviointeja tehdään otospohjaisesti ja saatuja tuloksia hyödynnetään arvioitaessa opetussuunnitelman perusteissa määriteltyjen tavoitteiden saavuttamista.

Yhteiskuntaopin ja historian tutkimusaineisto mahdollisti ensimmäistä kertaa kahden erillisen oppiaineen välisen vertailun samojen oppilaiden ja koulujen tuottamalla aineistolla ja kahdella rinnakkaismittauksella, arvosanoilla ja seuranta-arvioinnissa osoitetulla osaamisella. Kerätty tutkimusaineisto ja kahden oppiaineen, yhteiskuntaopin ja historian, arvosanojen sekä ratkaisuosuuksien vertailu samojen oppilaiden tuloksilla tarjosi mahdollisuuden tutkia lähemmin siitä, kuinka validi osaamisen mitta arvosana on. Validius voidaan määritellä systemaattisten virheiden puuttumiseksi. Mikäli systemaattisia eroja on, ainakin toisen mittauksen validius voidaan kyseenalaistaa.

Tässä tutkimuksessa oppilaiden saamien yhteiskuntaopin arvosanojen ja seuranta-arvioinnissa osoittaman osaamisen validiutta koeteltiin tulkinta-argumenttien ja Toulminin (1958) kehittämän argumentointimallin avulla siten, että laadittiin kaksi keskenään kilpailevaa väitettä:

1. *arvosanat mittaavat validisti oppilaiden yhteiskuntaopin osaamista ja*
2. *oppilaiden arvosanat eivät ole validi mittari mittaamaan opetussuunnitelman perusteissa määriteltyä yhteiskuntaopin osaamista.*

Yhteiskuntaopissa ja historiassa kerätyllä aineistolla kyettiin osoittamaan säännönmukaisuus, joka muodostuu koulujen osaamistason ja oppilasarvioinnin välille: kun koulun osaamistaso osoittautui hyväksi, se osoittautui hyväksi myös keskimääräisiin arvosanoihin nähden.

Kahden eri oppiaineen samanaikaisella arvioinnilla, voitiin ratkaista kahden mittauksen eli arvosanojen ja seuranta-arvioinnin validiusongelma. Kouluissa, joissa yhteiskuntaopin arvosanat annettiin esimerkiksi tiukemmin perustein kuin seuranta-arvioinnissa osoitettu keskimääräinen osaaminen olisi antanut olettaa, annettiin tiukemmin perustein myös historian päättöarvosana. Siksi koulujen välillä tuntuu olevan sellaisia oppiaineesta riippumattomia arviointikäytänteiden eroja, jonka perusteella voidaan tulkita, että arvosanat eivät ole validi mittari mittaamaan koulujen välisiä osaamiseroja

Näyttää siltä, että opettajat suhteuttavat antamansa arvosanat valtakunnallisten kriteerien sijasta muiden oppilaidensa osaamistasoon. Kun perusopetuksen opetussuunnitelman perusteissa määritellyt kriteerit eivät koulutasolla toimi arvioinnin perustana, päättöarvioinnin yhdenvertaisuus ei toteudu. Yhdenmukaisen arvioinnin varmistamiseksi opetussuunnitelman perusteiden olisi tarjottava aiempaa enemmän tukea opettajan arviointityöhön. Tämä toteutuu esimerkiksi määrittelemällä päättöarvioinnin kriteerit tarkemmin ja luomalla opettajille erillistä materiaalia kriteeripohjaisen arvioinnin tueksi.

Asiasanat: perusopetus, oppimistulosten arviointi, päättöarvioinnin kriteerit, arvosanat, arviointi, validius, yhteiskuntaoppi, yhteiskunnalliset aineet

ABSTRACT

This study investigated assessment differences between schools in relation to the grades given to pupils and to the competencies revealed by the assessment of learning outcomes conducted by the Finnish National Board of Education. On the one hand, the reliability of the grades teachers gave were studied as a gauge of learning while on the other hand, this study evaluated the validity of the follow-up assessment conducted by the Finnish National Board of Education in 2011 to measure learning outcomes.

The National Core Curriculum for Basic Education for Social Studies and History determined the learning objectives for these outcomes. The grades given to pupils in the final assessment and the competencies shown in the assessment of learning outcomes in social studies were compared to grades and outcomes in history.

The background to this study has shown that since the beginning of the 21st century, frequent discrepancies have arisen in the assessment of learning outcomes with regard to competencies in mathematics and mother tongue. These discrepancies have been observed between the grades given to pupils and the competence shown in the follow-up assessment in different schools. Pupils at the same level of competence received a different grade. Similarly, the assessment of learning outcomes in social studies and history showed that systematic discrepancies in the social sciences exist between the competence shown in the follow-up study and the grades given by the schools.

This study aimed to answer the following questions:

1. How does the assessment of learning outcomes track grades?
2. Are there systematic discrepancies between these two evaluations, and if so, how do they occur?
3. Can one draw conclusions about the validity of grades from the follow-up assessment or will the validity problem of two measurements remain unsolved?
4. Can a simultaneous assessment of two subjects resolve the problem of two measurements, namely the problem of that relating to the validity of grades and the assessment of learning outcomes?

Finally, this study discusses the kind of conclusions that one can draw from grades, the norms that guide grading and common grading practices. The sample in this study comprised 4 158 pupils: 3 616 from Finnish- language schools and 545 from Swedish- language schools.

The theoretical part of this study compared the Finnish school system to the French, Swedish and American systems. In Finland, grades are based on the description of pupils' command of key knowledge and skills as outlined by the National Core Curriculum for Basic Education. Teachers have an independent status as evaluators. Learning outcomes are assessed through stratified sampling, which serves to evaluate the achievement of the objectives of the Core Curriculum.

The research data in social studies and history enabled, for the first time, the comparison of two separate subjects using data produced by the same pupils and schools as well as by using two parallel measurements: school grades and the competence shown in the follow-up assessment. This research enabled a closer study of the validity of school grades as a measure of competence. This is because the data comprised grades and results of the assessment of the same pupils in two different subjects. Validity can be defined as an absence of systematic errors. Should systematic errors occur, the validity of at least one of the meters can then be questioned.

In this study, the validity of the grades given by the schools and the learning outcomes of the pupils, demonstrated in the follow-up assessment, were tested by means of the interpretative argument and the model of validity of argument created by Toulmin (1958). This testing was carried out by formulating two competing statements:

1. *The grades given by schools are a valid gauge for measuring competence in the social sciences as described in the Core Curriculum and*
2. *The grades given by schools are not a valid gauge for measuring competence in the social sciences as described in the Core Curriculum.*

The data collected in social studies and history made it possible to demonstrate conformity between the schools' competence levels and grades: when a school's competence level was high, also the grades of its pupils were higher than the average.

Using simultaneous measurement of two subjects could solve the problem of validity between the grades and the follow-up assessment. It was found that in the schools where social studies grades were given using stricter criteria than the follow-up assessment let assume, also the grades in final assessment in history were given using this strict criteria. Therefore, there seem to be differences in grading between schools that are independent of the subject taught. According to this result one can assume that the school grades are not a valid measurement to gauge the differences in competences between schools.

It seems that teachers adjust their grades to the general competence level of their class. When the description of key knowledge and skills does not serve as a basis for evaluation, one can argue that there is no resultant equity in final grading. To ensure consistent final assessment, the Core Curriculum should provide teachers with more support for evaluation, by for example, defining grading criteria more clearly or by creating separate material for criteria-based evaluation.

Key words: basic education, assessment of learning outcomes, criteria for final assessment, grades, evaluation, social studies, social sciences

Tämä tutkimus sai alkunsa, kun Opetushallitus ryhtyi valmistelemaan yhteiskunnallisten aineiden oppimistulosten seuranta-arviointia yhteistyössä Helsingin yliopiston Koulutuksen arviointikeskuksen ja opettajankoulutuslaitoksen kanssa loppuvuodesta 2010. Seuranta-arvioinnin tulokset herättivät minussa kiinnostuksen jatkaa tutkimusta siitä, mihin Opetushallituksen raportoimat historian ja yhteiskuntaopin oppimistulosarviointit päättyivät. Olen saanut nauttia omina kouluvuosinani sivistykselliseen tasa-arvoisuuteen ja yhdenvertaisuuteen pyrkivästä peruskoulusta ja käsitykseni näihin arvoihin perustuvan koulutusjärjestelmän vahvuuksista on syventynyt vuosikymmenten saatossa, kun olen toiminut opetusalan erilaisissa tehtävissä. Työskentely Opetushallituksen oppimistulosten arviointiyksikössä avasi silmäni näkemään sen, miten oppimistulosten arviointi kytkeytyy suomalaisessa koulutusjärjestelmässä edellä mainitsemini sivistyksellisiin perusarvoihin.

Sain Opetushallitukselta tutkimuskäyttöön yhteiskunnallisten aineiden seuranta-arviointiaineiston ja haluankin osoittaa suuret kiitokset oppimistulosten arviointiyksikön silloiselle johtajalle, tutkimusprofessori emerita *Ritva Jakku-Sihvoselle* hänen väitöstutkimustani kohtaan osoittamastaan kannustavasta tuesta. Yksikön nykyistä johtajaa, opetusneuvos *Anu Räisästä* ja johtaja *Raakel Tihosta* kiitän lämpimästi siitä, että he kiinnostuivat tutkimukseni tuloksista ja mahdollistivat väitöstutkimukseni julkaisemisen Opetushallituksen julkaisusarjassa.

Erikoistutkija emeritus *Jorma Kuusela* haluan kiittää siitä, että olen saanut olla mestarin opissa: olen oppinut joka päivä uutta saadessani työskennellessäni hänen kanssaan. Erityisesti haluan kuitenkin kiittää häntä suurenmoisesta väitöskirjaohjauksesta. Ilman hänen korvaamattomia neuvojaan ja koko väitöskirjaprosessin ajan kestänyttä kärsivällistä ohjaustaan tämä työ ei ehkä olisi koskaan valmistunut. Osoitan suuret kiitokset myös tutkimukseni kahdelle muulle ohjaajalle, professori *Jukka Rantalalle* sekä historian ja yhteiskunnallisten aineiden didaktiikan yliopistonlehtori, dosentti *Jan Löfströmille* tutkimusaiheeseeni paneutumisesta sekä asian-tuntevasta ja rakentavasta ohjauksesta. Professori *Jarkko Hautamäkeä* kiitän hänen arvokkaista näkemyksistään tutkimukseni alkuvaiheessa ja professori *Tero Autiota* kiitän hänen väitöskirjani käsikirjoituksesta esittämistään hyödyllistä ja selkeistä kommentteista.

Kiitän tutkimukseni esitarkastajia professori *Antero Malinia* ja erikoistutkija, dosentti *Kari Törmäkangasta* perusteellisesta paneutumisesta työhöni sekä heidän esittämistään arvokkaista kommentteista ja rakentavasta palautteesta, jotka olen pyrkinyt ottamaan huomioon viimeistellessäni työtäni. Tutkimuksen taloudellisesta tuesta kiitän *opetus- ja kulttuuriministeriön yleissivistävän koulutuksen yksikköä* saamastani apurahasta ja *Opetushallitusta* saamastani tutkimusaineistosta sekä siitä, että tutkimukseni julkaistaan Opetushallituksen julkaisusarjassa.

Osoitan ystäväilleni FM *Johanna Kumeniukselle* sydämelliset kiitokset hänen tekevästään kielenhuoltotyöstä sekä FM *Paula Lehtonen-Mauffret'lle* ja M.A. *David Mauffret'lle* heidän käännöstyöstään. Julkaisun painokuntoon saattamisesta ja huolellisesta taittotyöstä kiitän lämpimästi Opetushallituksen julkaisusihteeriä *Sirpa Ropposta*.

Olen kiitollinen Helsingin normaalilyseon kannustavasta ja tutkimustyöhön myönteisesti suhtautuvasta työilmapiiristä. Erityiset kiitokset lausun ystäväilleni ja työtovereilleni FM *Virpi Seppälä-Pekkaselle*, FT, dosentti *Marko van den Bergille* ja KT *Lasse Hongistolle* heidän suurenmoisesta tuestaan.

Kiitos äiti *Saara*, isä *Ahmed* ja sisareni *Yasmine* siitä, että olette aina tukeneet ja kannustaneet minua elämässä eteenpäin. Aviomiestäni *Anttia* kiitän hänen rajattomasta uskostaan kykyihini sekä hänen epätietoisyydestä ja kannustavasta suhtautumisestaan väitöskirjatyöhöni. Tiedän Antti, etten olisi koskaan saavuttanut tätä hetkeä ilman sinua!

Porvoon Sommarissa 9.9.2013

Najat Ouakrim-Soivio

1 JOHDANTO

Suuri joukko toisen asteen opettajista tunnistaa tilanteen, jossa opiskelija pettyy uudessa oppilaitoksessa saamiinsa arvosanoihin. Opettajalle voi olla myös tuttu tilanne, jossa hän on arvosanaa antaessaan katsonut uuden opiskelijansa tai opiskelijoidensa edellisiä arvosanoja ja miettinyt sitä, että arvosanat eivät vastaa opiskelijoiden osaamistasoa. Kyse ei aina ole siitä, että opettajat opettavat ja arvioivat oppilaidensa tai opiskelijoidensa osaamista eri tavoitteita, sisältöjä tai kriteereitä painottaen tai siitä, että oppilaiden tai opiskelijoiden osaamistaso ja opiskelumotivaatio olisivat oppilaitoksen vaihtuessa yllättäen kohentuneet tai heikentyneet.

Tullessaan uuteen kouluun tai oppilaitokseen oppilas tai opiskelija huomaa varsin nopeasti, kuinka vaativaa tai lievää arviointikulttuuria opetusryhmässä tai oppilaitoksessa harjoitetaan suhteessa hänen aiempaan luokkaansa tai kouluunsa. Vaikka arvosana on vain osa oppilaan arviointia, se on usein ainoa tieto, jonka koulun ulkopuolinen toimija saa oppilaasta tai opiskelijasta, kun tämä siirtyy esimerkiksi peruskoulusta jatko-opintoihin tai työelämään. Siksi ei ole yhdentekevää, miten kouluissa tulkitaan Perusopetuksen opetussuunnitelman perusteissa (2004) määriteltyjä päättöarvioinnin kriteerejä annettaessa oppilaille päättöarvosanoja.

Suomessa peruskoulun alkuvaiheesta lähtien oppilaan arviointiin ja arvosanan antamiseen on liittynyt kysymys arviointimenetelmien validiudesta. Peruskoulun opetussuunnitelmakomitean mietinnössä 1970-luvun alussa linjattiin vaatimuksia koulutyön tulosten arvioimiseksi. Arviointimenetelmien tuli olla päteviä eli valideja, menetelmien tuli antaa tarkkaa ja virheetöntä tietoa oppilaan suorituksista eli niiden tuli olla reliaabeleja ja arviointimenetelmien täytyi olla hyviä erottelukyvyltään (Komiteanmietintö 1970: A4, 162–163).

Yli neljäkymmenen vuoden jälkeen komiteanmietinnössä esitetyt tavoitteet ovat edelleen ajankohtaisia. Uutta ei myöskään ole se, että arvosanojen validiuteen kytkeytyy mitä suurimmassa määrin kysymys arvioinnin yhdenvertaisuudesta. Kyse ei ole siitä, ovatko oppilaat saaneet ”oikean arvosanan”, sillä sitä on lähes mahdotonta arvioida. Arvioinnin yhdenvertaisuudessa on kyse ennen kaikkea siitä, saavatko oppilaat yhdenvertaisen perustein ja suhteessa samaan osaamistasoon¹ opintojen aikaiset arvosanansa tai päättöarvosanansa, joilla he hakeutuvat jatko-opintoihin.

1 Osaamistasolla tarkoitetaan tässä tutkimuksessa seuranta-arvioinnissa saavutettua ratkaisuosuutta eli ratkaisuprosenttia, joka käsitellään perusteellisemmin luvussa 5, jossa se kuvataan osana yhteiskunnallisten aineiden seuranta-arviointia.

Kun opettamista ja arviointia kansallisesti ohjaava Perusopetuksen opetussuunnitelman perusteet (2004) astui voimaan, aiemmin suosituksiksi laaditut oppiainekohtaiset päättöarvioinnin kriteerit muuttuivat velvoittaviksi normeiksi. Oppiainekohtaisilta päättöarvioinnin kriteereiltä odotettiin paljon, sillä muutamaa vuotta aiemmin voimaan tulleissa perusopetuslaissa (628/1998) ja perusopetusasetuksessa (825/1998) oli kiinnitetty huomiota arvioinnin yhdenmukaisuuteen ja oppilaan yhdenvertaiseen kohteluun (Lankinen & Lahtinen 2009, 236). Oppiainekohtaisten hyvän osaamisen kuvausten arvosanalle 8 oletettiin toimivan ”mittatikkuna” ja varmistavan, että arvosanat annettaisiin yhdenvertaisin periaattein, ja että arvosanat olisivat valtakunnallisesti entistä vertailukelpoisempia, riippumatta siitä, missä koulussa ne oli annettu. Kuitenkin jo 2000-luvun alusta lähtien Opetushallituksen laatimissa seuranta-arvioinneissa on toistuvasti tuotu esiin äidinkielen (esim. Lappalainen 2004, 141; 2010, 88–92) ja matematiikan (esim. Mattila 2005, 110) tulosten raportointien yhteydessä, että koulujen välillä on osaamistasoeroja suhteessa oppilaille annettuihin arvosanoihin. Keväällä 2011 perusopetuksen 9.-luokkalaisilta kerätty ja maaliskuussa 2012 julkaistu yhteiskuntaopin ja historian seuranta-arviointi osoitti, että myös yhteiskunnallisissa aineissa eli historiassa ja yhteiskuntaopissa oli systemaattisia eroja seuranta-arvioinnissa osoitetun osaamisen ja koulujen antamien arvosanojen välillä. Koulujen seuranta-arvioinnissa osoitetun osaamisen ja arvosanojen väliset erot raportoitiin hyvin yleisellä tasolla. (Ouakrim-Soivio & Kuusela 2012, 110–111, 134.)

Koulujen väliset erot seuranta-arvioinnissa osoitetun osaamisen ja arvosanojen välillä ovat merkittävä yhteiskunnallinen kysymys, joka kytkeytyy oppilaiden yhdenvertaisuuteen. Niin kauan kauan kuin oppilaat valikoituvat jatko-opintoihin ja työelämään saamiensa päättöarvosanojen avulla, niiden tulisi olla valtakunnallisesti yhdenvertaisin perustein annettuja. Vaikka edellä on tuotu painokkaasti esiin päättöarvosanojen yhdenvertaisuuteen liittyviä ongelmia, eivät Opetushallituksen seuranta-arvioinnitkaan ole aivan ongelmattomia.

Opetushallituksella ei ole ollut mahdollisuutta analysoida koulujen välisiä oppilaan arvioinnissa ja seuranta-arvioinnissa ilmenneitä eroja, vaikka useissa oppiaineissa on jo vuosikymmenen ajan ollut viitteitä vaihtelevista arviointikäytännöistä. Tässä tutkimuksessa tarkasteltavat koulujen väliset seuranta-arvioinnissa osoitetun osaamisen ja annettujen arvosanojen väliset erot yhteiskuntaopissa ja historiassa tuovat tarpeellista tietoa siitä, kuinka hyvin yhdenvertaisuusperiaatteen näkökulmasta perusopetuksen päättöarvioinnin kriteerit toimivat oppilaan päättöarvosanaa annettaessa.

1.1 Tutkimuksen rakenne

Tämä tutkimus rakentuu siten, että luvuissa yhdestä viiteen tarkastellaan oppilaan arvioinnin ja siinä erityisesti arvosanan antamisen tehtäviä ja perusteita sekä kansallisia ja kansainvälisiä käytänteitä oppimistulosten mittaamisessa.² Tutkimuksen varsinainen empiriaosuus keskittyy lukuun kuusi, ja tutkimuksen kahdessa viimeisessä luvussa tarkastellaan saatuja tutkimustuloksia sekä pohditaan tutkimusta ja sen tuloksia laajemmin.

Ensimmäisessä luvussa lukijalle selvitetään, miten valtakunnallinen perusopetuksen opetussuunnitelman perusteet ja Opetushallituksen tekemät oppimistulosten arvioinnit kytkeytyvät osaksi koulutuspoliittista ohjausjärjestelmää. 1990-luvun lopusta lähtien Opetushallituksen perustehtäviin on kuulunut oppimistulosten arviointi, jolla pyritään toisaalta saamaan arviointitietoa koulutusjärjestelmän toimivuudesta, toisaalta taas tietoa siitä, mikä oppilaiden osaamistaso on eri oppiaineissa perusopetuksen nivel- ja päättövaiheessa. Oppimistulosten arviointien tehtävänä on antaa palautetta ja vaikuttaa sitä kautta perusopetuksen opetussuunnitelmien kehittämistyöhön.

Johdantoluvussa pyritään myös määrittelemään oppilaan arviointiin ja arvosanan antamiseen sekä oppimistulosten arviointiin liittyviä keskeisiä käsitteitä sekä kytkeään ne osaksi aiempia tutkimuksia, joita tarkastellaan yhteiskuntaopin oppiaineen avulla. Yhteiskuntaopin on tutkimuskohteena kiinnostava, koska vuonna 2004 vahvistetussa perusopetuksen opetussuunnitelman perusteissa se erotettiin historian oppiaineen yhteydestä ja siitä tuli oma, itsenäinen oppiaineensa. Yhteiskuntaopissa kuten ei historiassakaan ole arvioitu oppimistuloksia ennen vuotta 2011, jolloin kyseisten oppiaineiden seuranta-arviointiaineisto kerättiin. Johdantoluvun lopussa taustoitetaan lukijalle, miten yhteiskunnallisten aineiden seuranta-arvioinnin tulokset ovat vaikuttaneet tähän tutkimukseen ja esitetään tämän väitöstutkimuksen tutkimuskysymykset.

Tutkimuksen toisessa luvussa keskitytään tarkastelemaan sitä, miten oppilaan arvosanoja sekä kansallisia arviointeja, kuten standardoituja testejä, päättökokeita tai seuranta-arviointeja, ohjeistetaan tai määrätään Suomessa ja kolmessa muussa arviointikulttuuriltaan erilaisessa maassa. Oppilaan arvosanoja ja kansallisia arviointeja tarkastellaan osana sentralisoitua tai desentralisoitua koulutusjärjestelmää, ja tarkastelun toisen ulottuvuuden muodostavat kaksi yleisintä arvosanan antamisen muodostamisperiaatetta eli kriteereihin perustuva arviointi tai suhteellinen arviointi. Luvussa olevat maakohtaiset esimerkit on valittu konkretisoimaan sitä, miten eri aikana tai eri koulutusjärjestelmissä arvosanan antamista ja kansallisia arviointeja on ohjeistettu ja mikä on opettajan asema arvioijana kyseisessä koulutusjärjestelmässä.

2 Oppilaan arvioinnilla tarkoitetaan sekä opintojen aikaista arviointia että päättöarviointia (POPS 2004, 260). Arvosanan antamisella viitataan tässä tutkimuksessa summatiiviseen arviointiin, jossa asetettuja tavoitteita, tietoja ja taitoja arvioidaan esimerkiksi numerolla tai kirjaimella. Kansallisilla oppimistulosten arvioinnilla tarkoitetaan puolestaan arviointitoimintaa, jolla pyritään saamaan tietoa koulutusjärjestelmän toimivuudesta sekä oppilaiden osaamistasosta.

Kolmannessa luvussa kuvataan sitä, miten suomalaisessa järjestelmässä ohjataan oppilaan arviointia ja mitä arvosanan antamisesta määrätään voimassa olevissa säädöksissä ja normeissa. Oppilaan arvioinnin perustan muodostavat perusopetuslaki (1998/628) ja perusopetusasetus (1998/825) sekä Perusopetuksen opetussuunnitelman perusteet (2004), joiden säädöspohjaa ja normeja avataan lukijalle. Koska tässä tutkimuksessa keskitytään tarkastelemaan arvosanojen ja seuranta-arvioinnin validiutta, lukija perehdytetään myös siihen, mihin tavoitteisiin, tietoihin, taitoihin ja kriteereihin yhteiskuntaopin opetus ja sen arviointi on perusopetuksen eri vuosikymmenten aikana perustunut ja mihin se perustuu nykyisin. Opetushallitus on itse arvostellut perusopetuksen oppilaan arviointia ohjaavia säädöksiä, normeja ja ohjeita niiden kerrostuneisuudesta oppilaan arviointia käsittelevissä artikkeleissa, jotka löytyvät Opetushallituksen verkkosivuilta. Koska näille artikkelimuotoisille teksteille ei sivustolla ole aina mainittu varsinaista tekijää tai muut viittaamisen tarvittavat tiedot, kuten tekstin julkaisuajankohta puuttuvat, viitataan tässä tutkimuksessa sähköisiin internetsivuihin tekstiin sijoitetulla alaviitteellä, jonka tiedot löytyvät aakkostettuna lähdeluettelosta sen mukaan, minkä organisaation sivustosta kulloinkin on kyse.³ Luvussa kuvataan tiivistä myös oppilaan arvioinnin ja arvosanan antamisen periaatteet peruskoulun alkuvuosista nykypäivään. Koska arviointi, samoin kuin arvosanan antaminen mittaavat asetettujen tavoitteiden ja sisältöjen saavuttamista, on luvussa pyritty kuvaamaan lyhyesti sitä, miten yhteiskuntaopin opetuksen tavoitteet, oppiaineelle määritellyt tiedot ja taidot sekä arviointiperusteet ovat muuttuneet peruskoulun voimassaolon aikana 1970-luvulta nykypäivään. Luvun lopussa pohditaan ohjausasiakirjojen näkökulmasta tämän tutkimuksen kannalta keskeistä kysymystä eli sitä, tukevatko nykyiset perusopetuksen päättöarvioinnin kriteerit riittävästi esimerkiksi yhteiskuntaopin arviointia ja arvosanan antamista, jotta oppilaiden yhdenvertaisuus arvosanaa annettaessa voisi toteutua.

Tutkimuksen neljäs luku pyrkii avaamaan lukijalle validius-käsitettä, sen muotoutumista kuuden viimeisen vuosikymmenen aikana erityisesti Cronbachin, Messickin ja Kanen työn tuloksena sekä käsitevalidiuden teoriataustaa. Luvun toisena painopisteenä on arvioinnin ja validiuden keskinäinen yhteys esimerkiksi erilaisten kokeiden laadinnassa tai luokkatyöskentelyn arvioinnissa. Luvun lopussa lukija johdellaan siihen, miten empiirisestä aineistosta saatuja tuloksia koetellaan Toulminin (1958) päättelymallia käyttäen sekä siihen, miten yhteiskuntaopin arvosanojen ja yhteiskuntaopin seuranta-arvioinnin validiutta vertaillaan validiuden teoriataustaa hyödyntäen.

3 Alaviitteiden käyttö tässä yhteydessä tekee tekstistä luettavamman kuin jos sivuston osoite liitettäisiin sellaiseen tekstiin. Opetushallituksen sivuston lisäksi tämä käytäntö koskee esimerkiksi Ranskan opetusministeriön verkkosivuilla olevia oppilaan arviointiin liittyviä tekstejä, joitain Ruotsin kouluviraston tekstejä, kuten opettajille suunnattua oppilaan arvioinnin tueksi laadittua sivustoa sekä kansallisille kokeille rakennettua sivustoa sekä Euroopan neuvoston sivustoja, joilta ei myöskään löydy tarvittavia viitetietoja.

Luvussa viisi tarkastellaan tämän tutkimuksen validiutta ja reliabiliutta. Lukijalle selvitetään muun muassa yhteiskunnallisten aineiden seuranta-arvioinnissa käytettyjen yhteiskuntaopin ja historian tehtävien laadintaprosessi ja seuranta-arvioinnin toteuttaminen sekä tulosten raportoiminen.

Luku kuusi muodostaa tutkimuksen varsinaisen empiriaosuuden, ja siinä keskitytään koulujen seuranta-arvioinnissa osoitetun keskimääräisen osaamisen ja oppilaiden saamien keskimääräisten arvosanojen välisiin eroihin yhteiskuntaopissa ja historiassa. Luvun alussa tarkastellaan oppilaiden välisiä arvosana- ja osaamiseroja eri taustamuuttujien mukaan luokiteltuna. Tämän jälkeen tutkimuksessa tarkastellaan koulujen eroja seuranta-arvioinnissa osoitetun osaamisen ja oppilaiden saamien arvosanojen välillä. Koska yhteiskuntaopin ja historian arvosanojen antaminen ja seuranta-arvioinnin tehtävät perustuvat samoihin Perusopetuksen opetussuunnitelman perusteisiin (2004) ja siinä määriteltyihin oppiainekohtaisiin kriteereihin, tulisi arvosanojen ja seuranta-arviointitulosten konvergoitua eli osoittaa samaan suuntaan. Kiinnostavaksi tutkimuksen tulokset muuttuvat, kun kaksi rinnakkaismitausta eivät konvergoidukaan oletetulla tavalla, vaan koulujen välille syntyy eroja. Koulujen välisiä seuranta-arvioinnissa osoitetun osaamisen ja oppilaiden saamien arvosanojen välisiä eroja on tutkittu Kuuselan (2006) kehittämällä menetelmällä.

Empirialuvun lopussa käytetään hyödyksi Toulminin argumentaatiomallia, kun arvioidaan seuranta-arvioinnin ja arvosanojen validiutta. Päättely etenee siten, että laaditaan kaksi keskenään kilpailevaa väitettä oletuksineen ja katsotaan, mihin päädytään, kun oletukset kilpailevine argumentteineen viedään loppuun asti. Luvun kuusi liitteeseen 1 on koottu taulukoita, joilla kuvataan aineiston tilastollisia tunnuslukuja ja korrelaatioita.

Seitsemännessä luvussa esitetään kootusti tutkimuksen keskeiset tulokset. Tutkimuksen viimeisessä eli kahdeksannessa luvussa pohditaan tutkimuksen tuloksia oppilaan arvioinnin ja siinä erityisesti arvosanan antamisen näkökulmasta sekä pohditaan päättöarvioinnin kriteereiden toimivuutta ja tutkimustulosten yhteiskunnallista merkitystä, kuten päättöarvioinnin yhdenvertaisuuden toteutumista. Luvun lopussa esitetään vielä mahdollisia jatkotutkimusaiheita, jotka nousevat esiin tämä tutkimuksen tuloksista ja pohdinnoista.

1.2 Oppimistulosten arviointijärjestelmä Suomessa

Tässä tutkimuksessa oppimistulosten seuranta-arvioinnilla⁴ tarkoitetaan eri oppiainesten oppimistulosten arviointeja, joita Opetushallitus tuottaa, raportoi ja seuraa. Nykyisin oppimistulosten seuranta-arvioinnit ovat osa Opetushallituksen perustettua koulutuksen kehittäjänä ja tuloksellisuuden edistäjänä. Oppimistulosten seuranta-arviointiin liittyvä tiedonkeruu toteutetaan valtakunnallisilla seuranta-arvioinneilla. Kerättävä tieto liittyy valtakunnallisissa opetussuunnitelman perusteissa määriteltyihin tavoitteisiin ja kriteereihin, joiden toteutumista seurataan otantaperusteisin arvioinnein (Metsämuuronen 2009, 7).

Opetushallituksen toteuttamien seuranta-arviointien tarkoituksena on tuottaa tietoa siitä, miten hyvin opetussuunnitelman perusteissa määritellyt koulutukselle asetetut tavoitteet on saavutettu (Jakku-Sihvonen 2011, 7). Seuranta-arvioinnin tehtävänä on myös antaa koulutuksen järjestäjille ja kouluille tietoa siitä, miten esimerkiksi oppilaiden keskimääräinen osaaminen ja koulun oppilasarvioinnin yleinen linja todistusarvosanoissa suhteutuvat kansalliseen tasoon (Ouakrim-Soivio & Kuusela 2012, 15; vrt myös Jakku-Sihvonen 2013, 17–18). Kyseessä ei ole koko ikäluokan päättökoe eikä seuranta-arviointien tarkoituksena ole määrittää yksittäisten koulujen tai oppilaiden keskinäistä järjestystä ja siksi Opetushallituksen kiinnostuksen kohteena eivät ole olleet yksittäiset koulut, opettajat tai oppilaat ja heidän suoriutumisensa, vaan yleinen osaamisen taso (Metsämuuronen 2009, 11; Jakku-Sihvonen 2010, 320; ks. myös Jakku-Sihvonen 2013, 16–18).

Vuonna 2011 järjestetty yhteiskunnallisten aineiden, historian ja yhteiskuntaopin, seuranta-arviointi oli osa perusopetuslaissa säädettyä koulutuksen arviointia. Vaikka yhteiskunnalliset aineet arvioitiin ensimmäisen kerran vasta tuolloin, seuranta-arvioinnit eivät ole uusi asia opetushallinnon alalla.⁵ Opetushallituksesta säädettyyn

4 Valtioneuvoston asetuksessa Opetushallituksesta (805/2008) määrätään Opetushallituksen tehtäväksi toteuttaa oppimistulosten seuranta-arviointeja. Tässä tutkimuksessa kyseistä seuranta-arvioinnin käsitettä käytetään synonyyminä oppimistulosten arvioinnille, koska seuranta-arviointi on asetuksessa käytetty virallinen käsite ja se on vakiintunut kuvaamaan Opetushallituksen toteuttamia oppimistulosten arviointeja. Oppimistulosten arviointi olisi kuitenkin käsitteenä selkeämpi, johdonmukaisempi, yksiselitteisempi ja se kuvaisi oikeammin sitä, mistä on kyse, kun oppimistuloksia arvioidaan.

5 Suomessa on järjestetty 1960-luvulta lähtien Historian ja yhteiskuntaopin opettajien liiton (HYOL ry:n) sekä nykyisen Nordea-pankin kanssa yhteistyössä taloustietokilpailuja. Tietokilpailukysymykset perustuvat yhteiskuntaopissa opetettaviin taloutta käsitteleviin sisältöihin, ei koko yhteiskuntaopin oppimäärään. Taloustietokilpailussa ei myöskään mitata opetussuunnitelman perusteissa asetettujen tavoitteiden saavuttamista, toisin kuin oppimistulosten seuranta-arvioinnissa. Koulut voivat halutessaan osallistua taloustietokilpailuun, seuranta-arvioinnin otokseen valikoitujen koulutuksenjärjestäjien ja koulujen on osallistuttava oppimistulosten arviointiin. Taloustietokilpailussa jokainen kilpailija kilpailee yksilönä ja kilpailussa koulun parhaan oppilaan tulosta verrataan muiden koulujen parhaisiin tuloksiin. Suomessa järjestetään myös muita perusopetuksen yhteiskuntaopin sisältöalueisiin liittyviä kilpailuja, kuten EU-tietokilpailu, jonka sisällöt ja toimintatavat muistuttavat paljon taloustietokilpailua. Oppimistulosten arviointien oppilaskohtaiset tulokset tietää vain oppilaan oma opettaja ja oppilas itse. Koulukohtaiset oppimistulokset annetaan vain koulutuksen järjestäjälle ja koulun omaan käyttöön ja Opetushallitus julkaisee ainoastaan valtakunnalliset tulokset oppimistulosten arviointiraportin yhteydessä. Oppimistulosten arvioinnit ja tietokilpailut poikkeavat tavoitteiltaan, periaatteiltaan ja toimintatavoiltaan toisistaan, eivätkä tietokilpailut mittaa perusopetuksen päättövaiheen yhteiskuntaopin oppiaineelle asetettujen sisältöjen, tavoitteiden ja kriteerien osaamista.

lakiin (182/1991)⁶ määriteltiin viraston tehtäviin kuuluvaksi koulutuksen tuloksellisuuden edistäminen ja koulutuksen järjestämisen seuraaminen. Opetushallituksen arviointitoiminnan perusteet on kirjattu Kansalliseen oppimistulosten arviointijärjestelmään (4/1998) ja Koulutuksen tuloksellisuuden arviointimalliin (7/1998).

Opetushallituksen oppimistulosten seuranta-arvioinnit ovat normiperusteisia sikäli, että yksittäisen oppilaan tuloksia verrataan samaan otokseen kuuluneiden oppilaiden tuloksiin (Metsämuuronen 2009, 17) eli näiden muodostamaan normiin. Toisaalta Opetushallituksen oppimistulosten seuranta-arvioinneissa on jo vuodesta 1998 alkaen arvioitu oppimistuloksia suhteessa opetussuunnitelman perusteissa asetettuihin kriteereihin niin, että hyväksytyin suorituksen sisällölliset kriteerit on ennalta määritelty (Koulutuksen tuloksellisuuden arviointimalli 1998, 56). Siksi oppimistulosten arviointia voidaan perustellusti pitää myös kriteeriperusteisena. Peruskoulua varten laadittiin kriteeriperusteista tietoa tuottava järjestelmä, jossa arvioidaan ennen kaikkea sitä, ovatko oppilaat esimerkiksi asuinpaikastaan, sukupuolestaan ja kotitaustastaan riippumatta saavuttaneet samat yleissivistävät tiedot, taidot ja valmiudet (Koulutuksen tuloksellisuuden arviointimalli 1998, 56). Kriteereillä tarkoitetaan tässä yhteydessä sitä perustetta, johon arvioija tukeutuu, kun hän tekee arvioinnin kohteesta päätelmiä (Koulutuksen tuloksellisuuden arviointimalli 1998, 24; Heinonen 2001a, 23).

Koulutuksen tasa-arvon pääkriteereiksi on nostettu alueellinen, sukupuolten välinen sekä sosiaaliryhmien välinen tasa-arvoisuus. Koulutuksen tasa-arvon lisäksi oppimistuloksissa tarkastellaan koulutuksen saatavuutta, jolla tarkoitetaan oppilaan yhdenvertaisia mahdollisuuksia jatkaa opintojaan peruskoulun jälkeen. Koulutuksen tasa-arvon voidaan katsoa toteutuneen hyvin, mikäli systemaattisia eroja ei ilmene näiden tarkasteltavien osa-alueiden suhteen. (Jakku-Sihvonen 1996, 3.) Tämän vuoksi Opetushallituksen toteuttamia eri oppiaineiden oppimistuloksia on arvioitu 1990-luvun lopulta alkaen ja arvioidaan yhä seuraavien taustamuuttujien suhteen: alueet ja siinä entiset läänit sekä tilastolliset kuntatyytit, kuten maaseutumaiset kunnat, taajamat ja kaupungit, sukupuoli sekä suomen- ja ruotsinkielisten koulujen oppilaat. Kevään 2011 oppimistulosten arviointien yhteydessä kysyttiin ensimmäistä kertaa myös huoltajien koulutustaustaa.

Lähtöoletuksena oppimistulosten arvioinnille on, että peruskoulussa opetetaan sellaisia tietoja ja taitoja, jotka on kirjattu perusopetuksen opetussuunnitelman perusteisiin ja ainekohtaisiin tavoitteisiin, sisältöihin ja kriteereihin. Tästä syystä yhteiskunnallisten aineiden tehtävät samoin kuin muidenkin oppiaineiden seuranta-arviointien tehtävät laaditaan perusopetuksen opetussuunnitelman perusteiden mukaan. Tietoa käytetään opetuksen tilan arviointiin, opetuksen, oppilasarvioinnin ja opetussuunnitelman perusteiden kehittämiseen ja koulutuksen tasa-arvon toteutumisen arviointiin. Systemaattisimmin Suomessa on seurattu 1990-luvun lopulta alkaen oppimistuloksia matematiikassa (Korhonen 1999, 2001; Mattila 2002, 2005; Hirvonen 2012), äidinkielessä ja kirjallisuudessa (Lappalainen 2000,

6 Lakia on uudistettu ja se on nykyisin Valtioneuvoston asetus (805/2008) Opetushallituksesta.

2001, 2004, 2006, 2011) sekä modersmålet och litteratur -oppiaineessa (Hannén 2000; Silveström 2002, 2003, 2004, 2006, 2008; Hellgren 2011). Suunnitelmallinen ja riittävän kattava oppimistulosten arviointi antaa mahdollisuuden myös seurata ja kehittää esimerkiksi oppilaan arviointia suhteessa asetettuihin tavoitteisiin ja arvioinnin kriteereihin.

Valtakunnallisia oppimistulosarviointeja on tehty otosperustaisina niin, että tulokset voidaan yleistää koskemaan koko ikäluokkaa. Ne mahdollistavat tarvittavan vertailun sukupuolten kesken, alueittain entisen läänijaon mukaan, erityyppisten kuntien sekä suomen- ja ruotsinkielisten koulujen tulosten välillä. Käytetyt otokset ovat niin suuria, että koko ikäluokan osaamisen yleistaso voidaan arvioida riittävällä tarkkuudella. Arviointiin osallistuu noin 5–10 % ikäluokasta ja noin 15 % kouluista. Ruotsinkielisestä koulutuksesta otetaan suhteellisesti suurempi otos vertailtavuuden parantamiseksi. Arviointi järjestetään yleensä 4 000–6 000 oppilaalle.

1.2.1 Perusopetuksen opetussuunnitelman perusteet osana ohjausjärjestelmää

Koulujen toiminnan suunnittelua ja toteutusta ohjataan erityisesti opetussuunnitelmalla, jossa yleensä ilmaistaan kyseisen koulun tai kouluasteen tavoitteet ja sisällöt sekä oppilasarvioinnin periaatteet. Koulukohtaisten opetussuunnitelmien taustalla ovat valtakunnalliset opetussuunnitelman perusteet, joiden rakenne on säilynyt vuosikymmenten ajan suhteellisen samanlaisena. Opetussuunnitelman perusteissa linjataan annettavan opetuksen keskeiset tavoitteet, kuvataan oppiaineet sekä niiden tavoitteet ja sisällöt. Painotus pedagogiikan ja oppiainesisältöjen suhteen on vaihdellut eri aikoina, samoin se, kuinka ohjaavia opetussuunnitelman perusteet ovat eri aikakausina olleet. Opetussuunnitelmien rakenne samoin kuin niiden malli on pysynyt uudistuksesta toiseen kuitenkin hyvin samanlaisena. (Vitikka 2009, 23.)

Vitikan (2009) mukaan Perusopetuksen opetussuunnitelman perusteet (2004) ovat kiinteä osa ohjausjärjestelmää, sillä se on koulutyötä, opetusta ja opettajan työtä ohjaava asiakirja (Vitikka 2009, 23, 65–66). Opetussuunnitelma kytkeytyy ohjausjärjestelmään esimerkiksi oppikirjojen kautta, jotka Heinosen tutkimuksen mukaan ohjaavat keskeisesti opetuksen suunnittelua ja toteutusta (Heinonen 2005, 57), vaikka niiden luonnetta ohjausjärjestelmään kuuluvana osana ei aina tule ajatelleeksi (Vitikka 2009, 67). Heinosen (2005) tutkimus osoitti myös, että koulujen opetussuunnitelmien sisällöt olivat usein täysin yhtenevät oppikirjojen sisältöjen kanssa ja monissa kouluissa oli jopa muutettu omia, koulukohtaisia suunnitelmia vastaamaan enemmän oppikirjojen sisältöjä (Heinonen 2005, 5). Jo edellisen, vuoden 1994 perusopetuksen opetussuunnitelman perusteiden ollessa voimassa, Korkeakoski (1997) oli tehnyt samanlaisen huomion kuin Heinonen siitä, että opettajan päivittäistä työtä peruskoulussa eivät niinkään ohjaa kasvatustavoitteet ja opetussuunnitelma, vaan oppiaineiden yleistavoitteet ja sisällöt, joita ohjataan enemmän oppimateriaalien kuin opetussuunnitelmien avulla (Korkeakoski 1997, 116).

Opetussuunnitelman avulla on määritelty koulutuksen tavoitteet ja pyritty takaamaan koulutukselle asetettujen tavoitteiden toteuttaminen (Rokka 2011, 36). Suomalaisessa koulutusjärjestelmässä opetussuunnitelman perusteet kiinnittyvät sekä hallintoon että didaktiikkaan. Opetussuunnitelman hallinnollista puolta edustavat poliittinen päätöksenteko sekä lakien ja asetusten laatiminen. (Rauste-von Wright & Wright 1994, 145.) Didaktinen osaaminen ilmenee esimerkiksi siinä, että opetussuunnitelmassa esitetään opetuksen tarkoitus (Kansanen 2004, 25). Virkamiesvalta tulee puolestaan esiin opetussuunnitelmaprosesseissa, sillä Suomessa opetussuunnitelmien valmistelusta ja perusteiden laadinnasta vastasi vuoteen 1991 Kouluhallitus ja tämän jälkeen sitä on ohjannut ja ohjaa edelleen Opetushallitus. Opetussuunnitelman perusteissa pyritään ennakoimaan niitä tietoja, taitoja ja valmiuksia, joita oppilaat tarvitsevat siirtyessään jatko-opintoihin ja työelämään. Siksi opetussuunnitelmien perusteissa määritellään paitsi yleiset kasvatukselliset myös opetukselliset tavoitteet.

Malisen mukaan tutkijan tulisi määritellä käyttämänsä opetussuunnitelman käsite (Malinen 1989, 7), vaikka opetussuunnitelma on käsitteenä kompleksinen (Heinonen 2005, 26; ks. Autio 2002, 160; Rokka 2011, 35) ja jopa ristiriitainen (Autio 2007, 43). Opetussuunnitelmakäsitteen määrittely ilmentää kirjoittajan valitsemaa näkökulmaa ja käsitteen käyttötilanne määrittää sen, miksi jokin toinen opetussuunnitelmaa kuvaava määritelmä on parempi kuin jokin toinen (Kosunen & Huusko 2002, 202; ks. Rokka 2011, 36). Vaikka opetussuunnitelman luonne tekee sen rajaamisen ja yksiselitteisen määrittelyn vaikeaksi (Konttinen, Suortti & Väyrynen, 1980, 12–13), voidaan opetussuunnitelmaa kuvata esimerkiksi niin, että se on aikaan, paikkaan ja yhteiskuntarakenteisiin sidottu kouluinstituution ohjausväline tai niin, että se on toiminnallinen dokumentti ja opettajan työväline (Atjonen 1993, 15; Kosunen 2002; Rinne 1984, 95–101). Opetussuunnitelma on toisaalta yhteiskunnallis-poliittinen ohjausväline (McKernan 2008, 5), toisaalta opetuksen työväline (Vitikka 2009, 50). Opetussuunnitelma voidaan määritellä etukäteissuunnitelmaksi toimista, joilla pyritään toteuttamaan koululle asetettuja kasvatustavoitteita (Kosunen & Huusko 2002, 203).

Tässä tutkimuksessa opetussuunnitelmalla viitataan peruskoulun ja perusopetuksen opetussuunnitelmien perusteisiin opetusta ja oppilaan arviointia normittavina ja ohjaavina asiakirjoina. Peruskoulun opetussuunnitelmien perusteilla tarkoitetaan peruskoulun opetussuunnitelmakomitean mietintöjä (1970)⁷ sekä vuosien 1985 ja 1994 opetussuunnitelmien perusteita, ja vuoden 2004 perusteista käytetään puolestaan sen virallista nimeä eli Perusopetuksen opetussuunnitelman perusteet tai lyhennetysti perusopetuksen opetussuunnitelma.

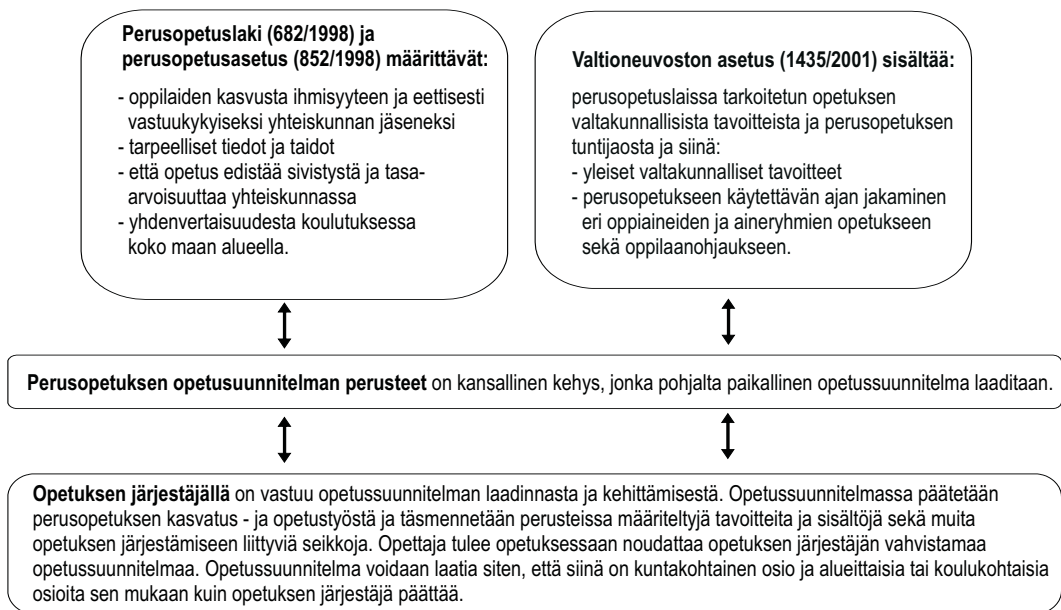
7 Komiteamietintöjen pohjalta vahvistettiin vuonna 1972 valtakunnalliset peruskoulun opetussuunnitelman perusteet.

Peruskoulu-uudistuksen myötä vuonna 1970 peruskoulun opetussuunnitelmakomitean mietinnöissä (Komiteanmietintö 1970: A4; Komiteanmietintö 1970: A5) luotiin perusta sille muodolle, jollaisena perusopetuksen opetussuunnitelmien perusteet ovat näyttäytyneet usean vuosikymmenen ajan (POPS 1985; POPS 1994; POPS 2004). Suomalainen perusopetuksen opetussuunnitelma on yhä tavoite- ja sisältöpainotteinen, siitä huolimatta, että siinä määrätään sekä koulun toimintakulttuurista että kasvatustehtävästä. Jo ensimmäisestä peruskoulun opetussuunnitelmasta lähtien valtakunnallisessa ohjausasiakirjassa on näkynyt myös oppiainejakoisuus. Opetussuunnitelman perusteissa otetaan kantaa opetusmenetelmiin, vaikka niiden valinnasta ja hyödyntämisestä opetuksessa opettajat päättävät vapaasti (Uusikylä & Atjonen 2000, 46). Vuoden 2004 Perusopetuksen opetussuunnitelman perusteissa käsitellään myös oppimis- ja tiedonkäsitystä, oppimisympäristöä ja toimintakulttuuria (POPS 2004, 16–17).

Tämän tutkimuksen kannalta on oleellista ymmärtää, miten voimassa oleva lainsäädäntö ja Perusopetuksen opetussuunnitelman perusteet (2004) sekä oppimistulosten seuranta-arvioinnit suhteutuvat toisiinsa ja muodostavat osan perusopetuksen ohjausmekanismista. Voimassa oleva perusopetuslaki (628/1998) ja asetus (825/1998) edellyttävät Opetushallituksen vahvistamien valtakunnallisten opetussuunnitelman perusteiden noudattamista. Koulutuksen järjestäjä laatii oman kunta-, koulu- tai oppilaitoskohtaisen opetussuunnitelmansa valtakunnallisten perusteiden pohjalta.

Suomalaisessa koulutusjärjestelmässä opetussuunnitelma on sekä hallinnollinen että pedagoginen ohjausväline, jolla säädellään koulujen toimintaa. Opetussuunnitelma on nykyisin luonteeltaan normatiivinen ja sen määräykset velvoittavat koulutuksen järjestäjiä. Opetushallitus laatii ja vahvistaa opetussuunnitelmien valtakunnalliset perusteet, joissa on oppiaineittain määritellyt tavoitteet, sisällöt sekä vuoden 2004 jälkeen oppiainekohtainen arviointi hyvän osaamisen kriteerien avulla. Se, kuinka paljon perusopetuksen eri oppiaineissa annetaan opetusta, määritellään valtioneuvoston tuntijakopäätöksessä.

Perusopetuksen järjestämistä ohjataan opetusta koskevan lainsäädännön, opetussuunnitelman valtakunnallisten perusteiden sekä paikallisen opetussuunnitelman avulla, jota havainnollistaa kuvio 1.



KUVIO 1. Perusopetuksen ohjusjärjestelmä. Lähde: Numminen & Ouakrim-Soivio 2009, 25.

1990-luvun lopulta lähtien on voimistunut näkemys siitä, että oppimistulosten seuranta-arvioinnit ovat saaneet yhä virallisemmän aseman koulutuspolitiikassa ja hallinnollisessa ohjauksessa (Linnakylä 2002, 33). Samansuuntainen näkemys oli myös entisellä Opetushallituksen pääjohtajalla Jukka Sarjalalla, jonka toteamus Perusopetuksen opetussuunnitelman perusteita (2004) valmisteltaessa kuvaa hyvin sitä, miten oppimistulosten seuranta-arvioinneista saadut tulokset vaikuttivat esimerkiksi opetussuunnitelmaprosessiin ja siihen, missä muodossa nykyisin voimassa oleva Perusopetuksen opetussuunnitelman perusteet (2004) kirjoitettiin ja julkaistiin. Sarjala totesi arviointitulosten ja käytännön kokemusten antaneen aiheita tarkentaa opetussuunnitelmien täsmällisyyttä, esimerkiksi selkeämpien vaatimustasokuvausten avulla. (Sarjala 2002, 13.) Hän korosti myös arviointitoiminnan koulutuspoliittista merkitystä luotettavan tiedon saamisessa ja siinä, miten valtakunnalliset tavoitteet on saavutettu (Sarjala 2002, 11).

Opetushallitusta on aika ajoin arvosteltu sen kaksoisroolista. Toisaalta sen tehtäviin kuuluu laatia perusopetuksen opetussuunnitelman valtakunnalliset perusteet, toisaalta taas arvioida perusteiden toteutumista ja niiden toimivuutta käytännössä eri oppiaineiden oppimistulosten seuranta-arviointien avulla (Lankinen 2007, 188–189).

1.2.2 Oppilaan arviointi ja arvosanat

Arvosanan antaminen (englanniksi *grading*, ruotsiksi *att ge vitsord*, ranskaksi *notation*) on osa opintojen aikaista arviointia tai päättöarviointia. Numeroarvosana kuvaa osaamisen tasoa, sanallisella arvioinnilla voidaan kuvata oppilaan edistymistä ja oppimisprosessia. Oppilaan arvioinnissa pyritään erottamaan toisistaan osaamisen ja edistymisen arviointi, joista annettava arviointipalautte on erilainen. Osamista arvioidaan numerolla tai kirjaimella ja edistymistä arvioidaan usein sanallisella palautteella. Kumpikin arviointi, sekä numeroarviointi että sanallinen arviointi, perustuvat oppiaineittain määriteltyihin kuvauksiin oppilaan hyvästä osaamisesta opintojen nivel- tai päättövaiheessa. Oppiainekohtaiset päättöarvioinnin kriteerit määrittelevät kansallisesti oppilaan arvioinnin tieto- ja taitotason, johon oppilaan arviointi perustuu. Oppilaan arviointi toteutetaan koko maassa samojen kriteerien pohjalta. (Vitikka 2009, 159.)

Suomessa oppilaan arvioinnilla on toteava, motivoiva, ohjaava ja ennustava tehtävä (Koppinen, Korpinen, Pollari 1994; Helakorpi & Olkinuora 1997; Lahdes 1997; Törmä 1998, 11; Atjonen 2007, 66). Oppilaan arviointi voi olla tyypiltään diagnostista, prognostista, formatiivista tai summatiivista siihen perustuen, miksi ja milloin arvioidaan, mihin kriteereihin arviointi perustuu, miten arviointi suoritetaan ja kenelle arviointipalautetta annetaan (Linnakylä & Välijärvi 2005, 26). Diagnostisella arvioinnilla kartoitetaan usein oppilaan tai opiskelijan oppimisedellytyksiä. Prognostisessa arvioinnissa tarkastellaan oppilaan tai opiskelijan oppimisedellytyksiä hänen tulevan kehityksensä ja menestymisensä näkökulmasta. Formatiivisella arvioinnilla hankitaan tietoa opiskelun aikana siitä, miten oppilas tai opiskelija on edistynyt suhteessa asetettuihin tavoitteisiin, kun taas summatiivisella arvioinnilla kootaan jälkepäin tietoa siitä, miten asetetut tavoitteet on saavutettu (Heinonen 2001a, 22; Metsämuuronen 2009, 8).

Arviointi on mukana kaikissa oppimisen ja opettamisen vaiheissa, sillä arvioinnin tehtävänä on sekä tukea että edistää oppimista ja opettamista (Koppinen, Korpinen & Pollari 1994, 8). Aiemmin arviointikäsitteestä erotettiin oppilasarvostelun käsite. Arvostelulla tarkoitettiin arvosanan eli numeron tai kirjaimen tai sanallisen arvostelun antamista suoritusten perusteella (Koppinen, Korpinen & Pollari 1994, 10). Voimassa olevat säädökset puhuvat oppilaan arvioinnista myös silloin, kun tarkoitetaan esimerkiksi arvosanan antamista (POPS 2004, 260–265).

Vaikka arviointi-käsitettä käytetään käyttäytymistieteissä ja koulukielessä useissa eri merkityksissä, tässä tutkimuksessa oppilaan arvioinnin käsitteellä ja arvosanan antamisen käsitteellä on kaksi, selkeästi eri merkitystä. Oppilaan arviointi on käsitteenä laajempi kuin arvosanan antaminen. Vain osa oppilaan arvioinnista liittyy arvosanan antamiseen. Oppilaan arvioinnilla tarkoitetaan tässä tutkimuksessa koko sitä arviointiprosessia, jonka tarkoituksena on tukea oppilaan oppimista ja antaa oppilaalle, opettajalle ja huoltajalle tietoa oppilaan edistymisestä. Arvioinnilla pyritään

toisaalta tukemaan oppilaan oppimisprosessia, toisaalta antamaan opettajalle palautetta siitä, miten oppilaat ovat eri tietoja ja taitoja omaksuneet. Arvioinnin tulisi kuvata realistisesti, mutta kannustavasti oppilaan saavuttamia tietoja, taitoja ja edistymistä. Arvosanan antaminen perustuu yhteisiin säädöksiin ja normeihin siitä, miten asetettuja tavoitteita sekä opittuja tietoja ja taitoja arvioidaan. Oppilaan arvioinnin yksi keskeisimmistä tehtävistä liittyy kuitenkin arvosanan antamiseen opiskelun aikana sekä opintojen päättövaiheessa. (POPS 2004, 260–265.)

Suomalaisessa opetussuunnitelman perusteissa oppilaan arviointi jaetaan opintojen aikaiseen arviointiin ja päättöarviointiin (POPS 2004, 260–265; Vitikka & Hurmerinta 2011, 62). Opintojen aikaisessa arvioinnissa kuvataan arvioinnin tehtävät, periaatteet, opinnoissa eteneminen, arvioitavat oppiaineet, työskentelyn arviointi, käyttäytymisen arviointi, oppilaan itsearviointi, erityistä tukea tarvitsevan oppilaan arviointi sekä maahanmuuttajaoppilaan arviointi (POPS 2004, 260–268). Opintojen aikaisen arvioinnin tehtävä on oppilaan persoonallisen kasvun tukeminen auttamalla oppilasta muodostamaan realistinen kuva oppimisestaan ja kehittymisestään. Arvioinnin periaatteissa esitetään, että arvioinnin tulee kohdistua oppilaan oppimiseen ja edistymiseen oppimisen eri osa-alueilla. Lisäksi todetaan, että oppilaan edistymistä, työskentelyä ja käyttäytymistä arvioidaan suhteessa opetussuunnitelman tavoitteisiin ja kuvauksiin oppilaan hyvästä osaamisesta. (Vitikka 2009, 159.)

Oppilaan arviointi perustuu voimassa olevaan lainsäädäntöön sekä muihin opetuksen järjestäjiä, kouluja ja opettajia sitoviin normeihin, kuten Perusopetuksen opetussuunnitelman perusteisiin (2004), jossa on määritelty oppilasarvioinnin yleiset periaatteet, ainekohtaiset päättöarvioinnin kriteerit ja niihin perustuvat hyvän osaamisen kuvaukset. Vuoden 2004 Perusopetuksen opetussuunnitelman perusteissa on oppiainekohtaisen hyvän osaamisen kuvausten ja päättöarvioinnin kriteerien lisäksi määryksiä oppilaan arvioinnin yleisistä periaatteista.

1.2.3 Seuranta-arvioinnin lähtökohdat

Suomen kielessä arviointi-sana kattaa laajasti kaikenlaisen arvioinnin. Arviointikäsitteellä kuvataan usein suorituksen tasoa, mittaamista tai palautetiedon antamista. Englannin kielessä vastaava merkitys on käsitteellä *assessment*, ruotsissa *bedömning*-sanalla (Opetushallitus 1998, 8; Heinonen 2001a, 22) ja ranskassa sanalla *évaluation*.

Opetushallituksen oppimistulosten seuranta-arvioinnit (englanniksi *assessment of learning outcomes*) ovat luonteeltaan summatiivisia (Koulutuksen tuloksellisuuden arviointimalli 1998, 56). Summatiivista arviointia käytetään yleensä kehittämistar-koitukseen. Seuranta-arvioinnit tuottavat perusopetuksen nivel- tai päättövaiheessa tietoa koulutusjärjestelmän ja opetussuunnitelmien kehittämiseksi (Jakku-Sihvonen 2010, 318). Arviointi voi olla myös sisäistä, kuten esimerkiksi luokassa opettajan aloitteesta tapahtuvaa, tai ulkoista, kuten Opetushallituksen seuranta-arvioinnit (Heinonen 2001a, 23; Metsämuuronen 2009,8). Ulkoisella arvioinnilla tarkoitetaan yleensä, että arvioija on arvioitavan toiminnan tai arvioitavan organisaation ulkopuolelta (Koulutuksen tuloksellisuuden arviointimalli 1998, 14).

Suomessa oppilaiden osaamisen arvioiminen on perustunut ns. koululähtöiseen arviointiin (*School-Based Assessment, SBA*), kun taas esimerkiksi Yhdysvalloissa tai Englannissa käytössä on tutkintoperusteinen järjestelmä (*Examination System, ES*). Koululähtöisessä arvioinnissa opettaja arvioi formatiivisesti oppilaan edistymistä ja ymmärtämistä. Formattiivista oppimistulosten arviointia toteutetaan opetuksen kuluessa ja kun halutaan kehittää tai vaikuttaa oppimisprosessiin. (Heinonen 2001a, 22; Brookhart 2007, 43; Metsämuuronen 2009, 8.) Opettaja voi formatiivisen arvioinnin avulla mukauttaa opetustaan ja opetusmenetelmiään niin, että ne sopivat paremmin oppilaille havaittuihin tarpeisiin. Summatiivisten arviointien kautta valvotaan esimerkiksi sitä, mitä oppilaat ovat oppineet. (Metsämuuronen 2009, 9.)

Oppimistulosten mittaaminen voi olla joko suhteellista eli normiperusteista (*norm-referenced*) tai kriteeriperusteista (*criterion based*). Normiperusteisella tarkoitetaan sitä, että yksittäisen oppilaan tulosta verrataan vertailuryhmän muodostamaan normiin. Opetushallituksen oppimistulosarviointi on normiperusteista sikäli, että yksittäisen oppilaan tuloksia verrataan muiden samaan otokseen kuuluneiden oppilaiden tuloksiin eli näiden muodostamaan normiin. Seuranta-arvioinnissa hyödynnettävien tehtäväsarjojen vaikeustaso määritellään esitestattujen osioiden perusteella etukäteen siten, että seuranta-arvioinnista tulisi mahdollisimman erotteleva. Kriteeriperusteinen mittaaminen perustuu ennalta asetettuihin ulkoisiin kriteereihin, jotka joko saavutetaan tai niitä ei saavuteta. Koska oppimistulosten seuranta-arviointien tehtäväsarjat rakentuvat oppiaineen tavoitteisiin ja päättöarvioinnin kriteereihin arvosanalle 8, voidaan Opetushallituksen seuranta-arviointeja pitää myös osittain kriteeriperusteisina. (Metsämuuronen 2009, 17.)

1.2.4 Yhteiskunnallisten aineiden oppimistulosten arviointi osana perusopetuksen ohjausjärjestelmää

Tässä tutkimuksessa yhteiskuntaopilla tarkoitetaan oppiainetta, jonka opetusta kouluissa annetaan yleensä perusopetuksen 9. vuosiluokalla. Yhteiskunnallisilla aineilla viitataan tässä tutkimuksessa Opetushallituksen oppimistulosten seuranta-arvioinnissa historialle ja yhteiskuntaopille annettuun nimitykseen, jolla se määritteli yhteisellä nimellä näiden kahden oppiaineen yhteisen seuranta-arvioinnin. Yhteiskunnalliset aineet -nimitys lienee siirtynyt Opetusministeriön koulutuksen arviointisuunnitelmasta, jossa se määritteli vuosina 2010–2011 arvioitavaksi yhteiskunnalliset oppiaineet (Opetusministeriö 2008, 9). Yhdysvalloissa käytetään yhteiskunnallisista aineista nimitystä *social studies*, ja siihen kuuluvat historian ja maantiedon lisäksi psykologia, yhteiskuntaoppi ja Yhdysvaltain hallinto. Ruotsissa yhteiskunnallisten aineiden ryhmän (*sambällsorienterande ämnen*) muodostavat peruskoulussa historia, maantieto, uskontotieto ja yhteiskuntaoppi. Edellisten käsitteiden lisäksi on hyvä määritellä yhteiskunnallinen opetus, joka ei ole yhden yksittäisen oppiaineen sisältämää opetusta, vaan sitä annetaan useissa eri oppiaineissa ja osin kouluopetuksen ulkopuolella. Suutarisen mukaan yhteiskunnallista opetusta voidaan kutsua laajasti myös yhteiskunnalliseksi kasvatukseksi (Suutarinen 2000, 35).

Yhteiskunnallisten aineiden seuranta-arviointi noudatti Opetushallituksessa jo vakiintunutta ulkoista prosessia, joka yleensä käynnistyy poliittisen päätöksenteon pohjalta, kuten ministeriön toimeksiannosta ja päättyy seuranta-arviointiraportin luovuttamiseen (Metsämuuronen 2009, 10). Opetusministeriö linjasi Koulutuksen arviointisuunnitelmassa vuosille 2009–2011 yhdeksi toteutettavista seuranta-arvioinneista yhteiskunnallisten aineiden arvioinnin. Kyseisen seuranta-arvioinnin valmistelu alkoi Opetushallituksessa keväällä 2010.

Historian ja yhteiskuntaopin seuranta-arviointi noudatti Opetushallituksessa vakiintunutta oppimistulosten arviointihankkeen mallia (Metsämuuronen 2009, 10). Otoksen valmistelu, kysymysten ja mittariston laadinta sekä aineiston kuvaaminen ja kyseisten oppiaineiden osaamistulokset on raportoitu yksityiskohtaisesti Opetushallituksen julkaisemassa koulutuksen seurantaraportissa (2012:3) Historian ja yhteiskuntaopin oppimistulokset perusopetuksen päättövaiheessa 2011 (Ouakrim-Soivio & Kuusela 2012).

Oppimistulosten arvioinnissa voidaan Metsämuurosen mukaan erottaa ulkoisen prosessin lisäksi sisäinen prosessi, joka liittyy metodiikkaan ja käsittää seuranta-arvioinnissa käytettävän mittarin laadinnan sekä aineiston analysoinnin ja sen tulokintavaiheet (Metsämuuronen 2009, 9).

Opetushallituksen oppimistulosten sisäiseen prosessiin kuuluvat:

- seuranta-arvioinnin tehtäväsarjan esivalmistelu, asiantuntijoiden kuuleminen ja opetussuunnitelmaan perustuvien painotusten valinta
- otannan valmistelu, tehtäväosioiden laadinta ja muokkaaminen
- tehtävien esitetaus ja alustava osioanalyysi
- validiteettitarkastelut ja mittarin rakentaminen
- mittaus eli varsinainen seuranta-arvioinnin toteuttaminen, aineiston syöttö ja tarkistus sekä pikapalautteen antaminen otoskouluille
- lopullinen osioanalyysi, reliabiliteettitarkastelut ja osioiden pankittaminen (Metsämuuronen 2006, 115, 180–183; 2009, 17.)

Sisäisen prosessin viimeinen vaihe on oppimistulosten raportointi ja tuloksista tiedottaminen. Edellä esitettyä sisäistä ja ulkoista prosessia noudatettiin myös yhteiskunnallisten aineiden seuranta-arvioinnissa (Ouakrim-Soivio & Kuusela 2012, 15–16). Oppimistulosten arviointien tarkoituksena on antaa luotettavaa tietoa osaamisen yleisestä tasosta perusopetuksen nivel- ja päättövaiheessa. Oppimistulosten arviointeihin laadittavat tehtävät perustuvat Perusopetuksen opetussuunnitelman perusteisiin (2004) sekä siinä määriteltäviin tavoitteisiin ja sisältöihin.

Peruskoulun opetussuunnitelman perusteet (1994) ja Perusopetuksen opetussuunnitelman perusteet (2004) sekä vuodesta 1998 alkaen tehdyt eri oppiaineiden seuranta-arvioinnit liittyvät kiinteästi toisiinsa, sillä seuranta-arviointi antaa tietoa koulutuksen tilasta esimerkiksi niille, jotka valmistelevat tai päättävät suomalaisen koulutuspolitiikan linjauksista ja koulutussektorilla harjoitetusta politiikasta (Heinonen 2001b, 54). Poliittiset päätöksentekijät tarvitsevat arviointitietoa poliittisten ratkaisujen ja niistä johdettujen toimenpiteiden toteutumisesta ja tuloksista.

Pickford ja Brown ovat esittäneet, että opetussuunnitelmaprosessin tulisi toimia siten, että ensin tutkittaisiin oppimistulokset, joista johdettaisiin opetussuunnitelman perusteisiin oppiaineiden tavoitteet ja sisällöt (Pickford & Brown 2006, 21). Suomalaisessa perusopetuksen ohjausjärjestelmässä arviointitietoa hyödynnetään poliittisten päätösten valmistelussa (Niemi 2009, 8–9). Käytännössä se on tarkoittanut sitä, että esimerkiksi vuoden 2001 ja 2012 tuntijakotyöryhmien valmistelutyössä on hyödynnetty eri oppiaineiden seuranta-arviointien tuloksia, joiden avulla tuntijakoa valmistelleet virkamiehet ja poliittiset päättäjät ovat saaneet ajantasaista tietoa opetussuunnitelman perusteiden ja ohjausjärjestelmän toimivuudesta.

Vuoden 2001 tuntijakopäätöstä valmistellut työryhmä jätti oman muistionsa (Opetusministeriö 2001:11), jossa oli jo hyödynnetty muutaman vuoden ajan Opetushallituksen toimesta kerättyjä oppimistulosten arviointeja. Äidinkielessä pidettiin huolestuttavana sitä, että ammatilliseen koulutukseen hakeutuvista pojista kaksi viidestä ei ollut saavuttanut tyydyttävää luku- ja kirjoitustaitoa. Lukutaito oli etelässä

selvästi parempaa kuin pohjoisessa. Matematiikassa ei esiintynyt merkittäviä oppimistulosten eroja koulujen eikä eri alueiden välillä, mutta yli viidenneksellä oppilasta on selviä puutteita matematiikan perustaidoissa. (Opetusministeriö 2001, 21.) Opetusministeriön tuntijakotyöryhmä esitti vuonna 2001 äidinkielen ja kirjallisuuden viikkotuntikertymän vähimmäistuntimäärän korottamista kahdella tunnilla (Opetusministeriö 2001, 26). Matematiikalle esitettiin lisättäväksi yksi vuosihiikotunti, koska seuranta-arviointitulokset osoittivat, että osa oppilasta hallitsi matematiikan perustaitoja heikosti (Opetusministeriö 2001, 28).

Opetushallituksen seuranta-arviointien vakiintuminen osaksi koulutusjärjestelmän toimivuuden arviointia näkyy siinä, että kymmenen vuotta myöhemmin vuoden 2012 tuntijakoesityksen, Tulevaisuuden perusopetus – valtakunnalliset tavoitteet ja tuntijako, perusteluteksteissä oli tarkasteltu niiden perusopetuksessa opettavien aineiden seuranta-arviointituloksia (Opetus- ja kulttuuriministeriö 2012, 15–27), joita oli tehty eri oppiaineissa 2000-luvun aikana. Poikkeuksina olivat luonnontieteet, joiden seuranta-arviointi oli tehty edellisen kerran vuonna 1998, aihekokonaisuudet sekä yhteiskunnalliset aineet, jotka arvioitiin ensimmäisen kerran keväällä 2011. Koska historian ja yhteiskuntaopin oppiaineista ei ollut aiemmin seuranta-arviointitietoa asiantuntijavirkamiehille ja poliittisille päättäjille, saatettiin yhteiskunnallisten aineiden seuranta-arvioinneissa nousseet keskeiset tulokset (Ouakrim-Soivio & Kuusela 2012, 133–135) heille tiedoksi jo syksyllä 2011 sekä vielä laajemmin seuranta-arviointiraportin ollessa jo viimeistelyvaiheessa tammikuussa 2012. Samalla tavalla toimittiin myös aihekokonaisuuksien arvioinnin, matematiikan ja luonnontieteiden seuranta-arviointien kanssa, joiden tulokset julkaistiin vasta tuntijakotyöryhmän esityksen jälkeen.

1.3 Aiempia tutkimuksia

Yhteiskuntaoppia oppiaineena on Suomessa tutkittu verrattain vähän (Löfström 2008, 28; Löfström 2012b, 173). Syynä yhteiskuntaopin tutkimuksen vähäisyyteen saattaa olla oppiaineen heikko asema (Löfström 2012b, 172) tai se, että Suomessa yhteiskuntaoppi on eriytynyt omaksi oppiaineekseen vasta 2000-luvun alussa. Myös sillä, että edelleen suurimmalla osalla yhteiskuntaopin opettajista on laudatur-tasoiset opinnot historiasta ja approbatur- tai cum laude -tasoisia opintoja yhteiskuntaopista, saattaa vaikuttaa taustalla (Löfström 2001a, 15). Esimerkkinä historian ja yhteiskuntaopin tiivistä oppiainekytöksestä on Virran (1995) väitöstutkimus, jossa hän tutki historian ja yhteiskuntaopin tietoa abiturienttien reaali-koetustauksissa. Virran tutkimuksessa käy ilmi, että hyvissäkin yhteiskuntaopin vastauksissa lainattiin ja sovellettiin historian tiedonrakenteita (Virta 1995, 173–174.) Artikkelissaan (1993) Virta on tarkastellut yhteiskuntaopin oppilasarviointia osana yhteiskuntaopin oppimista ja pohtinut laajemmin arvioinnin kehittämistä ja uudistamista myös muissa julkaisuissaan (Virta 1999; 2006).

Vaikka yhteiskuntaopissa tai historiassa ei ole tehty oppiaineiden oppimistuloksia kartoittavaa seuranta-arviointia ennen huhtikuuta 2011, on Opetushallitus julkaissut 1990-luvun lopulla ja 2000-luvun alussa joitain arviointeja, joissa sivuttiin nuorten historian tai yhteiskuntaopin tietoja, taitoja ja niiden soveltamista. Esimerkiksi Opetushallituksen (1996) julkaisussa ”Toteuttaako peruskoulu tasa-arvoa?” arvioinnin yhtenä painopisteenä oli peruskoulun yhteiskunnallinen vaikuttavuus sekä oppimistulosten analysointi oppiaineittain ja aihealueittain. Historiassa tutkittiin, miten oppimistulokset ilmentävät tasa-arvoa. Arvioinnin tuloksena historiassa nousi esiin historiatietoisuuden näkökulma, jossa on kyse paitsi siitä, miten ihmiset samaistuvat menneisyyteen, myös siitä, miten he suhtautuvat yhteiskunnan instituutioihin ja niiden toimintaan. (Ahonen 1996, 376.)

Vuonna 1998 julkaistiin Ahosen Suomessa toteuttama tutkimus, joka oli osa kansainvälistä Youth and history -tutkimusta (Ahonen 1998). Eurooppalaisten nuorten historiatietoisuutta kartoitettiin aineistolla, joka kerättiin kaikkiaan 29 eri maasta. Tutkimukseen osallistuneet olivat 1980-luvun alussa syntyneitä, haastatteluhetkellä 15–16-vuotiaita nuoria. Tutkimuksen tulokset niin Suomessa kuin muuallakin Euroopassa olivat samansuuntaisia: politiikka ei kiinnostanut eikä siihen haluttu osallistua. Ahosen mukaan tutkimustulokset kuvastivat sitä, että politiikka käsitteenä oli muuttumassa, sillä nuoret eivät tutkimuksen mukaan sitoutuneet puolueisiin tai vaaleihin. Tutkimuksessa kartoitettiin lisäksi nuorten käsityksiä demokratiasta, tasa-arvosta ja hyvinvointivaltioista, sosiaaliturvasta sekä maahanmuutosta. Youth and history -tutkimus osoitti, että suomalaisnuorille naisten ja miesten tasa-arvo oli 1990-luvun lopulla vähemmän tärkeää kuin muille pohjoismaalaisille tai länsieurooppalaisille nuorille. Pohjoismaalaiset ja etenkin suomalaiset nuoret olivat maahanmuuttajia kohtaan melko varautuneita. Myytti yltiöavoimesta pohjoismaisesta demokratiasta ei tutkimustulosten valossa enää näyttänyt pitävän paikkaansa. Tutkimus osoitti, että 1990-luvun puolivälissä Pohjoismaissa ja Suomessa nuoret pitivät hyvinvointia vielä tärkeämpänä kuin esimerkiksi Länsi-Euroopassa asuvat nuoret. Suomalaisnuorilla oli kuitenkin eurooppalaisia ikätovereitaan ristiriitaisempi asenne hyvinvointivaltioon, ja he suhtautuivat muita eurooppalaisia nuoria kovemmin heikompien auttamiseen. Suomalaisnuorten suhtautuminen hyvinvointivaltioon muistutti Ahosen mukaan keynesiläistä talousajattelua, ja hän arvelikin sen kumpuavan 1980- ja 1990-lukujen köyhyyttä paheksuvasta retoriikasta. Sosiaaliturvaa ja hyvinvointivaltion rakenteita nuoret pitivät saavutettuna etuna ja pysyvänä rakenteena, jonka puuttumista oli vaikea kuvitella. (Ahonen 1998, 153- 210.)

Vaikka tapakasvatuksen arviointi (2002) ei suoraan liity yhteiskunnallisten aineiden arviointiin, oli siinä joitain aihealueita, jotka taustoittivat sitä, mitä tietoja ja taitoja peruskoulun päättövaiheessa olleilla nuorilla oli 2000-luvun alun Suomesta, suomalaisista perinteistä sekä Suomessa tehdyistä rikoksista ja niistä annetuista rangaistuksista. Muutamissa yksittäisissä uskonnonvapauteen liittyvissä tehtävissä oppilaat saavuttivat jopa 95 %:n ratkaisusuuden, kun taas kansalais- ja ihmisoikeudet sekä poliittinen itsemääräämisoikeus tuottivat oppilaille ongelmia. Pojat osasivat vastata kansalais- ja ihmisoikeuksiin liittyviin tehtäviin tyttöjä paremmin, kun taas

suomalaisten perusoikeudet olivat paremmin tytöillä kuin pojilla tiedossa Suomea ja suomalaisuutta koskevassa kokonaisuudessa 9.-luokkalaisten ratkaisuosuus oli 56 %. Kyselyä tehtäessä Suomessa asui noin 85 000 ei-syntyperäistä suomalaista, ja esimerkiksi heidän osuutensa tiesi 31 % vastanneista oppilaista, ruotsinkielisten prosenttiosuuden koko väestöstä tiesi sen sijaan 58 % vastanneista. Rikoksia ja rangaistuksia käsitellessä tehtävissä hieman yli puolet oppilaista tiesi, että liikenteessä päin punaista valoa kävelemisestä voi saada sakkoja. Lähes saman verran piti sitä tekona, josta ei tule rangaistusta. Sen sijaan 90 % vastanneista tiesi, että television katselusta ilman maksettua tv-lupamaksua voi saada sakkoja. Lähes jokainen (95 %) vastanneista oppilaista tiesi, että huumeiden myynnistä seuraa sakkoja ankarampi rangaistus. Piraattituotteiden myymisen rangaistavuus tuotti sen sijaan vaikeuksia, sillä 2 % oppilaista jätti kokonaan vastaamatta sitä koskevaan kysymykseen ja tehtävän vastanneistakin vain puolet osui oikeaan. (Rusama 2002, 131–165.)

Tärkeän viitekehysten suomalaisten nuorten yhteiskuntaopin osaamisesta, yhteiskuntaopin soveltamisesta sekä yhteiskuntatiedollisista valmiuksista antavat kansainväliset CIVED-tutkimus (1999), ICCS-tutkimus (2010) ja OECD:n julkaisusarja Education at a Glance.

1.3.1 Yhteiskunnallisten aineiden kansainväliset tutkimukset CIVED ja ICCS

Suomi osallistui vuonna 1999 kansainväliseen Civic Education Study -tutkimukseen, joka suomeksi tunnetaan nimellä Nuori kansalainen eli CIVED-tutkimus. Suomalaisena vastuulaitoksena on toiminut Koulutuksen tutkimuslaitos, joka on osallistunut järjestön työskentelyyn 1950-luvun lopulta lähtien. Tutkimus tehtiin kansainvälisen koulutustutkimusjärjestö IEA:n johdolla, ja siinä noudatettiin tiukoja kansainvälisiä standardeja (IEA 1999), jotka edellyttivät muun muassa kaksivaiheista ryväotantaa ja tehtäväosioiden esitestausta sekä vaikeustason analysointia. Vuoden 1999 Civic Education Study oli toinen vaihe laajemmasta, kaksivaiheisesta tutkimuksesta. Ensimmäinen vaihe käynnistyi vuonna 1994, jolloin IEA päätti toteuttaa yhteiskunnalliseen opetukseen ja oppimiseen kohdistuvan tutkimuksen (Brunell 2002, 8).⁸ Tutkimuksen pääkenttävaihe toteutettiin huhti-toukokuussa 1999, ja se raportoitiin vuonna 2001.⁹ Kansallinen raportti julkaistiin vuonna 2001 nimellä Suomen peruskoulun päättövaiheessa olevien nuorten yhteiskunnalliset tiedot, taidot, käsitykset, asenteet ja toiminta 28 maan kansainvälisessä vertailussa (Suutarinen ym. 2001).¹⁰

8 Tutkimuksen ensimmäisen vaiheen aikana kerättiin laaja-alaista tietoa yhteiskunnallisen opetuksen tilanteesta, sisällöistä ja menetelmistä tutkimukseen osallistuneissa maissa. Ensimmäisen vaiheen tulokset on julkaistu teoksessa *Civic education across countries: twenty four national case studies from the IEA Civic Education Project* (Torney-Purta, Schwille & Amadeo 1999). Teos sisältää muun muassa Suomen osuuden.

9 Tutkimukselle on oma sivusto osoitteessa: <http://ktl.jyu.fi/arkisto/civics/civics.htm>

10 Toisen vaiheen otantaa, mittareiden kääntämistä, aineiston keruuta ja testauksen laadunvalvontaa on kuvattu teoksessa *Citizenship and Education in Twenty-Eight countries: Civic Knowledge and Engagement at Age Fourteen* (Torney-Purta, Lehman, Oswald & Schulz, 2001).

Vuonna 2009 tutkittiin kansainvälisessä ICCS-tutkimuksessa nuorten yhteiskunnallisia tietoja, osallistumista ja asenteita kaikkiaan 38 maassa. ICCS (*International Civic and Citizenship Education Study*) tunnetaan Suomessa nimellä ”Nuorten yhteiskunnallisen osaamisen, osallistumisen ja asenteiden tutkimus”. ICCS-tutkimus perustui keskeisiltä osin vuonna 1999 toteutettuun CIVED-tutkimukseen ja näin se mahdollisti osaamisen ja asenteiden muutoksen seuraamisen. Tutkimuksella pyrittiin kuvaamaan nuorten yhteiskuntaa koskevia tietoja, keskeisimpien ilmiöiden ja käsitteiden ymmärtämistä sekä valmiuksia osallistua yhteiskunnalliseen toimintaan ja oppilaiden asennoitumista kansalaisuuteen ja kansalaisyhteiskuntaan sekä heidän halukkuuttaan toimia sen aktiivisina jäseninä. (Suoninen, Kupari & Törmäkangas 2010, 143.) Vaikka suomalaisnuorilla oli hyvät tiedolliset valmiudet osallistua yhteiskunnalliseen toimintaan, he eivät olleet kansainvälisesti vertailun kiinnostuneita politiikasta tai yhteiskunnallisista kysymyksistä. ICCS-tulosten mukaan Suomi kuului niiden viiden maan joukkoon, joissa nuorten kiinnostus yhteiskunnallisia kysymyksiä kohtaan oli tilastollisesti merkitsevästi kansainvälistä keskiarvoa alhaisempi (Suoninen, Kupari, Törmäkangas 2010, 144). CIVED- ja ICCS-tutkimukset poikkeavat kahdella olennaisella tavalla yhteiskunnallisten aineiden seuranta-arvioinnista. Ensinnäkin CIVED- ja ICCS-tutkimuksia ei voida kohdentaa mihinkään erityiseen oppiaineeseen, kuten esimerkiksi yhteiskuntaoppiin tai minkään maan poliittiseen järjestelmään, eikä CIVED- ja ICCS-tutkimuksilla voida mitata sitä, miten opetussuunnitelmissa määritellyt tiedot ja taidot on saavutettu. Toiseksi sekä CIVED että ICCS tutkivat myös oppilaiden yhteiskunnallisia asenteita, mutta yhteiskunnallisten aineiden seuranta-arvioinnissa niitä ei puolestaan mitattu.

Nuorten yhteiskuntatiedollisesta osaamisesta ovat kirjoittaneet Löfströmin (2001; 2008; 2011; 2012) lisäksi Suutarinen (2000; 2001; 2002; 2006a; 2006b), van den Berg (2010, 2011) sekä Virta (1995; 1999; 2000; 2002; 2006; 2010). Keskustelua ja tutkimusartikkeleita ovat leimanneet vuoden 1999 CIVED-tutkimuksen tulokset sekä nuorten yhteiskunnallisen osaamisen, osallistumisen ja asenteiden ICCS-tutkimus vuodelta 2009, joiden mukaan suomalaisten nuorten tiedot ja taidot ovat maailman kärkitasoa, mutta esimerkiksi aktiivisuus yhteiskunnallisissa asenteissa on selvästi alle keskitason. Nuorten passiivisiin asenteisiin on löydetty syitä muun muassa opetussuunnitelmien, opetuksen ja oppikirjojen instituutiokeskeisyydestä ja yhteiskuntaopin opettajien tavasta opettaa yhteiskunnallisia asioita ikään kuin ne olisivat staattisia, ulkoapäin johdettuja ja jo päätettyjä asioita. (Suutarinen 2000; Virta 2000.)¹¹

11 Eränpalon mielestä nuorten passiivisista asenteista tulisi käydä keskustelua heidän elämismaailmastaan käsin. Luokan ilmapiiriä, ryhmadynamiikkaa ja vuorovaikutussuhteita kehittämällä voitaisiin luoda keskustelufoorumi, jossa ”uusi ja vanha” aktiivisuus voisivat kohdata. Eränpalo haastaa kehittämään yhteiskunnallista opetusta sosiaalisesti tiedostavammaksi. (Eränpalo 2012, 34.)

1.3.2 Oppimistulosten seuranta-arviointeihin ja arvosanan antamiseen liittyviä tutkimuksia

Opetushallitus on 2000-luvulla tehnyt säännöllisesti matematiikkaan (Korhonen 2001; Mattila 2002; Mattila 2005; Hirvonen 2012) sekä äidinkielen ja kirjallisuuden oppiaineiden päättöarviointeja (Lappalainen 2000; 2001; 2004; 2006; 2011). Perusopetuksen päättövaiheen seuranta-arviointeja on tehty 2000-luvulla myös A-kielissä: ranskassa, saksassa ja venäjässä (Väisänen 2003), ruotsin kielessä (Tuokko 2002; 2009), englannin kielessä (Tuokko 2003), uskonnossa, elämäntiedon tiedossa ja tapakasvatuksessa (Rusama 2002), liikunnassa (Huisman 2004; Palomäki & Heikinaro-Johansson 2011), musiikissa, kuvataiteessa ja käsityössä (Laitinen, Hilmola & Juntunen 2011), modersmållissa (Hannén 2001; Silverström 2004; 2006; Hellgren 2011), luonnontieteissä (Kärnä, Hakonen & Kuusela 2012) sekä historiassa ja yhteiskuntaopissa (Ouakrim-Soivio & Kuusela 2012).

Oppimistulosten seuranta-arviointeja ja niissä kerättyjä aineistoja on hyödynnetty esimerkiksi koulutuksen tasa-arvoon ja yhdenvertaisuuteen liittyvissä tutkimuksissa (Jakku-Sihvonen & Kuusela 2002; Kuusela 2003; 2006; Jakku-Sihvonen & Komulainen 2004; Gordon, Lahelma, Lyytinen, Niemivirta, Scheinin & Siimes 2004) sekä maahanmuuttajaoppilaiden oppimistulosten tarkasteluun (Kuusela, Etelälahti, Hagman, Hievanen, Karppinen, Nissilä, Rönnerberg & Siniharju, 2008). Rimpelän ja Berneliuksen (2010) toimittamassa monitieteellisessä tutkimuksessa on yhdistetty aiempaa tutkimustietoa sosiaalisesta eriytymisestä, oppimistuloksista ja oppilaiden hyvinvoinnista ja terveydestä pääkaupunkiseudulla. Kuusela (2010) on käyttänyt oppimistulosten arviointiaineistoa tutkiessaan oppilaiden sosioekonomisten taustatekijöiden ja siinä erityisesti huoltajien koulutustaustan yhteyttä oppimistuloksiin, ja Hautamäki (2010) on tutkinut oppilaiden, luokkien ja koulujen oppimaan oppimisen eroja.

Arvosanan antamisen ja oppiaineiden oppimistulosten välistä yhteyttä on Suomessa tutkittu vähän. Näin siitä huolimatta, että esimerkiksi Lappalainen (2004; 2006; 2010) on toistuvasti kiinnittänyt huomiota siihen, että koulujen antamat perusopetuksen päättöarvosanat ja Opetushallituksen seuranta-arviointitulokset eivät vastaa toisiaan. Kuusela on (2006) matematiikan oppimistulosten arviointiaineistolla osoittanut, että ”taipumus suunnata oppilasarviointia kohti keskimääräistä johtuu oman koulun käyttämisestä viiteryhmänä ja opettajaan kohdistuvista normidotuksista” (Kuusela 2006, 81). Hän osoitti matematiikan oppilasarvioinnista kerätyllä aineistolla, että suuri osa kouluista ankkuroi oppilasarviointinsa arvosanan 8 enemmän oman koulunsa keskiarvoon kuin Opetushallituksen kriteereihin (Kuusela 2006, 81). Kuusela on kehittänyt menetelmän (Kuusela 2006, 68; 73–76), jossa ”residuaalien sijasta oppilaille lasketaan todellisen standardipistemäärän ja heidän arvosanaansa vastaavan valtakunnallisen keskiarvon ero” (Kuusela 2006, 76). Tässä tutkimuksessa on käytetty Kuuselan kehittämää menetelmää, kun

oppilaan saamasta yhteiskuntaopin arvosanasta ja hänen yhteiskuntaopin seuranta-arvioinnissa osoittamastaan osaamisesta on muodostettu yhdistetty muuttuja mahdollisten systemaattisten osaamiserojen selvittämiseksi.

Mattila (2010) on käyttänyt lisensiaattityössään Opetushallituksen vuosina 1998–2005 keräämää matematiikan ja äidinkielen oppimistulosten arviointiaineistoa perusopetuksen päättövaiheessa. Tutkimustaan varten hän haastatteli otoskoulujen matematiikan ja äidinkielenopettajia ja opinto-ohjaajia heidän arvosanan antamiskäytänteistään. Mattilan tutkimuksen keskeisimpiä tuloksia oli, että arviointityylillä oli voimakas yhteys keskimääräiseen osaamistasoon koulussa ja eri vuosina arviointityylillä oli taipumus mukautua oppilaiden osaamistason vaihteluun samassakin koulussa. (Mattila 2010, 160.) Arviointityylit Mattila jakoi kuuteen ryhmään: ”tavanomainen, vaativa, lempeä, keskittävä, hajaannuttava ja armelias” (Mattila 2010, 159). Mattilan haastattelemat äidinkielen ja matematiikan opettajat kertoivat siitä, miten ja kuinka pitkältä ajalta he olivat ottaneet oppilaiden erilaisia suorituksia huomioon antaessaan oppineensa päättöarvosanaa (Mattila 2010, 92). Tutkimuksessa opettajat kuvasivat, miten tasovaikutuksena tunnettu arviointivirhe syntyi eli miten heikommassa ryhmässä tai ikäluokassa tietyn tasoinen oppilas sai korkeamman arvosanan kuin paremmassa ryhmässä opiskeleva oppilas (Mattila 2010, 160).

Eri opettajilla ja eri oppiaineissa käytänteet vaihtelivat huomattavasti siitä huolimatta, että opettajat noudattivat päättöarvosanan antamisperusteita ja opettajat tarkistivat, että he noudattivat päättöarvosanaa antaessaan Opetushallituksen määrittämiä päättöarvioinnin kriteereitä (Mattila 2010, 84). Mattilan haastattelemien opettajien mukaan päättöarvioinnin kriteereissä oli oppiainekohtaisia eroja. Opettajien mielestä matematiikan arviointi oli helpompaa kuin äidinkielen, koska valtakunnallisissa ohjeissa matematiikan tavoitteet oli ilmaistu selkeämmin ja yksityiskohtaisemmin kuin äidinkielessä (Mattila 2010, 167). Vaikka haastatellut opettajat totesivat yksimielisesti, että päättöarvosana ei peruskoulussa voi perustua ainoastaan kirjallisiin kokeisiin, suurin osa opettajista kertoi esimerkiksi koenumeroiden keskiarvon vaikuttaneen päättöarvosanaan. Antaessaan matematiikan tai äidinkielen päättöarvosanoja opettajat kertoivat laskeneensa koearvosanojen keskiarvojen lisäksi eri kurssiarvosanojen keskiarvot vuosiluokilla 7–9, mikäli oppiaine oli jaksotettu peruskoulun yläluokilla. (Mattila 2010, 91–92.)

Hildén on artikkelissaan (2008) pohtinut validiusargumentin (*validity argument*) rakentumista suullisen kielitaidon arviointiprojektissa. Artikkelissaan hän tarkastelee validiusargumentin avulla Perusopetuksen opetussuunnitelmien perusteisiin (2004) sisältyvää taitotasoasteikkoa. Arviointiasteikon ja sen pohjalta laadittujen puhetehtävien validiutta suullisen taidon mittareina hän koettelee Toulminin argumentaatiomallin avulla. Tässä tutkimuksessa on hyödynnetty Hildénin artikkelissaan käyttämiä validiusargumentointiin liittyvien termien käännöksiä (Hildén 2008, 757), kuten validiutta käsittelevässä luvussa 4 on esitetty.

Oppimistulosten arviointeihin liittyviä väitöstutkimuksia ovat tehneet muun muassa Lilja (2003), jonka tutkimus käsittelee peruskoulun matematiikan oppimistuloksiin yhteydessä olevia tekijöitä. Myös Näverin (2009) väitöstutkimuksessa hyödynnetään matematiikan oppimistuloksia, sillä hän on tutkinut, minkälaisia muutoksia peruskoulun päättövaiheen matematiikan osaamisessa on ollut havaittavissa viimeisen 20 vuoden aikana. Tutkimuksessaan Näveri totesi, että peruskoulun päättöluokkalaisten matematiikan todistusarvosanojen keskiarvoissa oli erittäin merkitseviä eroja ajallisessa tarkastelussa tyttöjen hyväksi, kun taas pojilla oli parempi osaamisen taso (Näveri 2009, 149; 159).

Salmion (2004) väitöstutkimuksessa on vuosina 1993–1995 ja 1998 kerättyä peruskoulun valtakunnallisten arviointihakkeiden aineistoa tarkasteltu kestävän kehityksen didaktiikan näkökulmasta. Tutkimuksessa arvioitiin niitä tekijöitä, jotka olivat vaikuttaneet siihen, että valtakunnallisia seuranta-arviointeja alettiin laatia ja panna toimeen sekä sitä, miten opetussuunnitelman perusteissa opetukselle asetettujen tavoitteiden toteutumista on voitu seuranta-arvioinneilla mitata. Tutkimuksen mukaan opetussuunnitelman perusteiden ja arvioinnin välillä pitäisi olla vielä vahvempi yhteys ja seuranta-arviointitulosten tulisi ohjata enemmän opetussuunnitelman perusteiden kehittämistä. Seuranta-arviointien sisällöt tulkittiin usein kouluissa tärkeiksi opetettaviksi asioiksi ja siksi kouluissa saatettiin opettaa seuranta-arvioinnissa kysytyjä asioita ja laatia samanlaisia koe- ja tehtävätyyppejä kuin valtakunnallisessa arvioinnissa oli käytetty. (Salmio 2004, 279, 283, 302.) Salmion mukaan seuranta-arvioinnit eivät saisi mitata ainoastaan opetussuunnitelmassa määriteltujen sisältöjen saavuttamista, vaan niiden tulisi kytkeytyä myös muuhun koulun tavoitteelliseen toimintaan. Salmion esittämien tutkimustulosten pohjalta olisi perusteltua, että seuranta-arviointitehtävät voisivat pikemminkin keskittyä korkeampitasoisen oppimisen, kuten tiedonhankinta- ja ongelmaratkaisutaitojen mittaamiseen kuin oppiainekohtaisten sisältöjen osaamisen mittaamiseen (Salmio 2004, 283).

Kansainvälisesti esimerkiksi Yhdysvalloissa oppilaan arvosanan antamista ja arvosanojen muodostumista on tutkittu paljon (esimerkiksi Brookhart 1991;1994; 2003; 2007; Ansley, 1997; Cizek 1996; 1997, Phye 1997; Calfee & Masuda 1997; McMillan, Myran, Workman 2002; Shepard 2006; Randall & Engelhard 2009; Gordon & Fay 2010; Guskey 2011). Oppimistuloksia ja niissä hyödynnettäviä mittareita, kuten standardoituja testejä ja niiden validiutta ovat tutkineet esimerkiksi Messick (1980; 1988; 1989a; 1989b; 1994; 1995), Kane (1990; 2001; 2006), Moss (2003; 2007) ja Llosa (2008). Luokkatyöskentelyn arviointiin liittyviä validiuskysymyksiä ovat tutkineet muun muassa Cizek, Fitzgerald ja Rachor (1995).

Brookhartin (1994) laatimassa meta-analyysissä kartoitettiin lähes 20 tutkimusta, jotka liittyivät arvosanan antamiseen ja opettajien arvosanan antamiskäytänteisiin. Muun muassa Cizek (1997; 2007) ja Phye (1997) ovat tutkineet ja teoretisoineet oppimisen, oppimistulosten ja arvioinnin välistä suhdetta sekä luokkatyöskentelyyn

liittyvää formatiivista arviointia. Calfee ja Masuda (1997) ovat tutkineet opettajaa arvioinnin asiantuntijana, jonka lisäksi Calfee (esim. 1987, 1995) on tutkinut arviointia ammatillisen toiminnan näkökulmasta. Opettajien arvosanojen antamiseen liittyviä käytänteitä ja arvosanojen muodostumisperusteita ovat kartoittaneet muun muassa Gullickson (1985) Wilson (1990), Cizek, Fitzgerald, Shawn ja Rachor (1995/1996) sekä McMillan, Myran ja Workman (2002). Ansley (1997) ja Abrams (2007) ovat tutkineet oppimistulosten arviointijärjestelmiä, Ansley standardoitujen koulusaavutuskokeiden (*standardized achievement tests*) ja Abrams formatiiviseen luokkatyöskentelyn arviointiin kehitettyjen testien näkökulmista.

Klapp Lekholm (2008) tutki vuonna 1987 syntyneitten ja peruskoulunsa vuonna 2003 päättäneitten lähes 100 000 ruotsalaisoppilaan arvosana-aineistolla sitä, miten oppilaiden ja koulujen erilaiset ominaisuudet vaikuttavat sekä yksilö- että koulutasolla oppilaiden arvosanoihin. Arvosanojen lisäksi Klapp Lekholmilla oli käytössään 9 000 oppilaan kansallisten kokeiden ja oppilaskyselyiden tuloksia. Tutkimuksessaan Klapp Lekholm osoitti, että arvosanoilla kuvattiin oppilaiden tietoja ja taitoja, mutta niillä arvotettiin myös oppilaiden motivaatiota ja heidän kiinnostuneisuuttaan kyseistä oppiainetta kohtaan, vaikka Ruotsin opetussuunnitelmien perusteissa oli ohjeistettu, että arvosanoilla arvioidaan vain oppiaineissa saavutettuja tietoja. Myös sukupuolella ja vanhempien koulutustaustalla ja esimerkiksi koulun koolla oli yhteys oppilaiden saamiin arvosanoihin, mitä voidaan pitää ongelmallisena oppilaiden päättöarvosanojen yhdenvertaisuuden kannalta. Klapp Lekholmin mukaan tyttöjen ja poikien arvosanaerot esimerkiksi ruotsin ja englannin kielissä selittyivät pitkälti sillä, että tyttöjen motivaatio ja kiinnostus kyseisten kielten opiskeluun oli suurempaa. Koulutasolla Klapp Lekholm osoitti, että arvosanoilla ja huoltajien koulutustaustalla oli vahva yhteys: kouluissa, joissa oppilaiden huoltajien koulutustausta oli matala, oppilaat hyötyivät arvosanan yleisestä ulottuvuudesta arvosanaa annettaessa. Klapp Lekholmin mukaan kyse oli siitä, että opettaja ottaa arvosanaa antaessaan huomioon kognitiivisten tietojen, taitojen ja kykyjen lisäksi ei-kognitiivisia, arvosanaan vaikuttavia tekijöitä. (Klapp Lekholm 2008, 92.)

Wickström (2006) on artikkelissaan luonut yleiskatsauksen siitä, miten ja millä perusteella Ruotsissa annetaan arvosanoja. Hän kuvaa vuoteen 2011 voimassa ollutta opetussuunnitelmaa ja siinä määriteltyihin kriteereihin perustuvaa oppilaan arviointia sekä kansallisten kokeiden ja arvioinnin suhdetta ruotsalaisessa koulutusjärjestelmässä. (Wickström 2006, 120–124.) Wickströmin mukaan, vaikka arvosanat annetaan suhteessa opetussuunnitelmassa asetettuihin oppiainekohtaisiin, kurssikohtaisiin ja vuosiluokittain määriteltyihin tavoitteisiin, arviointi on monimutkainen prosessi, joka vaihtelee opettajittain, oppiaineittain ja vuosiluokittain. Arvioinnin ja siinä arvosanan antamisen monimutkaisuutta lisäsi se, että oppilailla oli mahdollisuus täydentää suorittamiaan kursseja tai korottaa arvosanoja kuntien hallinnoimien, ulkopuolisten kokeiden avulla. (Wickström 2006, 120.) Wickström kuvaa artikkelissaan kansallisia kokeita diagnostisiksi, sillä niiden tarkoituksen on ollut kartoittaa oppilaiden vahvuuksia ja kehittämistarpeita. Kansallisilla kokeilla

on myös pyritty mittaamaan asetettujen tavoitteiden saavuttamista. Wickströmin mukaan useat opettajat ovat käyttäneet kansallisten kokeiden tuloksia apuna antaessaan oppilaille arvosanoja. Kansalliset kokeet ovat hänen mielestään ongelmallisia, sillä niitä on laadittu vain tiettyihin oppiaineisiin ja vain tietyille ikäryhmille. Lisäksi eri vuosien tulokset eivät ole keskenään vertailukelpoisia ja opettajat näyttivät tulkitsevan kansallisten kokeiden tuloksia hyvin eri tavoin (Wickström 2006, 121.)

Ruotsissa tällä hetkellä toimivista tohtorikouluista yksi, Karlstadin yliopiston hallinnoima tutkijakoulu (*Forskarskola för lärare i historia och samhällskunskap*), keskittyy erityisesti yhteiskuntatiedon (*samhällskunskap*) opetukseen, ja siellä on tehty yhteiskuntatietoon liittyviä opinnäytetöitä. Esimerkiksi Odenstadin lisensoitittu (2010) käsittelee erityyppisten kokeiden vaikutusta siihen, miten lukio-opiskelija oppii ja minkälaisia asioita lukiolainen opettelee yhteiskuntaopissa. Odenstadin tutkimuksen mukaan Ruotsin valtakunnallisella opetussuunnitelmalla näytti olevan vain vähän vaikutusta siihen, minkä tyyppinen ja minkä sisältöinen yhteiskuntaopin koe oli (Odenstadin 2010, 171–172).

Opetushallituksen julkaisemassa (2012) yhteiskunnallisten aineiden seuranta-arviointiraportissa on perusteellisesti kuvattu sekä yhteiskuntaopissa että historiassa saavutetut oppimistulokset (Ouakrim-Soivio & Kuusela 2012, 49–74) ja siksi niitä ei enää tässä väitöstutkimuksessa esitellä. Yhteiskunnallisten aineiden seuranta-arvioinnissa ilmeni, että kun koulut luokiteltiin neljään osaamiskvartiiliin (Q1–Q4) niiden yhteiskuntaopissa ja historiassa osoittaman seuranta-arviointimenestyksen mukaan, havaittiin, että kouluissa, joissa yhteiskuntaopin tai historian seuranta-arvioinneissa menestytettiin keskimäärin hyvin (Q4), annettiin yhteiskuntaopissa ja historiassa samoilla ratkaisuosuuksilla alempia arvosanoja kuin niissä kouluissa, joissa keskimääräiset ratkaisuosuudet yhteiskuntaopissa ja historiassa jäivät matalammiksi (Q1). (Ouakrim-Soivio & Kuusela 2012, 110–111.) Opetushallituksen seuranta-arvioinnissa julkaistut tulokset virittivät kiinnostusta ja lisäkysymyksiä esimerkiksi siitä, miten on mahdollista, että vaikka voimassa olevat Perusopetuksen opetussuunnitelman perusteet (2004) ja siinä määritellyt oppiainekohtaiset päättöarvioinnin kriteerit on laadittu ohjaamaan päättöarviointia valtakunnallisesti yhteismitallisemmaksi ja vertailukelpoisemmaksi, esitetyt tulokset yhteiskuntaopissa ja historiassa eivät tue käsitystä nykyisten päättöarviointikriteereiden toimivuudesta.

Opetushallituksen julkaisemat yhteiskunnallisten aineiden oppimistulokset käsitteivät yleisopetuksen oppilaiden lisäksi myös erityistä tukea saaneiden oppilaiden ja muuta kuin suomea tai ruotsia äidinkielenään puhuvien oppilaiden oppimistulokset. Tämän väitöstutkimuksen aineistoon kuuluvat ainoastaan yleisopetuksen oppilaat, kuten luvussa 5 on tarkemmin kuvattu, ja siksi tämän tutkimuksen analyysit on tehty kokonaan uudelleen käyttäen vain keväällä 2011 yleisopetuksessa olleiden ja otantaan kuuluneiden suomen- ja ruotsinkielisten perusopetuksen päättövaiheen oppilaiden aineistoja.

1.4 Tutkimuskysymykset

Tämä väitöstutkimus keskittyy analysoimaan yhteiskunnallisten aineiden seuranta-arviointia varten keväällä 2011 kerätyn aineiston pohjalta oppilaan arvioinnissa ilmenneitä koulujen välisiä eroja suhteessa seuranta-arvioinnissa osoitettuun osaamiseen.

Seuraavassa on koottu tiivistetysti ne kysymykset, joihin tämä tutkimus pyrkii vastaamaan:

1. Miten oppimistulosten seuranta-arviointi ja arvosanat vastaavat toisiaan?
2. Esiintyykö kahden arvioinnin välillä mahdollisia systemaattisia eroja ja jos esiintyy, miten ne ilmenevät?
3. Voidaanko seuranta-arvioinnin perusteella tehdä johtopäätöksiä arvosanojen validiudesta, vai jääkö kahden mittauksen validiusongelma ratkaisematta?
4. Voidaanko kahden oppiaineen samanaikaisella arvioinnilla ratkaista kahden mittauksen eli arvosanojen ja seuranta-arvioinnin validiutta koskeva ongelma?

Tutkimuksen lopuksi pohditaan, minkälaisia johtopäätöksiä voi aineiston perusteella tehdä oppilaiden arvosanoista, arvosanojen antamista ohjaavista normeista ja arvosanojen antamisen käytänteistä.

Tämän tutkimuksen taustalla on yhteiskunnallisten aineiden seuranta-arvioinnissa esille tullut tieto siitä, että sekä kummassakin oppiaineessa koulujen välillä on eroja seuranta-arvioinnissa osoitetun keskimääräisen osaamisen ja oppilaiden saamien keskimääräisten arvosanojen suhteen (Ouakrim-Soivio & Kuusela 2012, 110–111). Koulujen erilaisiin arviointikäytänteisiin on kiinnitetty huomiota muissakin Opetushallituksen oppimistulosten seuranta-arvioinneissa esimerkiksi matematiikassa ja äidinkiessä (ks. esim. Kuusela 2006; Lappalainen 2004; 2006; 2010; Mattila 2002; 2005), ja siksi tässä tutkimuksessa pyritään tarkastelemaan oppilaiden yhteiskuntaopin ja historian arvosanojen validiutta suhteessa seuranta-arvioinnissa osoitettuun osaamiseen. Lisäksi tiedetään, että yhteiskuntaopissa ja historiassa koulujen välinen vaihtelu oli poikkeuksellisen suurta (18,8 %) verrattuna muihin oppiaineisiin, joiden seuranta-arviointitulokset on raportoitu joko vuonna 2011 tai 2012 (Jakku-Sihvonen & Kuusela 2012, 11).

Ensimmäinen ja toinen tutkimuskysymys seuranta-arvioinnissa osoitetun osaamisen ja oppilaiden saamien arvosanojen suhteen pyrkii vastaamaan tutkimusaineiston perusteella yksityiskohtaisemmin siihen, vastaavatko oppilaiden yhteiskuntaopissa ja historiassa saamat arvosanat ja näiden oppiaineiden seuranta-arvioinnissa osoitettu osaaminen toisiaan oppilastasolla ja koulujen sisällä sekä koulujen välillä. Lisäksi tutkitaan, ovatko oppilaiden saamat yhteiskuntaopin ja historian arvosanat ja seuranta-arviointitulokset johdonmukaisessa yhteydessä taustamuuttujiin, vai ilmeneekö arvosanoissa tai seuranta-arviointituloksissa taustamuuttujien suhteen jotain odottamatonta. Yhteen seuranta-arviointiin saattaa sisältyä erilaisia satunnaisia virhelähteitä, mutta tämän tutkimuksen kannalta on olennaista tutkia, syntyykö seuranta-arvioinnin ja arvosanojen välille systemaattisia eroja. Mikäli systemaattisia eroja esiintyy, on kiinnostava tietää, miten ne ilmenevät.

Kolmannessa tutkimuskysymyksessä on viitattu kahden mittauksen ongelmaan, jolla tarkoitetaan sitä, että tutkimuksessa on käytettävissä vain kahden eri mittarin, arvosanojen ja seuranta-arvioinnin antamat tulokset. Ongelmallista näiden kahden mittarin validiustarkastelussa on se, että jos systemaattisia eroja esiintyy, ei yleensä voida päätellä, kumman osaamisen mittarin validius pettää, ellei käytettävissä ole mitään kolmatta kriteeriä tai mittaria, joihin arvosanoja tai seuranta-arviointia voitaisiin verrata. Kuten jo aiemmin tässä luvussa on esitetty, se, että yhteiskunnallisissa aineissa koulujen välinen vaihtelu oli poikkeuksellisen suurta, tarjoaa mahdollisuuden tarkastella kummankin oppiaineen muodostaman tutkimusaineiston pohjalta sitä, voidaanko seuranta-arvioinnin perusteella tehdä johtopäätöksiä arvosanojen validiudesta. Aiemmissa tutkimuksissa on yleensä lähdetty siitä oletuksesta, että juuri arvosannat eivät olisi valideja.

Neljäs tutkimuskysymys liittyy siihen, että yhteiskunnallisten aineiden tutkimusaineisto on sikäli ainutlaatuinen muihin Opetushallituksen tekemiin oppimistulosten seuranta-arviointeihin verrattuna, että yhteiskuntaopin lisäksi kerättiin samaan aikaan ja samojen koulujen oppilailta tutkimusaineisto myös historiassa. Yhteiskuntaopin ja historian seuranta-arviointi tarjoaa ensimmäisen kerran tutkimusasetelman, jossa samojen koulujen ja samojen oppilaiden osaamista voidaan mitata kahdessa eri oppiaineessa ja kahdella eri osaamisen mittarilla, arvosanoilla ja seuranta-arvioinnissa osoitetulla osaamisella. Kummatkin osaamisen mittarit perustuvat samaan kansalliseen normiin eli Perusopetuksen opetussuunnitelman perusteisiin (2004). Vaikka tämän tutkimuksen päämielenkiinto kohdistuu yhteiskuntaoppiin ja oppilaan arviointiin yhteiskuntaopissa, historiasta kerätty identtinen tutkimusaineisto antaa mahdollisuuden tutkia sitä, tarjoaako kahden oppiaineen samanaikainen arviointi mahdollisesti ratkaisun kahden rinnakkaismittarin validiusongelmaan.

2 ARVOSANOJEN JA KANSALLISTEN ARVIOINTIEN TARKASTELUA

Tässä tutkimuksessa on keskeistä opetussuunnitelman perusteissa (2004) määritellyn yhteiskuntaopin osaamisen mittareiden, oppilaiden arvosanojen ja seuranta-arvioinnin, validiuden tutkiminen. Siksi tässä yhteydessä tarkastellaan sitä, miten eri koulutusjärjestelmissä oppilaiden arvosanat määräytyvät ja miten kansallisia oppimistulosten arviointeja, kuten päättökokeita, kansallisia kokeita tai standardoituja testejä hyödynnetään koulutusjärjestelmä- ja oppilastasoilla.¹² Luvussa käsiteltävien neljän esimerkkimaan valossa näyttää siltä, että erilaisissa koulutusjärjestelmissä pyritään yhä etsimään parasta mahdollista menetelmää, jonka avulla oppilaalle olisi annettavissa mahdollisimman validi, ts. vertailukelpoinen ja yhteismitallinen arvosana.

Yleensä arvosanojen antamista ja kansallisia oppimistulosten arviointeja ohjaavat säädökset, normit ja kirjalliset ohjeet, joita tässä luvussa käsitellään arvosanan antamiseen ja kansallisiin oppimistulosten arviointeihin liittyvän tutkimuksen lisäksi. Koska eri koulutusjärjestelmissä arvosanan antamisesta ja kansallisista arvioinneista säädetään ja ohjeistetaan eri lähtökohdista käsin, pyritään maakohtaisin esimerkein konkretisoimaan sitä, millä perusteilla oppilaan osaamista arvioidaan ja arvosanoja annetaan sekä sitä, miten kansallisia arviointeja suoritetaan ja miten arvosanan antaminen ja kansalliset arvioinnit ovat mahdollisesti yhteydessä toisiinsa. Myös siinä, onko opettajan asema arvioitsijana itsenäinen vai epäitsenäinen, kun oppilaalle annetaan arvosanoja tai arvioidaan kansallisia oppimistuloksia, on järjestelmäkohtaisia eroja. Esimerkkimaiden avulla pyritään myös kuvaamaan kahta yleisintä arviointitapaa, suhteellista arviointia ja kriteereihin perustuvaa arviointia, jotka vaikuttavat arvosanojen antamisperiaatteisiin ja siihen, miten oppivelvollisuuskoulun päättövaiheessa mitataan kansallisia oppimistuloksia.¹³

- 12 Eri koulutusjärjestelmissä ulkoista oppimistulosten arviointia voidaan toteuttaa esimerkiksi päättökokeiden tai kansallisten kokeiden avulla, joita selvitetään tarkemmin tässä luvussa maakohtaisin esimerkein. Päättökokeilla tarkoitetaan koko ikäluokan arviointia, joka on edellytyksenä tietyn koulumuodon, esimerkiksi oppivelvollisuuskoulun, loppuunsaattamiseksi. Kansallisia kokeita järjestetään tietyille vuosiluokille tietyissä oppiaineissa, mutta ne eivät ole päättökokeiden tavoin edellytyksenä esimerkiksi päättötodistuksen saamiseksi. Oppimistulosten arviointi on oppijan saavutusten arviointia (Heinonen 2001a, 22), mutta järjestelmätasolla niillä voidaan toteuttaa arvioinnin kontrollitehtävää, jolloin tarkastellaan, onko oppimiselle tai opetukselle asetetut tavoitteet saavutettu. Järjestelmätason arviointi on yleensä tavoiteperusteista tai standardiperusteista. Tavoiteperusteisella arvioinnilla tarkoitetaan, että arvioinnissa mitataan etukäteen asetettujen tavoitteiden saavuttamista. Suomessa vuoden 1985 ja 1994 Peruskoulun opetussuunnitelman perusteisiin oli kirjattu tavoiteperusteinen arviointi, jossa asetetut tavoitteet määriteltiin sen mukaan, mitä ajateltiin, että oppilaiden oli hyvä osata järjestelmätason näkökulmasta (Salmio 2004, 107–108). Standardiperusteista arviointia käsitellään perusteellisemmin luvussa 2.4.
- 13 Suhteellisen ja kriteeriperusteisen arvioinnin lisäksi esimerkiksi Suomessa on annettu arvosanoja absoluuttiseen arviointiin perustuen, jolla tarkoitetaan, että kaikki oppilaat arvioidiin suhteessa yhteiseen muuttumattomaan normiin (Karjalainen 2001, 207) tai ehdottomaan suoritustasoon. Absoluuttisessa arvioinnissa oppilaan suoritusta ei verrata toisten oppilaiden suorituksiin vaan esimerkiksi opetussuunnitelman perusteissa asetettuihin yhteisiin tavoitteisiin. Opetushallitus pitää absoluuttista arviointia keinona antaa valtakunnallisesti yhteismitallisia ja vertailukelpoisia arvosanoja. Opetushallituksen näkemys absoluuttisesta arvioinnista löytyy osoitteesta: http://www.oph.fi/saadokset_ ja_ohjeet/ohjeita_koulutuksen_jarjestamiseen/perusopetuksen_jarjestaminen/perusopetuksen_oppilaan_arviointi/oppilaan_arviointi_menneina_vuosikymmenina

Arvosanat ovat yleisin, summatiivinen tapa arvioida oppilaiden suorituksia (Brookhart 1994, 229). Arvioinnin toteuttamistapa riippuu siitä, kuka tai mikä taho on ensisijainen tiedon tarvitsija (McMillan 2007, 5). Summatiivista arviointia hyödynnetään koulutuksen päättövaiheessa (Jakku-Sihvonen 2010, 317) tai silloin kun mitataan oppilaan saavutuksia (McMillan 2007, 1; Brookhart 2007, 43). Yhtä arvosanan antamiskäytäntöä ei ole mahdollista määritellä, mutta eri koulutusjärjestelmissä on yhdenmukaisia periaatteita, jotka liittyvät arvosanan antamiseen ja arvosanojen merkitykseen osaamisen mittareina. Eri maiden arvosanan antamista ohjaavat normit ja käytännöt arvioida oppilaita eroavat esimerkiksi siinä, ovatko arvioinnin keskiössä tiedot eli opetuksen sisällöt, kuten Yhdysvalloissa, vai arvioidaanko myös taitoja, kuten esimerkiksi Ruotsissa tai Suomessa (Vitikka & Hurmerinta 2011, 64).

Arvosanoilla on merkittävä yhteiskunnallinen tehtävä, sillä niillä pyritään yleensä kuvaamaan oppilaan oppimistuloksia. Useimmissa koulutusjärjestelmissä niillä myös arvotetaan oppilaan mahdollisuuksia siirtyä koulutuksen seuraavaan vaiheeseen. Yhtäältä arvosanat antavat tietoa oppilaan edistymisestä, toisaalta niiden tehtävänä on toimia instrumenttina, joilla osoitetaan, että oppilas on saavuttanut riittävän tieto- ja taitotason, kun hän siirtyy opinnoissaan eteenpäin tai työelämään. Arvosanat luovat yhtenäisen käytännön ja formaalin prosessin, joiden avulla arviointijat, kuten opettajat tai muut, kuten huoltajat tai työnantajat voivat kommunikoida. (Loyd & Loyd 1997, 481.) Arvosanoilla voidaan myös motivoida ja palkita oppilasta, tai arvosanoja voidaan käyttää oppilaan rankaisemiseen, ikään kuin vallan välineenä (Hills, 1991, 540–545). Wickströmin (2006) mukaan arvosanojen tehtävä on antaa tietoa, motivoida ja asettaa valintatilanteessa oppilaita paremmuusjärjestykseen (Wickström 2006, 117).

Se, mitä arvosanojen antamisessa painotetaan, kertoo yleensä myös siitä, mitä yhteiskunnassa arvostetaan ja mitä pidetään tärkeänä. Arvosanan muodostumiseen liittyvien periaatteiden ja arvosanan antamiskäytänteiden tulisi olla selkeitä ja ymmärrettäviä erityisesti oppilaille, mutta myös heidän huoltajilleen sekä muille arvosanatietojä hyödyntäville sidosryhmille. Arvosanan tulisi perustua oikeudenmukaiseen arviointiin, eikä se saisi riippua oppilaan sukupuolesta, sosio-ekonomisesta tai etnisestä taustasta. Arvosanan tulisi kannustaa ja haastaa oppilasta ja samalla antaa riittävästi tietoa oppilaan osaamisesta. (Loyd & Loyd 1997, 482.) Arvosanan antamiseen liitetään validiuden ja reliabiiliuden lisäksi oikeudenmukaisuuden ja yhdenvertaisuuden periaatteet (Gordon & Fay 2010, 93), jotka useissa maissa pyritään turvaamaan säädöksillä, normeilla ja kirjallisilla ohjeilla.

Useissa tutkimuksissa on ilmennyt, että opettajien arvosanojen antamiskäytännöt eroavat paljon koulumuodoittain, oppiaineittain sekä arvioinnissa käytettyjen menetelmien osalta (Gullickson 1985; Stiggins, Frisbie & Griswold 1989; Brookhart 1994; Cizek, Fitzgerald, Shawn & Rachor 1995/1996; McMillan, Myran & Workman 2002). Näiden tutkimusten valossa näyttää siltä, että yhtäältä olisi tarvetta tutkimukselle, jossa tarkasteltaisiin arvosanan antamisen perusteita, kuten

normeja tai muita arvosanan antamista määrittäviä formaaleja ohjeita, ja toisaalta tutkimukselle, jossa keskiössä olisivat arvosanan antaminen käytännön luokkatyössä ja osana opettajan päivittäistä työtä (Brookhart 1994, 229; Cizek, Fitzgerald, Shawn & Rachor 1995/1996, 173–175; Gordon & Fay 2010, 92).

Brookhartin (1994) tekemässä meta-analyysissä hyödynnettiin 19 tutkimusta, jotka oli julkaistu viimeisen kymmenen vuoden aikana ja joissa oli tutkittu opettajien arvosanan antamiskäytänteitä. Meta-analyysissä löytyi arvosanan antamiskäytänteisiin liittyvien erojen lisäksi myös paljon yhteisiä piirteitä. Opettajat pyrkivät antamaan arvosanoja oikeudenmukaisin perustein ja he pyrkivät selvittämään oppilailleen arvosanan muodostumisperiaatteet. Opettajilla oli taipumus esimerkiksi palkita paremmalla arvosanalla niitä oppilaita, jotka yrittivät kovasti (ks. Brookhart 2007, 48). Opettajien arvosanojen antamisen käytänteet ja perusteet, joilla he antoivat arvosanoja, vaihtelivat suuresti (Gullickson 1985; Wilson 1990; Cizek, Fitzgerald, Shawn & Rachor 1995/1996; McMillan, Myran & Workman 2002). Tutkimusten mukaan alakoulujen opettajat antoivat oppilaiden arvosanat useimmiten esimerkiksi observointien tai luokkatyöskentelyn perusteella, yläluokkien opettajat puolestaan antoivat arvosanan useimmiten joko edistymistä mittaavan testin tai opettajan itsensä laatiman kokeen perusteella. Ohjausasiakirjoilla ja arvioinnin ammattilaisten antamilla suosituksilla näytti tutkimusten mukaan olevan vain vähän vaikutusta opettajien arvosanan antamiskäytänteisiin (Brookhart 1994, 229). Joidenkin tutkimusten mukaan opettajat eivät pitäneet arvosanojen antamisesta, vaan se koettiin opettajan työhön kuuluvana ikävänä velvollisuutena (Green & Emerson 2007, 495). Myös opettajien arviointitiedot ja taidot esimerkiksi eri arviointimenetelmistä ja niiden hyödyntämisestä vaihtelivat suuresti (Phye 1997, 35–40).

Vaikka suhteellinen arviointi on yhä yleinen tapa arvioida, sitä käytettiin useissa maissa oppilaan arvioinnin ja arvosanojen antamisperiaatteena erityisesti 1950–1970-luvuilla. Suhteellisessa arvioinnissa arvioinnit suhteutetaan keskenään käsittelyvaiheessa. Koulujen keskinäinen vertailu on esimerkki suhteellisesta arvioinnista. (Koulutuksen tuloksellisuuden arviointimalli 1998, 13.) Tyypillisimpiä suhteellisen arvioinnin sovelluksia on Gaussin käyrä eli normaalijakauma, jota sovelletaan usein joko tahattomasti tai tarkoituksellisesti oppilaita arvioitaessa (Koulutuksen tuloksellisuuden arviointimalli 1998, 14.) Kun arvosanoja annetaan suhteellista arviointia noudattaen, oppilaan tiedot, taidot tai suoritus suhteutetaan muiden saman ryhmän oppilaiden suoritustasoon (Klapp Lekholm 2008, 19). Suhteellisen arvioinnin lähtökohtana on, että henkilöiden tiedot ja taidot vaihtelevat, mutta ne ovat suhteutettavissa saman ryhmän muiden jäsenten tietoihin ja taitoihin. Vertailuryhmänä, johon yksittäisen oppilaan tietoja tai taitoja verrataan, voidaan käyttää esimerkiksi koko ikäluokkaa tai opetusryhmää. Suhteellisen arvioinnin periaatteena on, että oppilaiden arvosanat voidaan asettaa normaalijakaumalle ja että keskiarvojakautuma seuraa arvosanajakautumaa. Normittamiseen ja mittaamiseen perustuvaa suhteellista arviointia on pidetty reliaabelina ja validina, kun on mitattu tietoja ja

taitoja. (Klapp Lekholm 2008, 18.) Gaussin käyrän mukaan arvioitua tulosta on pidetty suuressa oppilasjoukossa yhdenmukaisena ja vertailukelpoisena (Vänttinen 2011, 126). Normaali-jakauman käyttöä arvosanan antamisessa pienellä populaatiolla on kuitenkin kritisoitu, sillä sen käyttö perustuu tuloksen suhteuttamiseen koko väestöryhmässä, ei esimerkiksi yhden oppilasryhmän sisällä (Klapp Lekholm 2008, 19). Suomen lisäksi esimerkiksi Ruotsissa käytettiin aiemmin suhteellista arviointia sekä kokeiden arvioinneissa että arvosanoja annettaessa. Suhteellista arviointia käytetään usein koulutusjärjestelmissä, jotka ovat sentralisoituja ja joissa oppilaan arviointia ohjataan kansallisesti yksityiskohtaisesti. Ohjausasiakirjojen normit voivat olla pakollisia tai ne voivat olla suositusluonteisia, kuten luvuissa 2 ja 3 on suomalaista koulutusjärjestelmää esimerkkinä käyttäen kuvattu.

Vaikka suhteelliseen arviointiin on kohdistunut viime aikoina arvostelua, Euroopan komissio on pyrkinyt harmonisoimaan eurooppalaisten koulutusjärjestelmien käyttämiä arviointi- ja arvosana-asteikkoja luomalla oman ETCS-arviointijärjestelmänsä, joka perustuu suhteelliseen arviointiin. ETCS-järjestelmä pyrkii helpottamaan tutkintojen ja opintokokonaisuuksien suunnittelua, toteutusta, arviointia, tunnustamista ja validointia sekä opiskelijoiden liikkuvuutta.¹⁴ Alun perin ETCS-järjestelmä luotiin korkea-asteen koulutusmuotoja ja elinikäistä oppimista varten, mutta nykyisin esimerkiksi Ranskassa alakouluissa opettajat voivat käyttää oppilaan arvioinnissa kirjainarviointeja (A:sta F:ään), jotka perustuvat ETCS-arviointiasteikkoon ja sen viitekehukseen, jolloin oppilaan arviointi suhteutetaan saman luokan tai opetusryhmän muiden oppilaiden suorituksiin. Järjestelmässä oppilaat jaetaan niihin, jotka ovat suorittaneet arvioinnin hyväksytysti, ja niihin, jotka eivät ole. Oppilaista, joiden suoritus on hyväksytty, arvosanan A saavat ne oppilaat, joiden suoritus vastaa parhaita 10 % luokassa tai opetusryhmässä, arvosanan B:n seuraavat 25 % oppilaista, C:n seuraavat 30 % opetusryhmän tai luokan oppilaista, D:n seuraavat 25 % oppilaista ja E:n viimeiset 10 % oppilaista. Ne, joiden suoritus on hylätty, jaetaan kahteen ryhmään. Arvosanan FX saaneet ovat oppilaita, joilta vaaditaan hieman lisätyötä suorituksen hyväksymiseksi ja arvosanan F saaneet ovat oppilaita, joilta vaadittava työmäärä suorituksen hyväksymiseksi on huomattavasti suurempi. (ECTS 2009, 41–43.)

Kriteereihin perustuvaa oppilaan arviointimallia esittivät ensimmäisten joukossa Glacer ja Klaus (1962). Kriteereihin perustuvalla arvioinnilla uskotaan lisättävän arviointien oikeudenmukaisuutta. Kriteeriperusteista arviointia on pidetty oppilaan kannalta oikeudenmukaisena ja reiluna, koska kriteerien tavoitteet on määritelty etukäteen ja ne perustuvat opetussuunnitelmaan. (Jakku-Sihvonen 2001, 119; Atjonen 2007, 156.) Mitä yksityiskohtaisemmin ja selkeämmin kriteerit on kuvattu, sitä helpompaa on arviointituloksen luotettavuuden toteaminen. (Jakku-Sihvonen 2001, 119.) Kriteeriperusteisessa arvioinnissa toiminnan laatua ja tuloksia verrataan yhdessä sovittuihin kriteereihin. Kriteeriperusteista arviointia hyödynnetään

14 Euroopan komission ECTS-arviointijärjestelmää koskeva sivusto löytyy osoitteesta: http://ec.europa.eu/education/lifelong-learning-policy/ects_en.htm

usein diagnostisiin tarkoituksiin. Kriteereihin perustuvan arvioinnin keskeisimpiä tavoitteita on luoda järjestelmä, jossa arvosanan antamisella mitataan, miten oppilas hallitsee etukäteen määriteltyjä tieto- ja taitoalueita. Arviointikriteerien määrittely edellyttää, että kansallisista tavoitteista on päätetty yhteisesti ja että kriteerit on tulkittavissa yksimielisesti. (Koulutuksen tuloksellisuuden arviointimalli 1998, 13.)

Heinosen mukaan arvioinnin standardilla tarkoitetaan arvioinnin eri osapuolten yhteistä sopimusta, jossa arvioitavalle kohteelle on etukäteen määritelty tavoitearvoja, ominaisuuksia ja periaatteita (Heinonen 2001a, 23). Standardilla voidaan arvioinnin yhteydessä kuvata, miten hyvin tai huonosti kunkin kriteerin suhteen esimerkiksi kokeessa on onnistuttu (Heinonen 2001a, 23) tai standardi voidaan kytkeä asetettuun päämäärään tai kasvatus- ja opetustavoitteeseen (Virta 1999, 97).

Yhdysvalloissa arviointistandardien kehittäminen aloitettiin jo 1970-luvulla. 1980-luvun alussa keskushallinto ehdotti uuden standardi- ja arviointijärjestelmän luomista, jonka pohjalta syntyivät sekä Goals 2000 -asiakirja että No Child Left Behind -lainsäädäntö. Yhdysvalloissa on otettu huomioon standardien laadinnassa sitä koskeva tutkimus, kuten esimerkiksi Cizekin (1996) ja Kanen (1994; 1998) tutkimukset. Standardien määrittämisessä on pitkään vallinnut kaksijakoisuus sen mukaan, asetetaanko standardit testin (*test-centered methods*) vai suorittajan (*examinee-centered methods*) mukaan. Testeihin perustuvat standardit edellyttävät arviointiryhmän tekemää ennakoarviointia testeissä käytettävistä tehtävistä, kun taas suorittajan mukaan asetetuissa standardeissa kyse on siitä, miten testin suorittajat asetetaan paremmuusjärjestykseen. Uusin standardien asettaminen perustuu Hambletonin, Jaegerin, Plaken ja Millsin (2000) kehittämään menetelmään, jossa käytetään sekä testin että sen suorittajan mukaan asetettuja standardeja. Heidän kehittämässään menetelmässä standardien asettamisperusteita ovat testiosiot ja niiden pisteitys, testiin osallistuvan ominaisuudet, testiin osallistuvan työskentely ja arvioitsijan käsitys pisteitysprofiilista. (Hambleton & Pitoniak 2006, 440.)

Virta on teoksessaan viitannut Ravitchin luokitteluun opetuksen standardeista, joita ovat opetussuunnitelmien sisältöstandardit, suoritusstandardit sekä opetusjärjestelyjä ja oppimismahdollisuuksien tarjoamista koskevat standardit (Virta 1999, 97–98). Sisällölliset standardit määrittävät, mitä oppilaiden tulisi oppia ja mitä heille tulisi opettaa. Oppimista koskevat standardit ovat niitä, joita oletetaan kaikkien oppilaiden oppivan. Sisällöllisissä standardeissa painotetaan kriittistä ajattelua ja ongelmanratkaisua ja niiden kytkemistä ainekohtaiseen opiskeluun. Tietoon liittyvien standardien määrittämiseen liittyy kuitenkin ongelmia, kuten se, ketkä määrittävät standardit ja miten tietostandardit liitetään arviointiin. (Virta 1999, 97.) Suoritusstandardeilla määritellään oppilaan osaamista, tekemistä ja taitoja, jolloin kyseessä on tiedon soveltamistaidoista. Suoritusstandardien arvioimiseksi oppilaiden tulee osoittaa taitonsa erilaisissa tehtävissä. (Virta 1999, 98.) Standardien asettaminen perustuu arvioon siitä, mitä oppilaiden tulee esimerkiksi osata tai tehdä. Standardeilla pyritään erottelemaan oppilaiden erilaiset osaamis- ja taitotasot toisistaan

siten, että pisteityksissä määritellään tietyn standardin edellyttämä vähimmäistaso. (Kolen 2006, 168–169.) Oppimisen mahdollisuuksiin (*opportunity to learn*) liittyvissä standardeissa on kyse koulutuksen tasa-arvon toteutumisesta. Oppimisen mahdollisuuksiin liittyvät standardit määrittävät ehdot, jotka koulutuksen järjestäjän tai koulun on saavutettava, jotta oppilaat voisivat puolestaan saavuttaa heille asetetut sisältö- ja suoritusstandardit. (Virta 1999, 98.) Virran mukaan kaikille yhteisten standardien ongelmana on, että oppilaiden kyvyistä, taidoista, tiedoista ja kotitaustasta riippuen standardit voivat olla joillekin liian vaativia, toisille taas liian helppoja (Virta 1999, 102).

American Evaluation Association -yhdistys on määritellyt kahdeksan standardia, jotka kohdistuvat arvioinnin sisältöihin. Atjonen on kuvannut kyseisiä standardeja hyvän ulkoisen arvioinnin mukaisiksi. Nämä arvioinnin standardit kiinnittyvät arvioinnin osuvuuteen eli validiuteen, sattumanvaraisuuden vähentämiseen eli reliaaabeliuteen, testien kehittämiseen ja siihen, että ymmärretään, mihin jokin testi on suunniteltu, testin oikeudenmukaisuuteen, skaalojen, normien ja pisteityksen vertailtavuuteen sekä testien yhteismitalliseen toteutukseen, pisteitykseen ja raportointiin, kielelliseltä taustaltaan erilaisten yksilöiden testaamiseen sekä poliittisten päättäjien vastuuseen siitä, että testejä käytetään moneen tarkoitukseen. (Atjonen 2007, 137–138.)

Samalla tavalla, kun oppiminen ja oppimistulokset poikkeavat jo käsitteinä toisistaan (Cizek 1997, 2), on olennaista erottaa se, mitataanko oppimista ja esimerkiksi siihen kytkeytyviä kognitiivisia taitoja vai oppilaiden saavuttamia oppimistuloksia, jotka perustuvat etukäteen määriteltyihin tavoitteisiin (ks. Abrams 2007, 79; vrt. Koulutuksen tuloksellisuuden arviointimalli 7/1998). Joissain koulutusjärjestelmissä oppivelvollisuuskoulun päätteeksi annettavan todistuksen arvosanat voivat olla yhteydessä kansallisten päättökokeiden tuloksiin, kuten esimerkiksi Ranskassa, tai opettajat voivat itse päättää, missä määrin kansalliset kokeet vaikuttavat oppilaalle annettavaan arvosanaan, kuten Ruotsissa. Käsillä olevan tutkimuksen esimerkimaiden, Ranskan, Yhdysvaltojen, Ruotsin ja Suomen avulla pyritään kuvaamaan myös sitä, kuinka eri maiden koulutusjärjestelmissä arvosanan antamista ohjataan, joko keskitetysti, kuten Ranskassa, tai laajoin tulkinnallisoin vapauksin, kuten Suomessa tai Ruotsissa, joissa opettajien asema arvioijina on itsenäinen. Suomea käytetään esimerkkinä, kun kuvataan suhteellista oppilaan arviointia, Ruotsissa nykyinen oppilaan arviointijärjestelmä perustuu puolestaan ennalta määriteltyihin kriteereihin, ja Ranska on valittu esimerkiksi koulutusjärjestelmästä, jossa käytetään samaan aikaan sekä suhteellista että kriteeriperusteista arviointia. Yhdysvallat on puolestaan otettu tähän tutkimukseen kuvaamaan kompleksista ja monimuotoista

koulutusjärjestelmää, jossa käytetään sekä standardeja, kriteereitä¹⁵ että suhteellista arviointia ja jossa osavaltioilla on toisaalta paljon itsenäistä valtaa, mutta toisaalta kansallisesti on yhdenmukaisesti säädetty esimerkiksi matematiikan, englannin ja luonnontieteiden (*science*) oppimistulosten mittaamisesta.¹⁶

Testien¹⁷ ja kokeiden sisällöillä, sillä mitä ja miten mitataan, viestitetään myös, mitä arvostetaan (Kane 2006, 637). Oppilaan suorituksen arvioinnin tai arvostuksen antamisen lisäksi eri koulujärjestelmät poikkeavat toisistaan myös siinä, miten oppilaiden osaamista mitataan kansallisesti. Joissain maissa kansallisia oppimistulosarviointeja ei järjestetä lainkaan tai ne järjestetään otospohjaisesti kuten Suomessa. Ruotsissa oppimistuloksia mitataan kansallisin kokein vain tietyissä oppiaineissa ja arvioinnit on porrastettu pidettäväksi koko peruskoulun ajan nivelvaiheittain. Ranskassa on puolestaan järjestetty kansallisia päättökokeita vuodesta 1987 lähtien kaikille suomalaista yläkoulua (*collège*) vastaavan koulutason päättövaiheessa oleville oppilaille, ja oppivelvollisuuskoulun jälkeen oppilaille annettava päättötodistus edellyttää päättökokeiden suorittamista.

Yhdysvalloissa Goals 2000 -asiakirjan¹⁸ ja No Child Left Behind (NCLB) -lainsäädännön¹⁹ taustalla on ollut pyrkimys yhdenmukaistaa amerikkalaisten ala- ja yläkoulujen opetusta laatimalla kansallisiin standardeihin perustuvat määritelmät siitä, mitä oppilaiden tulisi oppia ja miten oppimista tulisi mitata (Mathinson, Ross & Vinson 2006, 100). Oppimistulosten innokkaalla mittaamisella on nähty olevan yhteys 2000-luvun Yhdysvalloissa taloudellis-yhteiskunnalliseen ajattelutapaan, jonka mukaan koulutusta ja oppimistuloksia mitataan, jotta voidaan arvioida koulujen tilivelvollisuutta (*accountability*) (Pinar 2004, 204–206; Koretz & Hamilton 2006, 530–578; Mathinson, Ross & Vinson 2006, 99, 102; Abramson 2007, 80).

15 Virran mukaan standardin käsitteellä on monia merkityksiä. Standardia käytetään, kun kuvataan sitä, miten suoritus suhteutuu asetettuun vaatimustason, esimerkiksiin, malliin, normiin tai periaatteeseen (Virta 1999, 97, 142). Atjosen mukaan standardeilla voidaan tarkoittaa jonkin organisaation esittämää suositusta siitä, miten jokin asia tulisi tehdä (Atjonen 2007, 109, 135). Jakku-Sihvonen määrittelee standardit yksityiskohtaisiksi, tarkoiksi ja laajoiksi vaatimustasokuvauksiksi (Jakku-Sihvonen 2001, 119). Standardien ei tarvitse olla ehdottomia. Ne saattavat muuttua ajan myötä ja ne voivat heijastaa esimerkiksi yhteiskunnan odotuksia. Standardeilla voidaan määritellä toiminnalle ideaali tai malli. (Heinonen & Jakku-Sihvonen 2001, 102.) Etenkin Yhdysvalloissa standardien käyttö on yleistynyt laajoina vaatimustasokuvauksina. Kriteerin Virta puolestaan määrittelee vaatimustasoksi, jonka perusteella oppilaan osaamista eli tietoja, taitoja tai tuotoksia arvioidaan (Virta 1999, 96, 141). Heinosen mukaan kriteerit muodostavat perusteen, kun arvioija tekee arvioitavasta kohteesta arvottavia johtopäätöksiä (Heinonen 2001a, 23). Jakku-Sihvonen erottaa standardit ja kriteerit toisistaan siten, että standardien voidaan ajatella määrittelevän vaatimustasoa, kun kriteerit toimivat yksityiskohtaisina vertailuperusteina tai mittavälineinä, kun arvioidaan, onko standardit saavutettu (Jakku-Sihvonen 2001, 119).

16 No Child Left Behind -laki, jossa muun muassa määritellään kyseisten oppiaineiden oppimistulosten säännöllisestä mittaamisesta, löytyy sivustolta: <http://www2.ed.gov/policy/elsec/leg/esea02/107-110.pdf>

17 Cizekin mukaan testillä voidaan viitata henkilön käyttäytymisen pisteittämiseen, jota havainnoidaan objektiivisissa, yhdenmukaisissa ja kontrolloiduissa olosuhteissa. Testillä ei tarkoiteta ainoastaan esimerkiksi kynäpaperi-koetta tai monivalintatehtäviä, vaan testi voi koostua suullisista tehtävistä, esityksestä, oikein/väärin-väittämistä tai esseestä. (Cizek 2007, 107.) Atjonen määrittelee testin siten, että testi toteutetaan ja pisteitetaan yhteismittallisesti ja tulokset raportoidaan samalla tavalla (Atjonen 2007, 138).

18 Kyseinen asiakirja perustuu lainsäädäntöön Goals 2000, Educate America Act of 1994, joka löytyy osoitteesta: <http://www2.ed.gov/legislation/GOALS2000/TheAct/index.html>

19 Perustuu säädökseen No Child Left Behind Act of 2001, joka löytyy osoitteesta: <http://www2.ed.gov/nclb/landing.jhtml>

Yhdysvalloissa No Child Left Behind -lainsäädännön on myös nähty vahvistavan erityisesti englannin kielen, matematiikan ja luonnontieteiden asemaa, sillä näiden aineiden oppimistuloksia mitataan kansallisesti vuosittain peruskoulun vuosiluokilla kolmannelta kahdeksanteen ja kerran lukion aikana. Sen sijaan muiden oppiaineiden oppimistulosten mittaamiseen ei ole yhtenäistä, kansallista järjestelmää, vaan eri osavaltioilla on omat, toisistaan poikkeavat mittaamiskäytänteensä. Joidenkin oppiaineiden oppimistulosten mittaaminen ja säännöllinen seuraaminen on johtanut oppiaineiden väliseen kilpailuun. Julkisuudessa on jopa esitetty, että No Child Left Behind -lainsäädännön vuoksi Yhdysvalloissa panostetaan yhä enemmän matematiikkaan, englantiin ja luonnontieteisiin, kun taas yhteiskuntaopin, historian, maatiedon ja monien muiden aineiden resurssit ja asema ovat heikentyneet, koska niiden oppimistulosten mittaaminen on vapaaehtoista. (Ravitch 2010, 108; Dillon 2006.)

Siinä, miten kansallisia arviointeja hyödynnetään, on myös suuria maa- ja koulutusjärjestelmäkohtaisia eroja. Yhtäältä kansalliset arvioinnit voivat ohjata koulujen omaa pedagogista kehittämistyötä, kuten Suomessa. Toisaalta esimerkiksi Yhdysvalloissa, koulut voidaan asettaa paremmuusjärjestykseen niiden kansallisten oppimistulosten mukaan ja koulujen tilivelvollisuutta voidaan mitata juuri oppimistulosten avulla. Yhdysvaltain arviointitulosten mittaamiskäytänteet ovat moninaiset ja saatujen tulosten käyttötarkoitus riippuu paljon osavaltioiden itsensä tekemistä linjauksista. Esimerkiksi Teksasissa, Virginiassa, Mississippissä ja New Yorkissa lukio- ja yliopistolaisten oppimistulokset vaikuttavat suoraan heidän opintosuorituksiinsa ja valmistumismahdollisuuksiinsa, kun taas Michiganissa ja Kentuckyssa oppimistuloksilla on yhteys ennen kaikkea koulujen tilivelvollisuuteen. New Yorkissa peruskoulun oppimistulokset puolestaan vaikuttavat osavaltion opetushallintoon, mutta niillä ei ole merkitystä yksittäisten koulujen kannalta. (Grant 2010, 43.) Ranskassa kansallinen päättökoe yläkoulun päättyessä ei toimi esimerkiksi sisäänpääsykokeena lukioon tai muuhun toisen asteen oppilaitokseen, vaan sillä osoitetaan, että oppivelvollisuuskoululle asetetut tiedot ja taidot on saavutettu.²⁰

Kansallisilla kokeilla saatetaan pyrkiä kalibroimaan oppilaan arviointia ja annettuja arvosanoja, kuten Ruotsissa on toimittu (Wickström 2006, 117, 121; Klapp Lekholm 2008, 23). Yhdysvalloissa osavaltioiden järjestämät laajat oppimistulosten arvioinnit perustuvat siihen, miten oppilaat ovat saavuttaneet ja oppineet opetussuunnitelmissa asetettuja tavoitteita ja kognitiivisia taitoja (Abrams 2007, 79). Suomalaisessa ja ruotsalaisessa koulutusjärjestelmässä arvosanojen ja kansallisten oppimistulosten arviointien välillä oleva yhteys perustuu siihen, että kummassakin osaamistason määrittely perustuu kansallisissa opetussuunnitelmien perusteissa määriteltyjen keskeisten tavoitteiden, tietojen ja taitojen osaamisen mittaamiseen (Klapp Lekholm 2008, 22–23; Jakku-Sihvonen 2010, 319–320). Suomalaisessa koulutusjärjestelmässä arvosanojen tehtävänä on ennen kaikkea kuvata yksittäisen oppilaan osaamistasoa, kun taas kansallisten seuranta-arviointien tehtävänä on

20 Ranskan opetusministeriön sivustolta, osoitteesta: <http://www.education.gouv.fr/cid2619/le-diplome-national-du-brevet.html>, löytyy voimassa olevat oppilaan arviointia ja päättötodistusta käsittelevät ohjeet ja säännöt.

mitata ja tuottaa tietoa siitä, miten koulutusjärjestelmä tai opetussuunnitelman perusteet toimivat, kuten johdantoluvussa on jo esitetty. Oppimistulosten seuranta-arvioinnissa ei kuitenkaan ole sellaista rakenteellista tekijää, miksi sitä ei voisi käyttää oppilaan arvioinnin tukena määriteltäessä yksittäisen oppilaan arvosanaa.

Arviointitiedoilla, kuten arvosanoilla ja kansallisilla oppimistulosarvioinneilla on merkitystä sekä yksilö- että koulutusjärjestelmätasolla. Yhtäältä arvosanoilla annetaan oppilaille palautetta hänen osaamisestaan ja suoriutumisestaan suhteessa asetettuihin tavoitteisiin tai muihin oppilaisiin, toisaalta arvosanoilla kuvataan muille arviointitietoa tarvitseville, kuten opettajalle, huoltajille tai muille oppilaitoksille oppilaan suoritus- ja taitotasoa. Arvioinnin ja opetuksen välistä yhteyttä on toisinaan kuvattu arkikielen ilmauksilla ”sitä opetetaan, mitä arvioidaan” tai ”sitä saa, mitä mittaa” (Atjonen 2007, 61; vrt. Calfee & Masuda 1997, 70). Arviointimenetelmillä ja arviointivaatimuksilla vaikutetaan myös opetuksen sisältöihin ja tapoihin oppia (Pickford & Brown 2006, 1) ja siksi ei ole yhdentekevää, mihin teoreettiseen viitekehykseen ja järjestelmään arvosanan antaminen ja oppimistulosten mittaaminen perustuvat.

Tässä tutkimuksessa oppilaan arvioinnin teoreettista taustaa ja erilaisia koulutusjärjestelmiä tarkastellaan maakohtaisin esimerkein myös siitä näkökulmasta, kuinka sentralisoidusti eli keskitetysti tai kansallisesti yhdenmukaisin normeihin ja ohjeisiin tai desentralisoidusti eli hajautetusti oppilaan arviointia ja oppimistulosten arviointia hallinnoidaan. Yleensä desentralisoidun koulutusjärjestelmän tunnusmerkkinä voidaan pitää opettajan pedagogista vapautta, joka ilmenee säädöksissä, normeissa ja kirjallisissa ohjeissa. Opettajan asema arvioijana on desentralisoidussa järjestelmässä määritelty ohjausasiakirjoilla itsenäiseksi. Useissa maissa, kuten Ruotsissa tai Suomessa se tarkoittaa, että opettaja itse arvioi oppilaansa suoritukset, myös kansalliset oppimistulosten arvioinnit, ja antaa arvosanat. Myös sentralisoidussa koulutusjärjestelmässä on säädetty oppilaan arvioinnista kansallisesti yhdenmukaisin ohjein sekä siitä kuka tai ketkä muut tahot osallistuvat oppilaan arviointiin esimerkiksi oppivelvollisuuskoulun päättövaiheessa. Sentralisoituun järjestelmään kuuluu usein, että opettajan rooli arvioijana on epäitsenäinen. Esimerkiksi päättökokeen arvioinnin voi suorittaa joko kokonaan tai osittain ulkopuolinen arvioitsija tai arviointiryhmä.

Oppilaan arviointi ja sitä ohjaava järjestelmä ovat aina kontekstisidonnaisia (Klapp Lekholm 2008, 12–13). Siksi arvosanan antamista ja kansallisia arviointeja on tärkeää tarkastella näkökulmasta, joka antaa vastauksen kysymyksiin miksi ja miten arvosanoja annetaan ja kansallisia oppimistulosten arviointeja laaditaan (vrt. Cresswell 2000, 71–72) ja mihin arvosanoja ja oppimistuloksia hyödynnetään. (Jakku-Sihvonen 2010, 317–318.) Seuraavassa tarkastellaan maakohtaisin esimerkein, miten oppilaan osaamista arvioidaan arvosanoilla ja kansallisilla oppimistulosten arvioinneilla ja miten oppilaan arvosanojen määräytymistä ja kansallisia oppimistulosten arviointeja ohjataan erilaisissa koulutusjärjestelmissä.

2.1 Suhteellinen arviointi ranskalaisessa sentralisoidussa järjestelmässä

Ranskalaista koulutusjärjestelmää hallinnoidaan sentralisoidusti. Valtio päättää ja hallinnoi opetussuunnitelmien yksityiskohdista aina kansallisiin päättökokeisiin asti.²¹ Kansallisesti yhdenmukaisin ja tarkoin kirjallisin ohjein sekä ulkopuolista arviointiryhmää (*jury*) tuotosten arvioinnissa hyödyntäen ylläpidetään arvioinnin objektiivisuuden ja oikeudenmukaisuuden periaatteita korostavaa järjestelmää. (Wolf 2000, 24.)

Ranskalaisessa koulutusjärjestelmässä perusopetuksen oppilaat arvioidaan lukuvuoden aikana useaan kertaan. Opettajan lukuvuoden aikana suorittamassa arvioinnissa pääpaino on siinä, mitä sisältöjä oppilas on oppinut. Arviointi voi olla suullista tai kirjallista ja opettaja antaa suorituksesta joko kirjainarvioinnin, kuten esimerkiksi alakoulussa, tai numeroarvioinnin, kuten usein toimitaan yläluokilla. Opettaja voi itse valita, kumpaa arviointitapaa hän noudattaa.²² Koulut voivat halutessaan käyttää eurooppalaiseen viitekehykseen ja suhteelliseen arviointiin perustuvaa ETCS-järjestelmää, jossa oppilas saa suorituksestaan kirjainarvioinnin.

Mikäli oppilas arvioidaan numerolla, arviointi annetaan asteikolla 0–20, jossa 0 on heikoin ja 20 paras annettavista arvosanoista. Yleensä hyväksytyyn suoritukseen vaaditaan vähintään numeroarvosana 10. Toisinaan alakouluissa saattaa olla käytössä arviointiasteikko 0–10, jolloin alin hyväksytty suoritus on 5. Huoltajat saavat tietoa lapsensa menestymisestä joko reissuvihon (*carnet de correspondance*), arviointivihon (*le livret scolaire*) tai neljännesvuosittain jaettavaan todistuksen avulla.

Alkuopetuksen päättyessä (CE1-luokkatason jälkeen eli noin kahdeksanvuotiaille oppilaille) ja alaluokkien päättyessä (CM2-luokkatason eli 11-vuotiaille oppilaille) järjestetään lukuvuoden lopussa kansallinen, diagnostinen arviointi (*l'évaluation nationale*) ranskan kielessä ja matematiikassa, jotta opettajat voisivat arvioida oppilaiden edistymistä ja suunnitella seuraavan lukuvuoden opetusta oppilaiden yksilölliset tarpeet huomioon ottaen. Kyseinen arviointi on jokaiselle oppilaalle pakollinen, ja sen tavoitteena on paitsi arvioida sitä, mitä oppilas on oppinut, myös tukea oppilaan tulevia opintoja, vahvistaa oppilaiden taitoja ja antaa huoltajille tietoa lapsen kehityksestä. Toukokuussa 2012 Ranskassa voimaan astuneen säädöksen mukaan alaluokkien kansalliset, diagnostiset kokeet arvioidaan kriteeriperusteisesti. Kummassakin arvioitavassa oppiaineessa, ranskan kielessä ja matematiikassa, on sisältökriteerien lisäksi asetettu CE1- ja CE2-luokkatasoille erilliset taitokriteerit. Esimerkiksi ranskan kielessä CE1-luokkatasolla olevan oppilaan lukutaitoa arvioidaan kriteerillä, jonka mukaan oppilas kykenee lukemaan ääneen ja ymmärtämään lukemansa sanat. Kirjoitustaidon arviointiin liittyvät puolestaan kriteerit, joiden

21 Ranskan koulutusjärjestelmän nykykuvaus löytyy sivustolta: https://webgate.ec.europa.eu/fpfis/mwikis/eurydice/index.php/France_fr:Aper%C3%A7u_des_principaux_%C3%A9l%C3%A9ments

22 Oppilaan arvioinnin voimassa olevat periaatteet löytyvät Ranskan opetusministeriön sivustolta: <http://www.education.gouv.fr/cid50506/les-parents-a-l-ecole.html#Note>

mukaan oppilas pystyy tuottamaan viidestä kymmeneen riviä tekstiä ja kirjoittamaan sanelusta yksinkertaisia sanoja. Vastaavasti CE2-luokkatasolle järjestettävän kansallisen kokeen sisältö- ja taitokriteerit ovat edellistä luokkatasoa vaativammat. Alaluokkien ranskan ja matematiikan kansallisissa arvioinneissa käytettävä arviointiasteikko on viisiportainen, ja eri tasot on määritelty seuraavin kuvauksin: odotetut vastaukset (*réponse attendue*), menestytty osittain ilman virheitä (*réussite partielle sans erreur*), menestytty osittain, mutta joukossa virheitä (*réussite partielle avec erreur*), muita vastauksia (*autres réponses*), ei vastauksia (*absence de réponse*). Oppilaiden diagnostisen kokeen tulokset käydään läpi yhdessä huoltajien kanssa, eikä arviointituloksia käytetä muuhun kuin oppilaan ja opettajan työskentelyn kehittämiseen.²³

Perusopetuksen yläluokilla oppilaita arvioidaan kuten alaluokillakin kouluvuoden aikana. Oppivelvollisuuskoulun päättyessä kaikki, yleensä noin 15-vuotiaat oppilaat, osallistuvat kansalliseen päättökokeeseen, jonka arvioi erillinen ja ulkopuolinen arviointiryhmä. Kansallisessa päättökokeessa arvioinnin kohteena ovat sekä tiedot että taidot, ja kokeita järjestään sekä suullisina että kirjallisina. Kaikille pakollisia kokeita on ranskan kielessä, matematiikassa ja historian, maantiedon ja yhteiskuntaopin yhdistelmäaineissa. Näissä aineissa kokeet ovat kirjallisia. Tämän lisäksi oppilaat osallistuvat pakolliseen vieraan kielen kokeeseen, joka on suullinen. Edellä mainittujen pakollisten kokeiden lisäksi oppilas valitsee kaksi oppiainetta seuraavista: fysiikka-kemia tai luonnontieteet, terveystieto tai ympäristökasvatus sekä kuvataide tai musiikki.²⁴

Ranskalaisessa järjestelmässä yhdistetään arvosanaa annettaessa sekä päättökokeen tulos että oppilaan kouluvuoden aikana osoittama osaaminen. Päättötodistukseen otetaan huomioon oppilaan osoittamat tiedot ja taidot, joissa oppilaalla on oltava vähintään taitotaso kolme. Oppilaan saavuttama arvosana kolmannen kouluvuoden aikana suoritettussa taidehistorian suullisessa kokeessa sekä kouluvuoden päättyessä suoritettu ranskan kielen koe, matematiikan koe sekä historian, maantiedon ja yhteiskuntaopin yhdistelmäkoee, joka painotetaan kertoimella kaksi. Tämän lisäksi arvioidaan koko kouluvuoden osalta kaikki muut oppilaan oppiaineet paitsi historia, maantieto ja yhteiskuntaoppi. Oppilaalle annetaan koulun käyntiin liittyvistä yleisistä asioista arvosana (*la note de vie scolaire*), joka otetaan huomioon päättötodistuksessa. Joka lukuvuosi arvioitavia koulun käyntiin liittyviä yleisiä asioita ovat myönteinen asennoituminen koulunkäyntiin ja muihin oppilaisiin, poissaolot ja huolellisuus, yhteisten sääntöjen noudattaminen, osallistuminen koulun muuhun toimintaan, kuten esimerkiksi kerhotoimintaan, oppilaskunta- tai tutor-toimintaan tai koulun liikenneturvallisuuskoulutukseen (ASSR) tai ensiapukurssille.

23 Oppilaan arvioinnin periaatteet ja niiden arvioinnin tavoitteet löytyvät Ranskan opetusministeriön sivustolta: <http://www.education.gouv.fr/cid262/evaluation-des-acquis-des-eleves-ce1-cm2.html#Objectifs>

24 Yläluokkien kansallisista kokeista löytyy tietoa Ranskan opetusministeriön sivustolta: <http://www.education.gouv.fr/cid2619/le-diplome-national-du-brevet.html>

Mikäli oppilas on opiskellut latinaa, kreikkaa, vierasta kieltä tai alueella puhuttavaa muuta kieltä kuin ranskaa tai hän on osallistunut työssäoppimisjaksolle, annetaan hänelle lisäpisteitä. Saadakseen kansallisen päättötodistuksen oppilaan tulee saavuttaa vähintään 10 pistettä 20:stä. Saadakseen suorituksistaan todistukseen maininnan oppilaalla tulee olla mainintaan ”aika hyvä” (*assez bien*) keskiarvo 12–14 pistettä enimmäispistemäärästä, joka on 20. Mainintaan ”hyvä” (*bien*) edellytetään pisteitä 14–16 ja mainintaan ”erittäin hyvä” (*très bien*) vaaditaan yli 16 pistettä. Heikosti suoriutunut oppilas (10–11,9 pistettä) ei yleensä saa virallista mainintaa, mutta hänelle voidaan kirjoittaa todistukseen suoritusmerkintä (*correct passable*). Hylätty suoritus on 0–9,9 pistettä.²⁵

2.2 Suhteellinen arviointi suomalaisessa desentralisoidussa järjestelmässä

Suomalaisessa järjestelmässä oppilaan arviointi on ymmärretty jo peruskoulun alkuvaiheista lähtien pitkälti prosessina, joka tukee oppilaa oppimista ja antaa oppimisesta palautetta. Peruskoulun alusta aina 1980-luvun puoliväliin normeihin ja säädöksiin oli kirjattu, että oppilaan arviointi perustui suhteelliseen arviointiin. Suomalaiseen järjestelmään on kuulunut se, että opettajalla on ollut laaja pedagoginen vapaus, kun hän on arvioinut oppilaita ja antanut arvosanoja. Peruskoulussa oppilaan arviointi perustui 1970-luvulta 1980-luvun puoliväliin Peruskoulun opetussuunnitelmakomitean mietinnön I-osaan, jonka mukaan oppilaan arvioinnin perustana tuli olla selkeät tavoitteet, joiden avulla oppilaan tietoja ja taitoja arvosteltiin.²⁶ Mietinnössä korostettiin arvioinnin ja arvosanojen merkitystä oppilaan jatko-opintojen kannalta (Komiteanmietintö 1970: A4, 158). Oppilaan arvioinnin validiuden ja yhdenmukaisuuden uskottiin parantuvan siirryttäessä aiemmin Suomessa käytössä olleesta absoluuttisesta arvioinnista suhteelliseen arviointiin (Komiteanmietintö 1970: A4, 162). Arvosanojen tehtävänä oli kuvata oppilaalle, hänen huoltajilleen ja peruskoulun suorittamisen jälkeen myös muille oppilaitoksille ja työnantajille oppilaan suoriutumista (Komiteanmietintö 1970: A4, 161).

25 Arvosanojen pisteityksistä ja arvosanakuvauksista löytyvät tiedot Ranskan kansallisia kokeita käsittelevältä sivustolta: <http://eduscol.education.fr/cid46835/brevet-organisation-et-modalites-d-attribution.html>

26 Peruskoulun alkuvuosikymmeninä ei puhuttu oppilaan arvioinnista vaan oppilaan arvostelusta.

Suomessa standardoitujen koulusaavutuskokeiden²⁷ katsottiin soveltuvan ennen kaikkea oppilasryhmien, ei niinkään yksittäisten oppilaiden suoritusten arviointiin (Komiteamietintö 1970: A4, 162). Komiteamietinnön mukaan standardoitujen koulukokeiden tuloksista saattoi tehdä päätelmiä oppilaan tai luokan suoritustasosta suhteutettuna koko maan vastaavaan suoritustasoon, joka saatiin otosoppilasryhmän koulusaavutuskokeiden tuloksista. Standardoitujen koulusaavutuskokeiden ja niistä saatujen tulosten ajateltiin auttavan myös opettajaa hänen oman työnsä kehittämisessä ja siinä, kuinka hyvin oppilaat olivat saavuttaneet oppimisen kannalta keskeisiksi määritellyt tavoitteet. Komitea suositteli, että opettaja hyödynsi itse laatimiaan koulukokeita standardoitujen koulusaavutuskokeiden rinnalla. Opettajien laatimien kokeiden oletettiin antavan opettajille enemmän didaktista vapautta, sillä hän saattoi laatimissaan kokeissa painottaa niitä sisältöjä tai opetuksen aihealueita, joita hän oli omassa opetuksessaankin painottanut. (Komiteamietintö 1970: A4, 165.) Opettajan laatimien kokeiden sisällön validiteetin haluttiin kiinnittää huomiota, ja opettajaa kehoitettiin laatimaan monta ja monipuolista tehtäväosiota sisältävä koe. Opettajan laatimien koetehtävien tuli kohdentua opetussuunnitelmassa asetettuihin tavoitteisiin. (Komiteamietintö 1970: A4, 163.) Kokeen reliabiliutta pyrittiin parantamaan kiinnittämällä huomiota hyvään kieliasuun, kokeen yksityiskohtaiseen ohjeistukseen, koetilanteeseen sekä riittävään määrään tehtäviä. Kokeen erotteluvuutta pyrittiin lisäämään laatimalla eritasoisia tehtäviä, myös helppoja tehtäviä. (Komiteamietintö 1970: A4, 164.)

Suhteellisen arviointijärjestelmän erilainen pedagoginen ote eri koulutusjärjestelmissä näkyy esimerkiksi siinä, miten vuoden 1970 Komiteamietinnössä opettajaa kehoitettiin olemaan korostamatta oppilaan tekemiä virheitä, koska sitä pidettiin oppimispsykologian periaatteiden vastaisena (Komiteamietintö 1970: A4, 166). Ranskassa taas kansallisten arviointien viisiportaisesta asteikosta kaksi kuvaa nimenomaan väärin vastausten määrää. Virheet ja niiden lukumäärä toimivat oppilaan arvioinnin yhtenä perusteena, sillä opettaja arvioi oppilaan tuotosta sen mukaan, onko suoritus osittain onnistunut, ei virheitä (*réussite partielle sans erreur*) vai osittain onnistunut, vaikka tuotoksessa on virheitä (*réussite partielle avec erreur*).²⁸ Suhteellisen arvioinnin pedagoginen lähestymistapa ilmeni Suomessa myös siinä, miten opettajia kehoitettiin konkretisoimaan oppilaan suorituksia esimerkiksi

27 Peruskoulun opetussuunnitelmakomitean mietinnössä (1970: A4) todettiin, että ”koulukokeet jaetaan joko opettajan itsensä laatimiin kokeisiin tai standardoituihin koulukokeisiin, jotka mittavat oppikurssien keskeisten kohtien hallintaa”. Standardoidut koulukokeet laadittiin tavoiteanalyysin, esikokeiden ja osioanalyysien jälkeen ja ne oli testattu tietyn luokkatason oppilasjoukolla, joka oli valittu otannan perusteella. Oppilaiden standardoiduista koulusaavutuskokeista saamien tulosten perusteella laadittiin normitaulukot, joihin opettaja saattoi suhteuttaa oman oppilasryhmänsä suoritukset. (Komiteamietintö 1970:A4, 165.) Standardoituja koulukokeita pidettiin merkittävänä ja hyödyllisinä arvioinnin yhdenvertaisuuden kannalta (Komiteamietintö 1970: A4, 165–166). Koulukokeita kehitettiin 1970-luvulla lähinnä lukemista ja ala-asteen matematiikkaa sekä joiltain osin vieraita kieliä varten. Kokeita laadittiin 1970-luvulla Kasvatustieteiden tutkimuslaitoksen toimesta, joka paitsi teki niitä, myös tutki koetuloksia. 1970-luvulta lähtien standardoituja koulukokeita laadittiin myös ainepedagogisten järjestöjen sekä Kouluhallituksen ja myöhemmin 1990-luvulla Opetushallituksen virkamiesten yhteistyönä. (Opetushallitus 1997, 8.)

28 Ranskassa nykyisin käytössä olevat arvosanakuvaukset löytyvät opetusministeriön sivustolta: <http://www.education.gouv.fr/cid262/evaluation-des-acquis-des-eleves-ce1-cm2.html#Objectifs>

kehottamalla oppilaita piirtämään graafinen esitys omista koearvosanoistaan. Näin oppilas pystyi seuraamaan koearvosanojen vaihtelua lukuvuoden aikana. Toisena vaihtoehtona ehdotettiin, että opettajat siirtyisivät antamaan oppilaille suorituspisteitä, jotka oli laskettu oppilaan koesuoritusten tai tehtyjen tehtävien perusteella. Parhaimpana, mutta vaikeimpana vaihtoehtona komitea piti ratkaisua, jossa arvioinnin avulla oppilaalle osoitettiin käytännössä, miten hän on opinnoissaan edistynyt. (Komiteanmietintö 1970: A4, 166–167.) Peruskoulun ensimmäistä opetussuunnitelmaa linjannut komitea kiinnitti huomiota numeroarvioinnin ongelmallisuuteen, ja se totesi mietinnössään ristiriitaisesti, että arviointitapoja tuli kehittää siten, että oppilaat saisivat havainnollisemmalla tavalla tietoa omasta edistymisestään eivätkä vain suoriutumisestaan verrattuna muihin oppilastovereihin (Komiteanmietintö 1970: A4, 162).

Komiteanmietinnössä oppilaan arvosanojen antamisperiaatteita ohjeistettiin perusteellisesti. Opettajia suositeltiin käyttämään kokeiden arvioinnissa samaa arviointiasteikkoa kuin todistusarvosanoja annettaessa. Komitea totesi, että peruskoulujärjestelmän siirtyessä prognostisista kokeista diagnostis-didaktisiin kokeisiin saattoi olla eduksi, että kokeissa käytettäisiin erilaista arviointitapaa kuin todistusarvosanoja annettaessa. (Komiteanmietintö 1970: A4, 167.) Suomalaisesta desentralisoidusta koulutusjärjestelmästä kertoo Komiteanmietinnön ohjeistus, jossa opettajalle jätettiin vapaus valita, kumpaa arvosanakäytäntöä, absoluuttista vai suhteellista arviointia, hän halusi käyttää. Jos opettaja valitsi absoluuttisen arviointitavan, hän käytti kaksiluokkaista arviointia hyväksyty tai hylätty -asteikolla. Tällöin opettaja vertasi oppilaan suoritusta johonkin kriteeriin tai ehdottomaan suoritustasoon, jonka mukaan oppilaalle annettiin hyväksyty tai hylätty suoritusmerkintä. Mikäli opettaja valitsi suhteellisen arviointitavan, oppilaan suoritusta verrattiin muiden oppilaiden tasoon, jolloin oppilas sai vertailevan arvosanan. (Komiteanmietintö 1970: A4, 167.) Suoritusmerkintää pidettiin pedagogisesti tarkoituksenmukaisempana, koska se ohjasi oppilasta kilpailemaan, ei muiden oppilaiden, vaan oppikurssien kanssa. Hyväksytysti kurssin suorittaneita oppilaita pidettiin keskenään samassa asemassa, kun he saivat samanarvoisen arvosanan eli suoritusmerkinnän. Hyväksytyt suoritusmerkinnän arveltiin vähentävän oppilaiden keskinäistä kilpailua arvosanoista. (Komiteanmietintö 1970: A4, 170.) Komiteanmietinnön sisältämän ohjeistuksen ja koulujen arkikäytäntöjen välisiä eroja sekä opettajan itsenäistä asemaa arvioijana kuvaa se, että arvosanojen antamista suhteellisen arviointitavan mukaan jatkettiin eikä hyväksymismerkintä Suomessa syrjäyttänyt esimerkiksi todistuksissa numeroarvosanoja, vaikka valtakunnallisessa ohjausasiakirjassa hyväksymismerkintää suositeltiin.

Vaikka suhteellista arviointia on pidetty eräänlaisena objektiivisen arvioinnin mittana (Vänttinen 2011, 126; Klapp Lekholm 2008, 18), opetussuunnitelmakomiteaa arvelutti arvosanojen antamisen yhdenmukaisuus. Komitean huolta lisäsi silloiset tutkimukset, joiden mukaan opettajat antoivat samassa oppiaineessa samalla luokkatasolla samasta suorituksesta erilaisia arvosanoja. Yhdenvertaisuuden toteutumisen

kannalta pidettiin huolestuttavana, että eri oppiaineissa ja eri luokkatasoilla käytettiin arvosana-asteikkoa eri tavalla. (Komiteanmietintö 1970: A4,168.) Komitea piti arvosanojen antamiskäytänteiden epäyhtenäisyyttä merkinä arvosanojen epäluotettavuudesta (Komiteanmietintö 1970: A4, 169).

Vuoden 1970 Komiteanmietinnössä esitettiin seitsenluokkainen asteikko, jota alettiin soveltaa niissä oppiaineissa, joissa opettaja oli päättänyt suorittaa arvioinnin suhteellisesti. (Komiteanmietintö 1970: A4, 169). Komitea esitti arvosanojen jakautuman seuraavasti:

Arvosana	4	5	6	7	8	9	10
Suhteellinen osuus	3 %	10 %	22 %	30 %	22 %	10 %	3 %

Kun arvosana 4 on hylätty, jakautuma oli normaali keskiarvon 7 suhteen ja arvosanojen hajonta oli 1,33 arvosanaa kohden.

Mikäli opettaja päätti käyttää aiempaa asteikkoa 1–6 jakaumasta tuli seuraavanlainen:

Arvosana	1–2	3	4	5	6
Suhteellinen osuus	5 %	25 %	40 %	25 %	5 %

Komitean jäsenet ymmärsivät suhteellisen arvioinnin sisältämät haasteet ja esimerkiksi sen, että pienessä luokassa vertailtavuus vähenee. Tämän vuoksi opettajia ohjeistettiin niin, että arvosanajakaumaa ei pitänyt noudattaa kaavamaisesti. (Komiteanmietintö 1970: A4, 169.) Opettajan pedagoginen vapaus oppilaan arvioinnissa ja arvosanan antamisessa ilmeni ohjeistuksessa, jossa opettajaa kehoitettiin asettamaan luokkansa oppilaat paremmuusjärjestykseen sen mukaan, miten oppilaat hallitsivat tietoja ja taitoja, miten opettaja oli havainnoinut esimerkiksi oppilaiden luokkatyöskentelyä ja miten oppilaat olivat suoriutuneet opettajan laatimista kokeista (Komiteanmietintö 1970: A4, 169).

Standardoitujen koulukokeiden rooli oppilaan todistusarvosanoja annettaessa ilmeni siten, että sen jälkeen, kun oppilaiden tuloksista oli laskettu luokan keskiarvo, kunkin oppilaan todistusarvosana muodostettiin suhteessa oppilaan sijoitukseen standardoidun koulukokeen perusteella laaditussa paremmuusasteikossa. Komitea kehotti opettajia antamaan todistusarvosanat siten, että niistä laskettava keskiarvo oli sama kuin standardoidun koulukokeen osoittama luokan keskiarvo. Edellä kuvatulla menettelyllä komitea uskoi yhdistettävän opettajan oppilaasta hankkima tieto, jolla käsitettiin oppilaiden välisiä suoritusasoeroja, sekä standardoitujen koulukokeiden antama tieto, joka koski luokan yleistaso. (Komiteanmietintö 1970: A4, 169.) Kyseisellä menettelyllä uskottiin lisättävän luokkien ja

eri opetusryhmien välisten arvosanojen vertailtavuutta (Komiteanmietintö 1970: A4, 170). Komiteanmietinnön ristiriitaisia ohjeita oppilaan arvosanan antamisessa kuvaa muun muassa se, että huolimatta edellä esitetystä, yksityiskohtaisesta ohjeistuksesta oppilaiden laittamiseksi paremmuusjärjestykseen, opetussuunnitelmakehitea esitti toiveen vertailevien arvosanojen poistumisesta toteamalla, että ”komitea pitää kuitenkin suotavana, että vertaileva arvosanajärjestelmä tulevaisuudessa puretaan, ja siirrytään suoritusmerkintöjen käyttöön”. (Komiteanmietintö 1970: A4, 168.)

Opetussuunnitelmassa esitetty toive suhteellisen arviointikäytännön poistumisesta toteutui vasta 15 vuoden kuluttua, jolloin Peruskoulun opetussuunnitelman perusteissa (1985) todettiin, että ”Kaikkien aineiden arvostelussa luovutaan oppilaiden keskinäiseen vertailuun perustuvasta suhteellisesta arvostelusta. Oppilaan arvosanan muodostuminen ei siis ole riippuvainen toisten oppilaiden arvosanoista. Arvostelun on kuitenkin aina oltava oikeudenmukaista.” (POPS 1985, 29.) Suhteellisen arvioinnin päättymistä peruskoulussa perusteltiin sillä, että peruskoulu oli oppivelvollisuuskoulu, joka jokaisen lapsen tuli voida suorittaa. Lisäperusteluina muutokselle todettiin, että oppilasarvioinnin perustana oli opetukselle asetetut tavoitteet, joiden avulla pyrittiin selvittämään, kuinka pitkälle kukin oppilas omien edellytystensä rajoissa saattoi kussakin yksittäisessä asiassa edetä. Arvosanat ilmaisivat oppilaan onnistumista suhteessa asetettuihin tavoitteisiin. Tämän tavoitepohjaisen arviointimenettelyn lähtökohdana arvosanan antamisessa oli, että oppilaan saavutuksia ei verrattu enää muiden saavutuksiin, vaan jokaista oppilasta haluttiin kannustaa niin, että hän käyttäisi omia oppimisedellytyksiään mahdollisimman hyvin. (POPS 1985, 29.) Arvosanojen antaminen perustui jatkuvaan näyttöön ja summatiivisiin kokeisiin eikä oppilaan arvosana saanut laskea ilman erityistä syytä kahta arvosanaa kerralla (Mattila 2010, 12–13).

2.3 Kriteereihin perustuva desentralisoitu järjestelmä Ruotsissa

Ruotsissa siirryttiin suhteellisesta arvioinnista kriteeriperusteiseen arviointiin vuonna 1994. Oppilaan arvosana perustuu ruotsalaisessa koulutusjärjestelmässä opetussuunnitelman perusteissa kansallisesti määriteltyihin yksityiskohtaisiin kriteereihin ja niihin liittyviin taitotasokuvauksiin. Kunnilla, kouluilla ja opettajilla on laaja pedagoginen vapaus tulkita oppilaan arvioinnista annettuja normeja sekä päättää siitä, miten kansallisesti asetettuja kriteereitä tulkitaan arvioinnissa (Klapp Lekholm 2008, 20; OECD 2011, 49). Koulut saavat päättää, minkälaisen painoarvon he antavat kansallisten kokeiden (*nationella prov*) tuloksille ja miten kansallisten arviointien tulokset suhteutuvat opettajan antamaan arvosanaan (Skolveket 2007a, 10; OECD 2011, 49; Skolverket 2011b, 24–25).

Lukuvuoden 2011–2012 alusta Ruotsissa otettiin käyttöön uudistettu peruskoulun opetussuunnitelma.²⁹ Ruotsin opetussuunnitelman perusteet määrittelevät perusopetuksen tavoitteena olevien eri oppiaineiden tietojen lisäksi myös taitoja. Huomionarvoista uudistetuissa opetussuunnitelman perusteissa on, että perustetekstissä kuvatut oppiaineiden tavoitteet ja sisällöt on kiinnitetty eri taitotasoihin hyvin yksityiskohtaisesti. Arvioinnin kriteerit on porrastettu nivelvaiheittain vuosiluokkien 3, 6 ja 9 loppuun. Oppilaiden eri oppiaineissa saavuttama tieto- ja taitotaso arvioidaan kirjaimin A:sta E:hen. Hylätty suoritus arvioidaan kirjaimella F. Tieto- ja taitotasokuvaukset etenevät porrastetusti siten, että peruskoulun kolmannella luokalla on määritelty taitotasot matematiikassa, ruotsin kielessä, ruotsi toisena kielenä oppiaineessa, yhteiskunnallisesti orientoivissa aineissa (*sambällsorienterande ämnena*) sekä luonnontiedepainotteisissa aineissa. Kuudennella vuosiluokalla on määritelty edellisten lisäksi taitotasot myös vieraisa kielissä ja yhdeksännellä vuosiluokalla taitotasot on määritelty kaikissa oppiaineissa.³⁰

Ruotsin kouluviraston ohjeen mukaan oppilaan arviointi perustuu hänen osoittamiinsa tietoihin, jotka oppilas on voinut hankkia muulla tavalla tai muussa yhteydessä kuin oppitunneilla tai kurssin aikana. Opetussuunnitelman mukaan opettajan tulee käyttää arviointinsa perustana kaikkea sitä tietoa, joka hänellä on saatavilla oppilaan osaamisesta, suhteuttaa se kansalliseen vaatimustasoon sekä tehdä kattava arvio oppilaan osaamisesta ja antaa arvosana sen perusteella. Opettajan antaman arvosanan tulee myös perustua monipuolisiin arviointitapoihin.³¹ Opettajan arviointityön tukemiseksi Ruotsin kouluvirasto on laatinut eri perusopetuksen vuosiluokille, 1–3, 4–6 ja 7–9, oppiainekohtaista tukimateriaalia, jossa on yksityiskohtaisia ohjeita esimerkkeineen oppilaan arvioimiseksi.³² Perusopetuksen 9. luokan päättyessä eli päättöarvioinnissa kriteerit on opetussuunnitelmassa määritelty kolmelle arvosanalle, E:lle, C:lle ja A:lle. Päättöarvosanojen taitotasokuvauksissa arvosanoille E, C ja A ei ole eritelty taitoja ja sisältöjä toisistaan, vaan niitä on kuvattu lomittain, jolloin ne muodostavat yhtenäisen kokonaisuuden, kuten taulukossa 1 esitetään (Vitikka & Hurmerinta 2011, 64–65). Esimerkiksi perusopetuksen päättyessä yhteiskuntatiedossa (*sambällskunskap*) vaadittu osaaminen tasolla E on vähäisempää kuin tasolla C ja oppilaalta edellytetään lähinnä perustietojen ja taitojen hallintaa. Taitotasolla C oppilaalta odotetaan hyvää yhteiskunnallisten tietojen ja taitojen hallintaa ja tasolla A oppilaalta edellytetään erittäin hyvää ja monipuolista perehtyneisyyttä yhteiskunnallisiin ja taloudellisiin aiheisiin Taulukossa 1 on kuvattu tiivistetysti muutaman kriteerin avulla yhteiskunnallisten aineiden päätövaiheen taitotasokuvaukset arvosanoille E, C ja A. Näiden lisäksi taitotasosta D

29 Ruotsin kouluviraston (*Skolverket*) asetuksella 2011:19 määrättiin perusopetuksessa annettavien oppiaineiden sisällöistä ja taitotasokuvauksista.

30 Nykyisin voimassa olevan, lukvuonna 2011–2012 käyttöön otetun perusopetuksen opetussuunnitelman perusteden mukainen oppilaan arviointi on kuvattu Ruotsin kouluviraston sivustolla: <http://www.skolverket.se/bedomning>

31 Arvioinnin periaatteet löytyvät Ruotsin kouluviraston sivustolta: <http://www.skolverket.se/bedomning/betyg/att-satta-betyg/2.7109/betyg-i-gymnasieskolan-1.182171>

32 Opettajille suunnattu peruskoulun oppilaan arviointia ja arvosanan antamista käsittelevä tukimateriaali löytyy Ruotsin kouluviraston sivustolta: <http://www.skolverket.se/bedomning/betyg>

todetaan, että oppilas jää joiltain osin taitotasolle E, mutta pääasiallisesti hänen taitonsa sijoittuvat tasolle C. Taitotason B kuvauksessa todetaan, että oppilas jää joiltain osin taitotasolle C, vaikka pääasiallisesti hänen taitonsa sijoittuvat tasolle A. (Skolverket 2011a, 211.)

TAULUKKO 1. Taitotasojen E:n, C:n ja A:n kuvaus yhteiskunnallisissa aineissa perusopetuksen päättövaiheessa muutaman esimerkiksi pöimitun kriteerin suhteen. Lähde: Skolverket 2011a, 211.

Taitotaso E	Taitotaso C	Taitotaso A
<p>Oppilaalla on perustiedot yhteiskunnan rakenteista.</p> <p>Hän pystyy osoittamaan tutkimalla, miten yhteiskunnalliset, oikeudelliset, sosiaaliset, taloudelliset, poliittiset ja mediaan liittyvät rakenteet ovat syntyneet ja miten ne toimivat.</p> <p>Oppilas pystyy kuvaamaan yksinkertaisia yhteiskunnassa toimivia suhteita ja kuvatessaan yhteiskunnan toimintaa oppilas hyödyntää peruskäsitteitä.</p> <p>Oppilas osaa päätellä, miten yksilö ja yhteiskunta vaikuttavat toisiinsa ja hän osaa kuvata yhteiskunnassa toimivia yksinkertaisia suhteita.</p>	<p>Oppilas on hyvin perehtynyt yhteiskunnan rakenteisiin.</p> <p>Hän pystyy osoittamaan erittäin hyvin tutkimalla, miten yhteiskunnalliset, oikeudelliset, taloudelliset, poliittiset ja mediaan liittyvät rakenteet toimivat ja osaa kuvata suhteellisen monimutkaisia yhteiskunnan sisäisiä suhteita ja niiden vaikutusta yhteiskunnan eri rakenteisiin.</p> <p>Oppilas hallitsee melko hyvin käsitteitä, ja hän osaa antaa hyviä esimerkkejä ja perusteluja.</p> <p>Oppilas osaa arvioida ja ilmaista erilaisia näkemyksiä joidenkin sosiaalisten kysymysten osalta. Hän pystyy päättämään, perustelemaan ja argumentoimaan suhteellisen hyvin, ja hän kykenee tarkastelemaan asioita eri näkökulmista.</p>	<p>Oppilas on erittäin hyvin perehtynyt yhteiskunnan rakenteisiin.</p> <p>Hän pystyy osoittamaan erittäin hyvin tutkimalla, miten yhteiskunnalliset, oikeudelliset, taloudelliset, poliittiset ja mediaan liittyvät rakenteet toimivat ja osaa kuvata erittäin monimutkaisia yhteiskunnan sisäisiä suhteita ja niiden vaikutusta yhteiskunnan eri rakenteisiin.</p> <p>Oppilas hallitsee erittäin hyvin käsitteitä ja hän osaa antaa hyviä esimerkkejä ja perusteluja.</p> <p>Oppilas osaa käydä monipuolista keskustelua siitä, miten yksilöt ja yhteisö vaikuttavat toisiinsa ja hän osaa kuvata monimutkaisia suhteita ja erilaisia tekijöitä, jotka ovat tärkeitä yksilön oman elämän, mutta myös muiden kannalta.</p>

Ruotsin kouluviraston tehtävänä on kehittää kansallisia kokeita perusopetukseen ja toisen asteen koulutukseen. Kansallisten kokeiden tavoitteena on edistää yhdenvertaista ja tasapuolista arviointia sekä antaa koulujen ja kuntien lisäksi myös kansallisella tasolla mahdollisuus arvioida, missä määrin opetussuunnitelman perusteissa asetetut tieto- ja taitovaatimukset ovat täyttyneet. Kansallisilla kokeilla pyritään auttamaan oppilasta tai opiskelijaa suunnittelemaan konkreettisesti tulevia opintoja ja motivoimaan oppilasta tämän opinnoissa. Niillä pyritään myös tukemaan oppilaiden tavoitteen asettamista joko oppiaine- tai kurssikohtaisesti.³³

33 Kansallisten kokeiden tehtävämäärittely löytyy Ruotsin kouluviraston sivustolta: <http://www.skolverket.se/bedomning/nationella-prov-bedomningsstod>

Eri oppiaineissa järjestettävät kansalliset kokeet on laadittu siten, että ne mittaavat oppilaan osaamista suhteessa opetussuunnitelman perusteissa asetettuihin tavoitteisiin (OECD 2011, 48). Kansallisia kokeita järjestetään voimassa olevan opetussuunnitelman mukaan säännöllisesti seuraavissa oppiaineissa:

- 3. vuosiluokalla matematiikka, ruotsi ja ruotsi toisena kielenä,
- 6. ja 9. vuosiluokilla englanti, matematiikka, ruotsi ja ruotsi toisena kielenä ja kevästä 2013 lähtien biologia, maantieto, fysiikka tai kemia ja historia sekä yhteiskuntaoppi tai uskontotieto.³⁴

Kansallisiin kokeisiin osallistuminen velvoittaa kouluja ja opettajia. Opettajat korjaavat itse oppilaidensa koevastaukset saamiensa ohjeiden ja mallivastausten perusteella. Kouluilla on kuitenkin jonkin verran vapauksia kansallisten kokeiden arvioinnissa esimerkiksi siinä, korjaavatko oppilaiden omat opettajat kokeet vai korjaako joku toinen, ulkopuolinen opettaja ne. Ruotsissa kansallisten kokeiden tehtävänä on arvioida myös yksittäisten oppilaiden suorituksia. Kansallinen koe toimii opettajan arvioinnin tukena, mutta kouluviraston ohjeiden mukaan opettaja kuitenkin itse päättää viime kädessä, kuinka kansalliset kokeet vaikuttavat yksittäisen oppilaan arvosanaan. Ohjeissa painotetaan oppilaan arvioinnin vartaitavuutta, monipuolisuutta ja sitä, että arvosana ei saa perustua ainoastaan kokeessa osoitettuun näyttöön. (Skolverket 2011b, 24–25, 27.) Wickström näkee kansallisten kokeiden tehtävänä auttaa opettajia sisäistämään opetussuunnitelmassa asetettuja kriteereitä ja kalibroimaan arvosana-asteikkoaan siten, että arvosanat olisivat kansallisesti vertailukelpoisempia (Wickström 2006, 121). Suomalaisissa ja ruotsalaisissa oppimistulosten raportoinneissa on eroja esimerkiksi siinä, että Suomessa koulutuksen järjestäjän, yleensä kunnan, lisäksi vain koulu saa tietoonsa itseään koskevat, koulukohtaiset tulokset. Suomessa koulu yhdessä koulutuksen järjestäjän kanssa päättää sen, miten he hyödyntävät saamaansa arviointitietoa (Koulutuksen tuloksellisuuden arviointimalli 1998; Jakku-Sihvonen 2010, 320). Ruotsissa kansalliseen kokeeseen osallistuneiden koulujen ja kuntien tulokset raportoidaan, ja ne ovat nähtävillä erillisessä SIRIS-tietokannassa.³⁵

Ruotsin kouluvirasto on laatinut runsaasti tukemateriaalia kansallisia kokeita varten, jotka ovat koulujen ja opettajien käytettävissä sen verkkosivujen kautta.³⁶ Tukimateriaalin avulla on tarkoitus muun muassa konkretisoida kansallisesti arvioitavien oppiaineiden ydinsisältöjä, kehittää formatiivista ja summatiivista arviointia sekä tukea tasapuolisesti ja oikeudenmukaisesti arviointiprosessia ja arvosanan antamista.

34 Oppiaineet, joissa järjestetään kansallisia kokeita 6. ja 9.vuosiluokilla, ovat muuttuneet keväällä 2013. Uudet säädökset ja ohjeet löytyvät osoitteesta: <http://www.skolverket.se/bedomning/nationella-prov-bedomningsstod/bedomning-i-grundskolan/bedomning-i-arskurs-4-6/amnesproven-i-arskurs-6>

35 Verkkopohjainen tietokanta löytyy Ruotsin kouluviraston verkkosivujen yhteydessä olevasta osoitteesta: <http://siris.skolverket.se/apex/f?p=Siris:1:0>

36 Ruotsin kouluviraston oppilaan arvioinnin tukimateriaalin sivusto löytyy osoitteesta: <http://www.skolverket.se/bedomning/nationella-prov-bedomningsstod>

Tukimateriaali sisältää diagnostista materiaalia, esimerkkejä kansallisista kokeista sekä koetehtäviä. Ruotsin kouluviraston ylläpitämästä koepankista on saatavissa esimerkkitehtäviä niistä oppiaineista, joissa kansallisia kokeita järjestetään. Sellaisista oppiaineista, joissa kansallisia kokeita ei järjestetä, kuten vieraista kielistä, kotitaloudesta ja kuluttajakasvatuksesta, on saatavissa esimerkkitehtäviä.³⁷

OECD on laajentanut toimintaansa osaamisen arviointien alueelle. Se pyrkii tuottamaan tietoa, jota voidaan käyttää osanottajamaiden koulutusjärjestelmien kehittämiseen, tuloksellisuuden vertailemiseen sekä palvelemaan arkielämää (Jakku-Sihvonen 2010, 317). OECD:n asettama ryhmä arvioi vuosina 2010–2011 Ruotsin koulutus- ja arviointijärjestelmän toimivuutta oppilas-, koulu- ja kuntatasolla sekä kansallisella tasolla. OECD:n raportin laatinut ryhmä käsitteli arviointia kaikilla edellä mainituilla tasoilla, mutta tässä tutkimuksessa tuodaan esiin vain niitä keskeisiä kohtia, jotka liittyvät joko oppilaan arviointiin ja siinä erityisesti oppilaan arvosanan antamiseen tai Ruotsin kansallisiin arviointeihin. OECD:n raportti julkaistiin helmikuussa 2011, jolloin osa siinä arvioiduista osa-alueista perustui vielä vanhaan järjestelmään. Tässä yhteydessä käsitellään vain niitä raportin kohtia, jotka ovat ajankohtaisia uuden opetussuunnitelman astuttua Ruotsissa voimaan syksyllä 2011.

OECD piti raportissaan (2011) ansiokkaana, että SIRIS- ja koepankkijärjestelmät avoimuudessaan palvelevat opettajia, kouluja ja opetushallinnon viranomaisia niiden päivittäisessä arviointityössä (OECD 2011, 34, 51, 59, 81, 101). Arviointiryhmä esitti raportissaan sekä oppilasarvioinnin että kansallisten kokeiden osalta ulkoisen arvioinnin kehittämistä ja selkeyttämistä (OECD 2011, 45). OECD:n raportissa esitettiin useissa kohdissa kuitenkin huoli ruotsalaisesta, desentralisoidusta järjestelmästä, jossa opettajat arvioivat itse oppilaitaan sekä luokassa, arvosanoja antaessaan tai arvioidessaan kansallisia kokeita. Arviointiryhmän mukaan rehtoreilla ja opettajilla ei ole riittävää koulutusta suhteessa heidän tekemäänsä vaativaan arviointityöhön. Raportissa todettiin lisäksi, että Ruotsista puuttui yhdenmukainen ohjeistus, jolla kansallisia kokeita hyödynnettäisiin osana oppilaan arviointia. Ohjeiden puuttuminen vaarantaa arviointiryhmän mukaan oppilaiden arvosanojen reliabiliuden. Oppilasarvioinnin reliabiliutta tulisi OECD:n mukaan kehittää lisäämällä standardoituja suljettuja tehtäviä nykyisen luokkatyöskentelyyn perustuvan arvioinnin rinnalle, jolloin oppilaan arvosanojen ja kansallisten kokeiden tulosten vertailtavuus paranisi. (OECD 2011, 42–43, 45.)

OECD:n raportissa esittämiä kantoja tuki Ruotsin kouluviraston vuonna 2009 tekemä tutkimus, joka osoitti, että opettajien arvosanan antamiskäytänteissä sekä siinä, kuinka kansallisten arviointien tulokset otetaan huomioon arvosanoja annettaessa, oli suuria eroja. Käytänteissä oli eroja sekä koulujen sisällä että koulujen välillä. (Skolverket 2009, 32–38.) Myös aiemmissa tutkimuksissa tulivat esille opettajien erilaiset arviointikäytänteet ja arviointikriteerien erilaiset tulkintatavat sekä

37 Eri oppiaineiden tehtäviä löytyy Ruotsin kouluviraston sivuilta, osoitteesta: <http://www.skolverket.se/kursplaner-och-betyg>

niiden vaikutukset oppilaiden saamiin arvosanoihin (Wickström 2006, 124). Ruotsin kouluviraston laatima analyysi (2007) perusopetuksen päättövaiheessa olleiden oppilaiden saamista ruotsin, matematiikan ja englannin kielen arvosanoista ja kansallisten kokeiden tuloksista vuosilta 1998–2006 (Skolverket 2007a, 3, 34) osoitti, että koulujen välillä oli eroja kansallisessa kokeessa osoitetun osaamisen ja oppilaiden saamien keskimääräisten arvosanojen suhteen. Oppilastasolla johdonmukainen selitys olisi voinut olla se, että opettajilla oli arvosanaa antaessaan oppilaidensa osaamisesta laajempaa ja syvempää tietoa kuin mitä kansallisella kokeella voitiin mitata. Ruotsin kouluviraston mukaan kyseinen perustelu ei selittänyt riittävästi koulujen välisiä systemaattisia eroja annettujen arvosanojen ja kansallisten kokeiden suhteen. Toinen koulujen välisiä arvosana- ja osaamistasoeroja selittävä tekijä olisi saattanut olla se, että oppilaat olivat oppineet ja opettajat olivat opettaneet eri sisällöllisin painotuksin kuin mitä kansallisissa kokeissa oli esimerkiksi oppiaineiden sisältöjä mitattu. Aiemmat analyysit ovat kuitenkin osoittaneet eri oppiaineissa, että esimerkiksi oppimateriaalit ohjasivat ja yhdenmukaistivat opettajien opetusta. Lisäksi tavoitteiden ja sisältöjen opettaminen harmonisoituivat myös siksi, että opettajat pyrkivät opettamaan niitä sisältöjä ja tavoitteita, joita he pitivät keskeisinä kansallisissa kokeissa menestymisen kannalta. (Skolverket 2007a, 37.)

Edellisten selitysten sijaan Ruotsin kouluvirasto löysi muita perusteluita sille, miksi koulujen väliset erot olivat suuria koulujen antamien keskimääräisten arvosanojen ja kansallisissa kokeissa osoitetun osaamisen suhteen: esimerkiksi matematiikassa, jossa kansalliset kokeet pidettiin perusopetuksen päättövaiheen oppilaille helmikuussa, kouluilla oli mahdollisuus antaa tehostettua tukea niille oppilaille, jotka suoriutuivat kansallisista kokeista heikosti ennen kuin oppilaat saivat matematiikan päättöarvosanan kevään lopussa. Näin ollen oppilaan osaaminen saattoi kohentua siitä, mitä helmikuussa osoitettu matematiikan kansallisessa kokeessa osoitettu osaaminen indikoi ja siksi oppilaat saivat lukuvuoden lopuksi myös parempia arvosanoja kuin kansallisen kokeen perusteella saattoi olettaa. (Skolverket 2007a, 37.) Toinen selitys liittyi siihen, että opettajat eivät mielellään antaneet hylättyjä arvosanoja oppilailleen. Opettajat antoivat hyväksytyjä arvosanoja hylättyjen sijaan, vaikka oppilaiden osaaminen ei vastannutkaan opetussuunnitelmassa asetettuja tavoitteita. Kouluviraston laatiman aiemman tutkimuksen mukaan näytti myös siltä, että opettajat tulkitsivat lievemmin opetussuunnitelman perusteissa kirjattuja tavoitteita kuin kansallisissa kokeita laadittaessa oli tulkittu. (Skolverket 2007a, 38.) Oppilaiden yhdenvertaisuusperiaatteen toteutumisen kannalta ongelmallista oli myös se, että kouluviraston laatimien aiempien tutkimusten valossa näytti siltä, että opettajat tulkitsivat arviointikriteereitä kovin eri tavoin antaessaan oppilailleen arvosanoja (Skolverket 2007a, 44).

OECD:n arviointiryhmä esitti, että kansallisten kokeiden reliabiliutta tulisi parantaa siten, että kokeet korjaisi ulkopuolinen arvioitsija tai arvioitsijat, jotka eivät tuntisi oppilaita. Ulkopuoliset arvioitsijat toisivat sekä arvosanojen antamiseen että kansallisiin arviointeihin yhdenmukaisuutta, vertailtavuutta ja yhdenvertaisuutta.

(OECD 2011, 42–43, 45, 48, 49, 54, 103.) OECD:n raportin esittämä kritiikki ja huoli ruotsalaisen desentralisoidun järjestelmän toimivuudesta esimerkiksi oppilaan arvioinnissa ja arvosanojen antamisessa yhdenvertaisin periaattein sekä kansallisten kokeiden reliabeliuden vahvistamiseksi on ymmärrettävää ja perusteltua, mutta samalla voidaan kysyä, eikö desentralisoidun koulutusjärjestelmän vahvuuksiin kuulu juuri opettajien pedagoginen vapaus ja vastuu arviointityön ammattilaisina.

Ruotsissa käytettävien kansallisten kokeiden rakennetta, jossa suljettujen tehtävien lisäksi oli avotehtäviä, arviointiryhmä piti hyvänä, koska opettajilla oli arviointiensa ja arvosanan antamisen perusteeksi monipuolisesti tietoa oppilaan kehittymisestä ja osaamisesta (OECD 2011, 48). Raportissa kannettiin huolta myös siitä, että kansallisten kokeiden reliabelius ja yleistettävyyys olivat heikkoja, koska kokeissa käytettiin standardoitujen suljettujen tehtävien lisäksi esimerkiksi tuottamistehtäviä ja suullisia tehtäviä. Tutkimukset ovat osoittaneet, että avotehtäviä ja suullisia tuotoksia on vaikea arvioida ja että erilaisten suullisten tai muiden tuotosten sekä avotehtävien arvioinnit ovat tehtäväsidonnaisia. (OECD 2011, 50–51.) Esimerkkinä avotehtävien pisteittämisen haastavuudesta oli matematiikan ja ruotsin kielen kansalliset kokeet, joissa tarkistettiin uudelleen 35 000 yhdeksäsluokkalaisten kokeet. Omien opettajien tekemissä matematiikan koetehtävien korjauksissa ei ollut tarkastuskorjauksien jälkeen huomautettavaa, sen sijaan ruotsin kielen avotehtävissä omien opettajien tekemissä korjauksissa ja tarkastajien korjauksissa erot olivat huomattavia. (OECD 2011, 51.) OECD:n arviointiryhmän mukaan amerikkalaiset tutkimukset (Baxter & Glaser 1998; Hamilton ym. 1997; Pellegrino ym. 1999) ovat osoittaneet, että tämäntyyppinen tuottamistehtäviin perustuva arviointitapa ei ennakkouskomuksista huolimatta ole mitannut kovinkaan hyvin oppilaiden taitoja, niissä kehittymistä tai oppimisprosesseja. Sen sijaan ne ovat saattaneet heikentää kansallisten kokeiden validiutta. (OECD 2011, 50.)

OECD:n arviointiryhmä katsoi, että Ruotsissa vuonna 2011 uudistettu opetussuunnitelman perusteet ovat tavoitteiltaan aiempaa selkeämmät ja perusteiden toivotaan tukevan aiempaa opetussuunnitelmaa paremmin koulujen yhdenmukaisia ja yhdenvertaisia arviointi- ja arvosanakäytänteitä. Ryhmä piti hyvänä, että arvosanoja aletaan antaa viimeistään kuudennen vuosiluokan jälkeen, entisen kahdeksannen tai yhdeksännen vuosiluokan sijaan (OECD 2011, 44, 47). Nykyisin voimassa olevan opetussuunnitelman valmistelu perustui osittain Ruotsin kouluviraston seuranta-aineistoihin, joita tehtiin esimerkiksi niissä oppiaineissa, joissa kansallisia arviointeja ei järjestetty. Kuten Suomessa, myös Ruotsissa yhteiskunnalliset aineet eivät kuulu toistaiseksi niiden oppiaineiden joukkoon, joissa järjestetään säännöllisesti kansallisia oppimistulosten arviointeja. Ruotsin kouluvirasto seurasi esimerkiksi vuosien 2001–2006 välisenä aikana sitä, kuinka eri oppiaineissa saavutettiin tieto- ja taitotavoitteet peruskoulun päättövaiheessa. Kouluviraston julkaiseman raportin mukaan oppiainekohtaiset ja valtakunnalliset erot olivat suuret: yhteiskuntaopissa tieto- ja taitotasoa ei saavuttanut vuosien 2001–2006 aikana tehdyissä arvioinneissa keskimäärin 7 % arviointiin osallistuneista oppilaista (Skolverket 2007b, 59).

Viimeisin kansallinen oppimistulosten arviointi suoritettiin yhteiskunnallisissa aineissa Ruotsissa vuonna 2003, ja sitä edeltäneet arvioinnit olivat vuosina 1992 ja 1995 (Oscarsson & Svingby 2005, 122–134; Odenstadt 2010, 18). Vuoden 2003 yhteiskunnallisten aineiden valtakunnallisen arviointiin osallistui Ruotsissa 197 koulua, 10 000 oppilasta ja 1 900 opettajaa (Oscarsson 2005, 5). Arvioinnissa pyrittiin keskittymään niin sanottuihin yleismaailmallisiin kysymyksiin, koska nuoret toimivat yhä enenevässä määrin globaalistuvassa maailmassa (Oscarsson 2005, 9).

Vuonna 2003 tehdyssä arvioinnissa oppilailta kysyttiin, aivan kuten vuosina 1993 ja 1995, minkälainen on hyvä yhteiskunnallisten aineiden osaamista mittaava koe. Yli puolet vastanneista oppilaista oli sitä mieltä, että kokeen ei tulisi mitata pinnallista tietoa, vaan sen tulisi olla sellainen, jossa oppilas voisi osoittaa osaavansa liittämällä asioita erilaisiin asiayhteyksiin sekä muodostamaan syy- ja seuraussuhteita. (Skolverket 2005, 29.) Yhteiskunnallisten aineiden arvioinnissa ilmeni, että kyetäkseen toimimaan monimutkaistuvassa maailmassa peruskoulun 9.-luokkalaisten tulisi kehittää taitojaan enemmän esimerkiksi demokratian osalta (Skolverket 2005, 48–49). Arvioinnissa tarkasteltiin opetussuunnitelmissa asetettuja tavoitteita suhteessa oppilaiden saamiin tuloksiin. Arviointi osoitti, että oppilailla oli riittämättömät tiedot ja taidot esimerkiksi politiikan, taloustiedon ja yhteiskunnallisten muutosten ymmärtämisen osalta (Oscarsson & Svingby 2005, 134; Oscarsson 2005, 63). Oppilaat toivoivat yhteiskunnalliseen opetukseen enemmän sisältöjä, jotka liittyvät heidän elämänalueisiinsa ja vähemmän abstrakteja asiasisältöjä, kuten poliittisia järjestelmiä tai poliittisia prosesseja esitteleviä aiheita (Oscarsson & Svingby 2005, 131; Oscarsson 2005, 63).

Kyseisessä arvioinnissa ilmeni, että oppilaiden tulokset olivat sitä paremmat, mitä enemmän oppilailla oli mahdollisuus vaikuttaa ja valita itseään kiinnostavia sisältöjä. Oppiakseen mahdollisimman hyvin oppilaat tarvitsivat selkeitä ja pysyviä rakenteita, ja heillä tuli olla selvä käsitys opiskelun tavoitteista. Kysyttäessä yhteiskuntaopin opettajilta, kuinka tärkeänä he näkevät opetussuunnitelman roolin ohjausasiakirjana, kaksi kolmasosaa vastanneista oli sitä mieltä, että heidän opetuksensa tulisi perustua vielä enemmän valtakunnallisen opetussuunnitelman perusteisiin sekä paikalliseen opetussuunnitelmaan ja niissä asetettuihin tavoitteisiin. Opettajien vastaus oli kouluviraston tutkimuksen kanssa linjassa sikäli, etteivät kansallinen opetussuunnitelman perusteet enempää kuin paikalliset opetussuunnitelmat tai kurssisuunnitelmat ohjanneet yhteiskuntaopin opetusta riittävästi (Skolverket 2005, 55–56). Pikemminkin vuoden 2003 arviointitulosten valossa näytti siltä, että yhteiskunnallisissa aineissa opetuksellisten tavoitteiden asettaminen oli varsin puutteellista (Oscarsson & Svingby 2005, 130).

2.4 Standardeihin ja suhteelliseen arviointiin perustuva osittain sentralisoitu, osittain desentralisoitu järjestelmä, tapausesimerkkinä Yhdysvallat

Yhdysvallat on esimerkki järjestelmästä, jossa opettajilla on suuri itsenäisyys arvosanojen antamisessa. Arvosanat eivät määräydy ainoastaan osavaltion valitseman standardoidun testin perusteella, vaan opettajien omilla kokeilla on merkitystä arvossanaa annettaessa (ks. esimerkiksi Brookhart 207, 48). Desentralisoidusta järjestelmästä kertoo se, että Yhdysvalloissa ei ole yhtä kansallista opetussuunnitelmaa, vaan jokaisella osavaltiolla on omat opetussuunnitelmansa, jotka usein toimivat sellaisinaan koulujen opetussuunnitelmina (Vitikka & Hurmerinta 2011, 20). Oppiaineille on laadittu monia keskenään kilpailevia standardoituja testejä, joista yleensä osavaltio valitsee sen, mitä käytetään (Brennan 2006, 11; Ferrara & DeMauro 2006, 579). Yhdysvalloissa 1990-luvun lopun ja 2000-luvun koulutuspoliittiset linjaukset ovat johtaneet matematiikan, äidinkielen eli englannin sekä luonnontieteiden opetuksen ja oppimistulosten seuraamiseen yhdenmukaisemmin normein kuin muissa oppiaineissa. Kansalliset arvioinnit ovat keskittyneet lukutaidon ja matemaattisten taitojen (*literacy and numeracy*) mittaamiseen (Vitikka & Hurmerinta 2011, 23) sekä luonnontieteiden (*science*) osaamisen mittaamiseen.

Standardoiduilla testeillä tarkoitetaan yhdenmukaisin pisteitysohjein hallintoituja testejä, jotka on laadittu niin, että testin kysymykset, niiden järjestäminen, arviointi ja pisteitystulosten tulkinta ovat mahdollisimman yhdenmukaisia (vrt. esim. Atjonen 2007, 137–138). Yleensä standardoidut testit arvioidaan suhteellisesti (Ansley 1997, 265), joka on esimerkiksi Yhdysvalloissa yleisin tapa arvioida testien tuloksia (Phye 1997, 40). Testit voidaan rakentaa myös niin, että ne arvioidaan suhteessa asetettuihin kriteereihin. Kriteereihin perustuvat standardoidut testit ovat usein paikallisten kouluviranomaisten kehittämisiä ja niitä voidaan käyttää esimerkiksi vuosittain, kun halutaan tietää, miten tietyille luokkatasolle asetetut tavoitteet on saavutettu. Esimerkiksi Iowassa on kehitetty kriteereihin perustuvia oppimistulostestejä (*criterion-referenced achievement tests*), joilla mitataan erityisesti tiettyjen sisältöalueiden osaamista. Suhteelliseen arviointiin perustuvat standardoidut testit pyritään puolestaan laatimaan joko osavaltiotasolla tai kansallisesti yhtenäisiin normeihin perustuen. (Phye 1997, 41.)

Yhdysvalloissa eri koulumuodot ja oppiaineet ovat erilaisessa asemassa sen mukaan, miten oppimistulosten arvioinneissa käytetään standardoituja testejä. Alakouluissa, joissa tavallisesti noudatetaan oppiainekohtaista opetussuunnitelmaa ja joissa yksi opettaja yleensä vastaa koko luokan opetuksesta, standardoidut testit sopivat Ansleyn mukaan oppilaiden osaamistason testaamiseen paremmin kuin yläkouluissa, joissa opetussuunnitelmien kirjo on suuri ja aineenopettaja vastaa yhden oppiaineen opetuksesta. Alakoulun opettajien suhtautuminen standardoituihin testeihin näyttää olevan yläkoulun opettajia myönteisempää, koska alakoulun opettajat ovat kokonaisvaltaisemmin mukana testausprosessissa ja he voivat hyödyntää

oppilaidensa testituloksia mielekkäämmin esimerkiksi vanhempien kanssa käytävissä keskusteluissa kuin yläkoulun opettajat, jotka eivät välttämättä saa aina oppilaidensa testituloksia edes tiedoksi. (Ansley 1997, 267–268.) Sekä ala- että yläluokilla suhteellisesti arvioitavien standardoitujen testien tehtävänä on yleensä arvioida oppilaan kasvua, vahvuuksia ja kehittämiskohteita suhteessa asetettuihin tavoitteisiin ja indikoida oppilaan menestymistä suhteessa opetusryhmän muihin oppilaisiin (Ansley 1997, 270).

Vaikka Yhdysvalloissa oppilaiden osaamisen arviointi erilaisin standardoiduin testein on yleistä, niillä on todettu olevan varsin harvoin yhteyttä oppilaan avosanaan tai sen antamisperusteisiin (Ansley 1997, 268, 270–271; Phye 1997, 40–41). Monet osavaltiot hyödyntävät esimerkiksi minimiosaamistasoa (*minimum competency*) mittaavia testejä, joiden suorittamisesta on muodostunut edellytys päättötodistuksen saamiselle. Suurimmalle osalla oppilaista arvosanojen perusteella laskettu keskiarvopistemäärä (*Grade Point Average*) on päättötodistusta tärkeämpi. Opettajilla on itsenäinen asema annettaessa oppilaille arvosanoja, mikä on puolestaan johtanut siihen, että arvosanoja ja keskiarvopisteitä joudutaan tulkitsemaan ikään kuin ne olisivat standardeja (Wolf 2000, 16). Sekaannusta aiheuttaa helposti myös se, että monet standardoidut testit, joilla mitataan oppilaan suoritustasoa, saatetaan tulkita eräänlaisina kelpoisuustesteinä (*aptitude test*), mitä ne eivät kuitenkaan ole. Eroa kuvanee parhaiten se, että yleensä koulussa suoritetaan standardoituja testejä vuosittain, mutta kelpoisuustestejä vain yhden tai kaksi kertaa oppilaan kouluran aikana. (Ansley 1997, 281.) Lukiovalmiuksia arvioivan SAT-testin (entiseltä nimeltään *Scholastic Aptitude Test*, nykyisin *Scholastic Assessment Test*) sekä GRE-testin (*Graduate Record Examination*) merkitys on viime vuosikymmenen aikana kasvanut. Näiden testien suorittamisesta on tullut eräänlainen hyväksyttävän suoritustason standardi (Wolf 2000, 17). Testin kuvaaminen standardoiduksi ei Wolfin mukaan kerro muusta kuin, että testit on normitettu suhteessa suureen populaatioon, jolloin tulokset voidaan hyvin suhteuttaa joko kansalliseen pistekeskiarvoon tai muuhun suhteutettuun asteikkoon (Wolf 2000, 17–18).

Yhdysvalloissa viimeisen kahdenkymmenen vuoden aikana tehdyt koulutuspoliittiset linjaukset auttavat ymmärtämään nykyisten kansallisten oppimistulosten mitaamiskäytäntöjen epäyhtenäisyyttä ja siihen kohdistunutta kritiikkiä. Varsinainen keskustelu oppimistulosten arvioinnista ja koulutuksen tuloksellisuudesta sai pontta jo 1960-luvulla, kun Colemanin ja muiden (1966) tutkimus 4 000 valtion ylläpitämän ala- ja yläkoulun ja yhteensä 645 000 oppilaan oppimistuloksista julkaistiin (Coleman ym. 1966, 8). Lähes kaksi vuosikymmentä Colemanin raportin julkaisemisen jälkeen, vuonna 1983, julkaistiin mietintö, *A Nation at Risk: The Imperative for Educational Reform*, jossa pääpaino oli amerikkalaisten lukioiden heikossa tasossa. Perusopetuksen ja sen rakenteiden katsottiin sen sijaan olevan kunnossa. (Ravitch 2010, 28.) Mietinnössä nähtiin, että Yhdysvaltojen taloudellista kilpailukykyä, teollisuutta, tutkimusta ja teknologisia innovaatioita uhkasi amerikkalaisten koulujen ja opetuksen heikko taso (Mathinson, Ross, Vinson 2006, 99; Evans

2004, 152; Ravitch 2010, 22). A Nation at Risk -mietintö keskittyi ennen kaikkea oppimistulosten mittaamiskäytäntöihin ja mittareiden kehittämiseen, ei niinkään opetussuunnitelmatyöhön, opettajien kouluttamiseen tai oppikirjojen kehittämiseen, vaikka niitäkin kyllä mietinnössä käsiteltiin (Ravitch 2010, 24, 27).

Vuonna 1989 presidentti Bush vanhemman käynnistämän koulutuspoliittisen ohjelman, America 2000, tavoitteisiin kuului, että vuoteen 2000 mennessä kaikille lapsille taattaisiin koulunkäynnin aloittamiseen liittyvät oppimisvalmiudet. Tavoitteeksi asetettiin myös, että uuden vuosituhannen alkuun mennessä amerikkalaiset opiskelijat olisivat maailman parhaita matematiikassa ja luonnontieteissä, vähintään 90 % opiskelijoista valmistuisi lukioista ja koko aikuisväestö olisi lukutaitoista. (Ravitch 2010, 31; Mathinson, Ross, Vinson 2006, 100.) 1980- ja 1990-lukujen koulutuspolitiikan kulmakivet A Nation at Risk ja America 2000 käynnistivät keskustelun siitä, miten Yhdysvaltojen taloustilanteen nopean heikkenemisen katsottiin osittain johtuneen sen huonotasoisista kouluista, huonolaatuisesta opetuksesta ja heikoista oppimistuloksista. Monet kasvatusalan asiantuntijat yhdessä suuren yleisön kanssa olivat sitä mieltä, että kouluja kohdannut ”kriisi” tuli ratkaista kehittämällä kansallisia opetussuunnitelman perusteita, palaamalla sellaisten perusaineiden, kuten historian tai maantiedon opettamiseen ja hyödyntämällä standardoituja testejä oppilaiden osaamistason mittaamiseksi (Evans 2004, 163). Osa asiantuntijoista ei kuitenkaan uskonut, että esitetyillä toimenpiteillä olisi vaikutusta, sillä he eivät olleet lähtökohtaisesti vakuuttuneita koulujen, oppimistulosten tai opetuksen huonosta tasosta. Esimerkiksi Kansallisen koulutusyhdistyksen, NEA:n asiantuntija Bracey piti oletuksia amerikkalaisten koulujen huonosta tasosta tai oppilaiden heikoista oppimistuloksista ”suurina valheina”, sillä hänen laatimansa analyysit osoittivat pikemminkin, että amerikkalaiset koulut eivät olleet koskaan aiemmin pärjänneet sekä kansallisesti että kansainvälisesti yhtä hyvin kuin kyseisenä ajankohtana. (Bracey 1991, 105–106.) Oppimistuloksia mittaavien SAT- testien tulokset olivat laskeneet pikemminkin siksi, että testejä tekeviä oppilaita oli aiempaa enemmän tai siksi, että kansainvälisiä vertailuja tehtiin puutteellisesti (Evans 2004, 163) ja testeissä oli paljon teknisiä puutteita (Sanders 1997, 228).

Sekä presidenttien Bush vanhemman että Clintonin kaudella amerikkalaisessa koulutuspolitiikassa kannatettiin vapaaehtoisuuteen perustuvia kansallisia tavoitteita ja mittareita. Vuonna 1994 säädetty Educate America Act -laki perustui Goals 2000 -esitykseen, jossa oppilaiden ja opiskelijoiden uskottiin yltävän sitä parempiin tuloksiin, mitä enemmän heiltä odotettaisiin. Esityksen tavoitteiden mukaan 4., 8. ja 12. vuoden jälkeen oppilaiden ja opiskelijoiden tulisi osoittaa osaamistonsa englannissa, matematiikassa, luonnontieteissä, vieraassa kielessä, yhteiskuntaopissa ja hallinnossa, taloustiedossa, historiassa ja maantiedossa sekä taideopinnoissa. Goals 2000 -ohjelman ja myöhemmin 2000-luvun alussa säädetyn No Child Left Behind -lain pyrkimyksinä on nähty opetuksen muuttaminen mitattavammaksi ja enemmän standardeihin perustuvaksi (Mathinson, Ross, Vinson 2006, 100). Kansallisin koulutuspoliittisin linjauksin opetusta on pyritty myös kehittämään holistisemmaksi ja määrittelemään opetuksen sisältöjä, menetelmiä ja oppimistulosten mittaamista

yhdenmukaisemmaksi (Mathinson, Ross, Vinson 2006, 101). Presidentti Clintonin kaudella Yhdysvalloissa päädyttiin siihen, että osavaltiot määrittivät itse omat arviointistandardinsa, joita esimerkiksi Ravitch on arvostellut sisällöllisesti väljiksi ja tavoitteiltaan epämääräisiksi (Ravitch 2010, 19).

2000-luvun alusta lähtien koulutuspolitiikkaa keskeisesti hallitseva No Child Left Behind -laki on edellyttänyt, että osavaltiot osallistuvat englannissa, matematiikassa ja luonnontieteissä kansalliseen oppimistulosten mittaamisprosessiin (*National Assessments of Educational Progress, NAEP*). Jokainen osavaltio on kuitenkin voinut valita omat oppimistulosten mittaamisen menetelmänsä ja mukauttaa ne määrittelemiinsä suoritusasoihin, kuten perustaitoihin, riittäviin taitoihin tai kehittyneisiin taitoihin. Osavaltiot päättävät itse sen, miten No Child Left Behind -lain sisältämä käsite riittävästä taitotasosta (*proficiency*) määritellään käytännössä. Julkista rahoitusta saavat koulut testaavat vuosittain oppilaansa vuosiluokilla kolmannelta kahdeksanteen ja kerran lukion aikana lukemisessa, matematiikassa ja lukuvuodesta 2005–2006 alkaen myös luonnontieteissä.³⁸ No Child Left Behind -laki velvoittaa osavaltioita ja koko amerikkalaista koulujärjestelmää toimenpiteisiin, joiden avulla 100 % vuosiluokkien 3–8 oppilaista saavuttaa osavaltion asettaman vaaditun suoritusason (*proficient performance*) lukuvuoteen 2013–2014 mennessä (Ferrera & De Mauro 2006, 579). Käytännössä tavoite tarkoittaa, että jokainen osavaltio joutuu laatimaan suunnitelman, jonka pohjalta osoitetaan, miten kaikki oppilaat ovat saavuttaneet riittävän taitotason matematiikassa ja lukemisessa lukuvuonna 2013–2014. Koulut, jotka eivät saavuta asetettuja tavoitteita kaikissa oppilasryhmissä ja vaadituissa oppiaineissa, luokitellaan SINI-kouluiksi (*school in need of improvement*). Tarvittaessa kouluja voidaan rangaista esimerkiksi erilaisin taloudellisin keinoin, koulun opetussuunnitelmaa tai henkilöstöä uudistamalla tai koulupäivien lukumäärää lisäämällä. (Ravitch 2010, 97–98.) Edellä kuvattu käytäntö on jatkunut myös 2010-luvulla presidentti Obaman kaudella.

Yhteiskunnallisten aineiden etujärjestö NCCS (*National Council for the Social Studies*), paheksui omassa julkilausumassaan No Child Left Behind -lainsäädännön aiheuttaneen yhteiskunnallisten aineiden opetuksen osalta kentän jakautumisen kahtia. Järjestön mukaan englannin, matematiikan ja luonnontieteiden opetuksen otetaan useissa osavaltioissa opetusaikaa erityisesti alakoulun yhteiskunnallisten aineiden opetuksesta. Julkilausumassa NCCS toteaa, että valtakunnallisten kokeiden puuttuminen yhteiskunnallisista oppiaineista on johtanut näiden oppiaineiden statuksen laskemiseen. Yhteiskunnallisten aineiden opettajat uskovat, että kansallisesti yhtenäistä opetusohjelmaa on lähes mahdoton aikaansaada ilman, että yhteiskunnallisille aineille laaditaan yhteiset kriteerit taitotasojen saavuttamisen testaamiseksi ja valtakunnallisten kokeiden laatimiseksi.³⁹

38 Oppiaineet löytyvät sähköisessä muodossa olevasta No Child Left Behind -laista osoitteesta: <http://www2.ed.gov/policy/elsec/leg/esea02/107-110.pdf> sivulta 21.

39 NCCS:n julkilausuma No Child Left Behind -lainsäädännöstä löytyy osoitteesta: <http://www.socialstudies.org/positions/nclbera>

Toisenlaisiakin näkökulmia on esitetty. Grantin mukaan No Child Left Behind -laki on vähitellen lisännyt yhteiskunnallisia aineita siihen joukkoon, joiden arvioinnissa käytetään osavaltioiden laatimia standardoituja testejä, siitä huolimatta, että mitään velvoitetta tähän ei ole. Vuonna 2004 noin puolet (23) osavaltioista ilmoitti käyttävänsä standardoituja testejä arvioidessaan yhteiskunnallisia aineita. (Grant 2010, 43.) Yhteiskunnallisten aineiden osalta näyttää siltä, että osavaltioiden laatimilla standardoiduilla testeillä on ollut vaikutusta lähinnä opettavien aihealueiden valintaan, mutta ei siihen, miten opettaja opettaa (Grant 2010, 44). Näyttää myös siltä, että opettajat käyttävät osavaltioiden standardoituja koetehtäviä, kuten monivalintatehtäviä, esseeaiheita tai lyhyitä vastauksia vaativia tehtäviä osana itse laatimiaan kokeita (Grant 2010, 45).

No Child Left Behind -lain yksi arvovaltaisimmista ja kovaäänisimmistä arvostelijoista on viime aikoina, paradoksaalista kyllä, ollut Ravitch, joka oli suunnittelemassa ja toteuttamassa sekä Bush vanhemman että Clintonin kauden kansallista järjestelmää oppimistulosten mittaamiseksi. Ravitch arvostelee poliittisia päättäjiä muun muassa siitä, ettei näillä ole ymmärrystä siitä, ettei standardoiduinkaan oppimistulosten mittaamisjärjestelmä ei ole riittävän tarkka ja erotteleva (Ravitch 2010, 153). Vuonna 1999 julkaistussa arviointiryhmän (*The Committee on Appropriate Test Use of National Research Council*) asiantuntijaraportissa varoiteltiin, että standardoidut testit eivät ole erehtymättömiä ja niiden suorituspisteet eivät välttämättä kerro oppilaiden taidoista ja kyvyistä (Heubert & Hauser 1999, 275–276). Kaksi vuotta raportin julkaisemisen jälkeen Yhdysvalloissa päätettiin kuitenkin hyväksyä laki, jonka seuraukset ovat olleet ankaria sekä kouluille että opettajille (Ravitch 2010, 153). Ravitchin mukaan standardoituja testejä saatetaan käyttää eräänlaisina oikopolkuina, joiden avulla oppimistulosten mittaamista ja niistä saatuja tuloksia käytetään muuhunkin kuin opetuksen laadun indikoimiseen ja opetuksen tason nostamiseen (Ravitch 2010, 154). Ravitch arvostelee osavaltioita myös siitä, että nämä osaavat peukaloida oppimistuloksia esimerkiksi madaltamalla pisteitystasoa edelliseen vuoteen verrattuna, jolloin aiempaa suurempi oppilasmäärä läpäisee kyseisen testin. Osavaltion opetusviranomaiset eivät kerro pisteitystason madaltamisesta, jolloin suuren yleisön silmissä näyttää siltä, että oppimistulokset ovat edelliseen kertaan verrattuna parantuneet. Esimerkiksi matematiikan tulokset ”kohe-nivat” edelliseen mittaamiskertaan verrattuna Buffalossa 28,6 %:sta 63,6 %:iin tai New Yorkissa 57 %:sta 81,8 %:iin. (Ravitch 2010, 157–158.)

Oppimistulosten mittaamisen nykykäytänteissä muun muassa Ravitch ja Koretz kritisoivat sitä, että oppilaat harjoittelevat standardoituja testejä varten oikeita vastaus-tekniikoita ja he käyttävät paljon aikaa väärin vaihtoehtojen poissulkemistekniikoiden opetteluun tai etukäteen räätälöityjen tekstien laatimiseen sen sijaan, että oppilaat opettelisivat eri oppiaineiden tietoja ja taitoja. (Koretz 2008, 253–255; Ravitch 2010, 107, 160.) Julkisuudessa on keskusteltu paljon myös siitä, että opettajat opettavat testejä varten ja kouluopetuksessa keskitytään siihen, mitä testeissä oletetaan arvioitavan (Grant 2010, 44). McNeilin (2000, 3) mukaan opetussuunnitelmien

kytkeminen standardoituihin testeihin on johtanut sekä opetussuunnitelmien kaventumiseen että myös opetussuunnitelmien standardointiin, joiden seurauksena kouluksellinen epätasa-arvo on lisääntynyt vähävaraisten, etniseen vähemmistön kuuluvien ja hyvätuloisten, valtaväestöön kuuluvien oppilaiden välillä (McNeill 2000, 6). Saamastaan kritiikistä huolimatta standardoiduilla testeillä on amerikkalaisessa koulumaailmassa vankka kannattajakunta. Esimerkiksi 62 % opettajista ja 61 % opetustoimen tarkastajista sekä koulujen rehtoreista pitää standardoituja testejä ”välttämättömänä pahana” ja 70 % amerikkalaisista on sitä mieltä, että standardoituja testejä tulee käyttää, jotta voidaan varmistaa, että oppilailla on riittävä akateeminen taitotaso (Ferrara & DeMauro 2006, 580–581).

Yhä useammat oppilaat ja heidän huoltajansa, opettajat ja opetusalan asiantuntijat ovat alkaneet vastustaa standardeihin perustuvaa opetusta ja oppimistulosten mittaamista sillä perusteella, että testit ovat riittämättömiä mittaamaan monipuolisesti oppilaiden osaamista. Tutkijat eivät kuitenkaan jaa yksimielisesti tätä suuren yleisön käsitystä. Ansleyn (1997) mukaan usein kuultu väite siitä, että monivalintatehtävillä ei voisi mitata kuin lyhytkestoiseen muistiin perustuvaa tietoa tai vaatimatonta ajattelutaitoa edellyttävää osaamista, ei tutkimusten perusteella näytä pitävän paikkaansa. Hyvällä kysymyksenasettelulla ja taitavalla monivalintatehtävien laatimisella voidaan mitata vaativia ajattelutaitoja edellyttävää osaamista. (Ansley 1997, 276.) Laaja testaaminen perustui jo alusta alkaen myös vaativampien kognitiivisten prosessien, kuten soveltamistaidon ja ongelmaratkaisutaidon mittaamiseen (Abrams, 2007, 80). Standardoitujen testien etuna pidetään sitä, että niillä voidaan kerätä suuri määrä tietoa edullisesti, nopeasti ja tehokkaasti (Calfée & Masuda 1997, 90).

Ravitchin lisäksi useat tutkijat ovat jo vuosikymmeniä esittäneet, että standardoitujen testien lisäksi oppilaita tulisi arvioida esimerkiksi luokkatyöskentelyn, osallistumisaktiivisuuden, kotitehtävien ja opettajan havaintojen perusteella (Ravitch 2010, 152; vrt. myös Gullickson 1985; Stiggins, Frisbie & Griswold 1989; Brookhart 1994; Cizek, Fitzgerald, Shawn & Rachor 1995; McMillan, Myran & Workman 2002). Yhdysvalloissa on käyty keskustelua siitä, miten erilaisia arviointitapoja tulisi kehittää, jotta voitaisiin lisätä esimerkiksi oppimisen arvioinnin validiutta (Virta 1999, 23). Monet tutkijat, kuten Cizek (ks. myös Glaser ym. 1987; Mitlevy ym. 1991; Nichols 1994; Phye 1997; Quellmalz & Hoskyn 1997) eivät näe standardoituja testejä ja muita arviointimenetelmiä toisiaan poissulkevinä, vaan toisiaan täydentävinä arviointitapoina. Cizekin mukaan arvioinnilla on perinteisesti keskitytty mittaamaan oppilaan tiedonhankintataitoja ja sitä, mitä tietoja ja kuinka paljon oppilas on saavuttanut. Uusia arviointitapoja tarvitaan, jotta voitaisiin esimerkiksi arvioida sitä, missä määrin oppilas on ymmärtänyt sisältöjä ja miten oppilas on muodostanut kognitiivisia rakenteita. (Cizek 1997, 13; vrt. Brookhart 2007, 56.) Abramsin mukaan laajamittaisten standardoitujen testien tavoitteena on mitata, missä määrin oppilaat ovat saavuttaneet opetussuunnitelmassa asetetut, kognitiiviset taidot, kun taas formatiivisen arvioinnin tavoitteena on tehdä oppimisprosessista läpinäkyvää ja osoittaa, missä oppilaalla on esimerkiksi vielä puutteita

(Abrams 2007, 79). Cizek painottaa, että on ensisijaisen tärkeää ymmärtää, että uudet arviointimenetelmät eivät korvaa vanhoja arviointitapoja, vaan niille kummallekin on yhä tarvetta. Vaihtoehtoisen arviointitavan käsitettä (*alternative assessment*) Cizek pitää huonona ja jopa harhaanjohtavana, sillä se luo mielikuvan, että erilaiset arviointitavat, kuten portfolioit, esseet, esitykset tai projektit olisivat vaihtoehtoisia tapoja kerätä samaa tietoa oppilailta. Cizekin mukaan esimerkiksi portfolioarviointi ja standardoidut oppimistuloksia mittaavat testit pyrkivät vastaamaan kahteen olennaisesti erilaiseen arviointikysymykseen: portfolioarvioinnin tavoitteena on arvioida yksittäisen oppilaan kasvua ja kehitystä suhteessa oppilaan itsensä asettamiin tavoitteisiin, kun taas oppimistulosten mittaamiseen laaditut standardoidut testipatteristot mahdollistavat oppilaiden oppimistulosten vertaamisen suhteessa muiden, jopa suuren oppilasjoukon tuloksiin (Cizek 1997, 14–15). Erilaisen arviointitiedon hyödyntäminen ei ole ristiriitaista, sillä esimerkiksi laajamittaisella testauksella tai formatiivisella arvioinnilla saatu erilainen arviointitieto palvelee eri tarkoitusta (Abrams 2007, 79). Erilaiset arviointitavoitteet pikemminkin heijastavat sitä, että suorituksen monipuoliseen arviointiin tarvitaan erilaista ja keskenään vertailukelpoista tietoa (Cizek 1997, 15; vrt. Phye 1997, 34; Brookhart 2007, 56).

Tutkijat ovat arvostelleet standardoituja testejä siksi, että niillä tuotettua tietoa on hyödynnetty johonkin sellaiseen, mihin niitä ei ole alun perin kehitetty ja laadittu tai testien tuloksia on tulkittu virheellisesti. Alun perin standardoitu testijärjestelmä sisälsi neljä osa-aluetta, joita olivat sisällölliset standardit, asetettujen standardien saavuttamisen mittaaminen, suorituskyvyn mittaaminen, jolla seurattiin, onko oppilas saavuttanut toivotun osaamistason, ja testitulosten seuraukset, jotka kytkeytyivät siihen, palkitaanko suoritusasteen saavuttamisesta vai rangaistaanko siitä, että asetettuja tavoitteita ei ole saavutettu. (Abrams 2007, 80–81.) Mathinsonin ja Freemanin (2003) mukaan standardoidut testit näyttävät tätä nykyään palvelevan enemmän ulkopuolisten tahojen, kuten opetusviranomaisten tavoitteiden toteutumista, eivätkä standardoidut testit motivoi oppilaita enempää kuin opettajiakaan. (Mathinson & Freeman 2003, 7, 17–19; Mathinson, Ross & Vinson 2006, 108.) Ravitch arvostelee puolestaan standardoitujen testitulosten hyödyntämistä. Hänen mukaansa standardoidut testit palvelevat vain sitä tarkoitusta, johon ne on aikoinaan laadittu, eikä niitä tulisi käyttää muuhun. Viidennen luokan lukemistaitoja mittaava testi kertoo vain lukemisen taitotasosta, se ei kerro esimerkiksi opettajan kyvystä opettaa lukutaitoa oppilaille. (Ravitch 2010, 153–153; vrt. myös Mathinson, Ross & Vinson 2006, 107–108; Mathinson & Framgoli 2006, 198–199.) Tutkijoiden kritiikki on kohdistunut ennen kaikkea standardoitujen testien validiteetin, sillä heidän mukaansa standardoidut testit ja niillä saadut tulokset ovat valideja, sikäli kun testejä ja niillä saatuja tuloksia on hyödynnetty siihen tarkoitukseen, mihin ne on alun perin suunniteltu eli mittaamaan sitä, mitä niiden on ollut tarkoitus mitata. Jotta testit olisivat valideja, niistä saadut tulokset tulee myös yleistää oikein ja oikeassa asiayhteydessä (Calfee & Masuda 1997, 91; vrt. Messick 1980, 1013; Kane 2001, 328).

Tässä tutkimuksen toisessa luvussa on pyritty esittämään, miten eri maiden koulutusjärjestelmissä oppilaan arvosanan muodostumista normitetaan ja ohjeistetaan eri tavalla siitä huolimatta, että arvosanojen tehtävänä on lähes poikkeuksetta kaikissa järjestelmissä kuvata yksittäisen oppilaan osaamista ja antaa tietoa oppilaan suoriutumisesta ja osaamisen tasosta. Luvussa esitetyt maakohtaiset esimerkit kertovat myös siitä, että erilaisissa koulutusjärjestelmissä pyritään löytämään mahdollisimman validi menetelmä, jolla oppilaan suorituksia voidaan arvioida. Maakohtaisin esimerkein on myös kuvattu sitä, millä eri tavoilla ja miksi oppimistuloksia mitataan ja miten oppimistulosten mittaamisella saavutettuja tuloksia hyödynnetään esimerkiksi oppilaan päättöarvosanaa annettaessa. Kansallisten arviointien tavoitteena on yleensä kuvata osaamisen tasoa suuremmassa oppilasjoukossa. Eri koulutusjärjestelmissä arvosanat määräytyvät eri tavoin. Kyse ei ole ainoastaan siitä, käytetäänkö esimerkiksi suhteellista arviointia vai kriteereihin perustuvaa arviointia, vaan myös siitä, kuinka yhdenmukaisin säädösin tai suurin tulkinnallisin vapauksin ohjausasiakirjoissa opettajien arviointityötä ja arvosanan antamista ohjataan.

Oppilaiden saamien arvosanojen validius eli se, että arvosanat kuvaavat mahdollisimman virheettömästi sitä, mitä niiden oletetaan kuvaavan eli oppilaiden osaamista, sekä arvosanojen vertailtavuus on useissa koulutusjärjestelmissä pyritty takaamaan erilaisin toimenpitein. Esimerkiksi Ranskassa oppimistuloksia mitataan kansallisilla, eri oppiaineiden päättökokeilla, joiden suorittaminen on myös edellytyksenä päättötodistuksen saamiselle. Päättökokeiden suoritusasolla on yhteys oppilaan päättöarvosanoihin ja arvosanojen validius ja vertailtavuus on pyritty turvaamaan ottamalla huomioon oppilaan suoritusaso kansallisessa päättökokeessa päättöarvosanaa annettaessa. Ruotsissa kansallisia kokeita suoritetaan tietyissä oppiaineissa ja yksittäiselle opettajalle tai koululle on jätetty paljon itsenäistä päätösvaltaa sen suhteen, miten kansalliset kokeet vaikuttavat oppilaan arvosanaan. Yhdysvalloissa osavaltiot itse päättävät, miten standardoitujen testien tuloksia hyödynnetään esimerkiksi oppilaan arvioinnissa. Tutkijoiden mukaan näyttää siltä, että testituloksia ei hyödynnetä juurikaan oppivelvollisuuskoulun päättyessä, mutta alaluokkien opettajat hyödyntävät testituloksia esimerkiksi huoltajien kanssa käymissään keskusteluissa (Ansley 1997, 268). Yhdysvalloissa standardoituja testejä pidetään toimivina, kun pyritään mittaamaan nopeasti, tehokkaasti ja edullisesti suuren oppilasjoukon osaamista (Calfee & Masuda 1997, 70), mutta ne eivät sovellu esimerkiksi sellaisten oppilaiden tuotosten arvioimiseen, jossa pyritään mittaamaan oppilaan kykyjä tuottaa omia kertomuksia (Ansley 1997, 276). Julkisuudessa standardoituja testejä on arvosteltu jyrkästi, koska monet osavaltiot käyttävät standardoituja testejä ja niistä saatuja tuloksia, aivan muuhun tarkoitukseen, kuten koulutuksen tilivelvollisuuden mittareina tai taloudellisten resurssien jakoperusteina, kun mihin ne on alun perin laadittu (Ravitch 2010, 153; vrt. myös Mathinson, Ross & Vinson 2006, 107–108; Mathinson & Framgoli 2006, 198–199).

Koska myös suomalaisen koulutusjärjestelmään on 1990-luvun lopulta kuulunut oppimistulosten arviointi, on tärkeää, että suomalaista oppimistulosten arviointijärjestelmää voidaan vertailla muiden maiden vastaaviin järjestelmiin. Näyttää myös siltä, että oppimistulosten arviointi on suomalaisessa koulutusjärjestelmässä vielä osin tuntematonta, sillä Opetushallituksen oppimistulosten arviointia, sen tavoitteita, periaatteita ja toteuttamistapoja ei kaikilta osin vielä tunneta. Opetushallituksen oppimistulosten arviointi ei ole sama asia kuin oppiaineen päättökoe (ks. esim. Atjonen 2007, 133). Suomalaisessa järjestelmässä oppimistulosten arvioinnin tavoitteena on saada tietoa siitä, miten perusopetuksen päättövaiheessa oppilaat osaavat kyseiselle oppiaineelle opetussuunnitelman perusteissa määriteltyjä tietoja ja taitoja. Oppimistulosten arvioinnissa käytettävät tehtävät on laadittu mittaamaan oppilaiden osaamista suhteessa opetussuunnitelman perusteissa määriteltyihin eri oppiaineiden tietoihin, taitoihin ja kriteereihin, aivan kuten eri oppiaineissa päättöarvioinnin kriteereillä pyritään arvioimaan, miten oppilas on saavuttanut asetetut tavoitteet (Päättöarvioinnin kriteerit 1999, 6).

Seuraavassa luvussa perehdytään tarkemmin siihen, mitä tietoja ja taitoja yhteiskuntaopille on opetussuunnitelman perusteissa eri vuosikymmeninä asetettu, miten yhteiskuntaopissa oppilaan yhteiskuntaopin arviointia ja arvosanan antamista on ohjausasiakirjoilla ohjeistettu, miten Perusopetuksen opetussuunnitelman perusteiden (2004) yhteiskuntaopille asetetut kriteerit määrittävät oppilaan osaamista perusopetuksen päättövaiheessa sekä miten kriteerit ohjaavat ja tukevat yhteiskuntaopin päättöarvosanan antamista.

3 OPETUSSUUNNITELMA- JA KRITEERIPERUSTEINEN OPPILAAN ARVIOINTI SUOMESSA

Oppilaan arviointi ei ole yksittäinen tekninen suoritus, vaan se on kiinteästi yhteydessä oppimiskäsitykseen ja opetussuunnitelmiin (Jakku-Sihvonen & Heinonen 2001, 78–80). Perusopetuksen opetussuunnitelman perusteissa (2004) oppilaan arviointia koskevat normit pohjautuvat osittain edellisten vuosikymmenten opetussuunnitelmien perusteisiin. Siksi tämän luvun alussa käsitellään Peruskoulun opetussuunnitelmien perusteiden (POPS 1985; POPS 1994) oppilaan arviointia käsitteleviä osuuksia sekä Perusopetuksen oppilaan arvioinnin perusteiden (1999) ja Perusopetuksen päättöarvioinnin kriteereiden (1999) arvosanan antamiseen liittyviä keskeisimpiä periaatteita. Luvun keskeinen tarkoitus on kuitenkin kuvata sitä, miten suomalaisessa koulutusjärjestelmässä oppilaan arviointia ja siinä erityisesti päättöarvosanan muodostumista säädetään ja ohjeistetaan. Koska oppilaan arvioinnin tulee kiinnittyä opetussuunnitelman perusteissa asetettuihin oppiaineen tavoitteisiin ja sisältöihin, käsitellään tässä luvussa myös sitä, mitä opetussuunnitelmien perusteissa (POPS 1985; POPS 1994; POPS 2004) on määrätty peruskoulun yhteiskuntaopin opetuksesta, sen tavoitteista, sisällöistä ja kriteereistä.

Oppilaan arvioinnin ohjeet ja arviointikäytännöt ovat olleet pitkään kerrostuneita, minkä Opetushallitus toteaa omalla oppilaan arviointia käsittelevällä Internet-sivustollaan ja siellä olevissa artikkeleissa.⁴⁰ Kun peruskoulun opetussuunnitelman perusteet vuonna 1994 uudistettiin, oppilaan arviointia käsittelevää osuutta ei juuri kehitetty (Virta 1999, 94). Vuoden 1985 Peruskoulun opetussuunnitelman perusteiden mukaan arvosanojen antaminen perustui opetukselle asetettuihin valtakunnallisiin ja kuntakohtaisiin tavoitteisiin, joihin oppilaan omaa edistymistä verrattiin (POPS 1985, 29). Mitä tämä käytännössä tarkoitti, sitä ei selkeästi määritely. Peruskoulun opetussuunnitelman perusteiden (1985) astuessa voimaan suhteellinen arviointi päättyi ainakin oppilaan arvosanan antamista ohjaavien säädösten ja normien tasolla, sillä arvosanojen määräytymistä suhteessa muuhun oppilaskäluokkaan pidettiin peruskoulun periaatteiden ja toiminta-ajatuksen vastaisena (Opetushallitus 1996,1).

Arvioinnin kahta erilaista tehtävää, oppimisen tukemista arviointipalautteen avulla ja osaamisen tason määrittelyä arvosanoilla, ei myöskään käsitelty vuoden 1994 perustetekstissä. Lisäksi perusteet määrittelyivät arviointia niin väljästi, että se käytännössä johti oppilaan arvioinnin valtakunnallisen ohjauksen vähenemiseen ja siihen, että arviointi perustui yhä enemmän paikallisiin ratkaisuihin ja päätöksiin. Peruskoulun oppilasarviointityöryhmä (1996) totesi vuoden 1994 Peruskoulun opetussuunnitelman perusteiden (1994) julkaisemisen jälkeisestä tilanteesta, että

40 Opetushallituksen verkkosivustolle on koottu oppilaan arviointiin liittyvä osio, joka löytyy osoitteesta: http://www.oph.fi/saadokset_ja_ohjeet/ohjeita_koulutuksen_jarjestamiseen/perusopetuksen_jarjestaminen/perusopetuksen_oppilaan_arviointi

peruskoulun oppilasarvioinnissa oli käytössä ”absoluuttis-suhteellis-yksilöllinen sekamalli” ja että ”koulussa annettavien numeroiden merkitys on niiden tulkit-sijan omassa harkinnassa”.⁴¹ Perusteiden väljyys ja arvosanojen laaja tulkinnalli-suus johtivat siihen, että oppilaan arvioinnin perusteita jouduttiin uudistamaan jo vuonna 1999. Perusteiden uusimisen osasyynä oli myös uudistunut perusopetus-laki (628/1998) ja -asetus (852/1998), jossa oppilaan arviointi jaettiin opintojen aikaiseen arviointiin ja päättöarviointiin (Perusopetusasetus 852/1998 § 10, § 12). Kyseinen jaottelu siirtyi sittemmin vuoden 1999 oppilaan arvioinnin perusteisiin (Perusopetuksen päättöarvioinnin kriteerit 1999, 5) ja myöhemmin Perusopetuk-sen opetussuunnitelman perusteisiin (2004). Uudessa laissa ja asetuksessa oppilaan arvioinnin tehtäväksi määriteltiin opiskelun ohjaaminen ja kannustaminen sekä oppilaan itsearviointitaitojen kehittäminen. Säädösten mukaan oppilaan oppimista, työskentelyä ja käyttäytymistä tuli arvioida monipuolisesti.

Opetushallitus julkaisi keväällä 1999 sekä Perusopetuksen oppilaan arvioinnin perusteet että Perusopetuksen päättöarvioinnin kriteerit. Päättöarvioinnin kriteerit olivat oppiainekohtaisia, ja niissä määriteltiin hyvän osaamisen kuvaus arvosanalle 8. Opetushallituksen tavoitteena oli parantaa päättöarviointikriteerien avulla arvo-sanojen valtakunnallista vertailtavuutta ja sitä kautta oppilaiden oikeusturvaa, kun he hakeutuvat jatko-opintoihin (Opetushallitus 1996, 1–2; Perusopetuksen päät-töarviointin kriteerit 1999, 5–6). Päättöarviointin kriteerien tuli palvella myös sitä, että oppilaan oppimistulokset vastaavat ja kuvaavat oppiaineen kulloistakin arvosanaa (Perusopetuksen päättöarviointin kriteerit 1999, 6). Päättöarviointin kriteereitä valmistelleen työryhmän ohjeistuksessa painotettiin, että ”kriteerit tulee johtaa opetussuunnitelman perusteista ja oppiaineiden tavoitteista, niiden tulee olla valideja ja reliaabeleja, ne auttavat mittaamaan mitä on tarkoitettu. Kriteerei-den tulee olla käyttökelpoisia koko maassa, niiden tulee olla ymmärrettäviä ja myös oppilaiden tulee ymmärtää kriteereitä ja hyödyntää niitä opiskelussaan.” (Opetus-hallitus 1996 13–14; Opetushallitus 1997.)

Jotta eri aineiden päättöarviointin kriteerit olisivat jollain tavalla yhteismitallisia, Opetushallitus ohjeisti päättöarviointin kriteereitä valmistellutta työryhmää niin, että eri oppiaineissa päättöarviointin kriteereiden ”mittatikku” rakennettiin suh-teuttamalla oppilasikaluokan osaamistaso niin, että puolet oppilaista olisi yltänyt päättöarvosanoihin 4–7 ja puolet arvosanoihin 8–10 (Opetushallitus 1996; Ope-tushallitus 1997; ks. Rantala 1999, 261). Opetushallitus painotti työryhmän muis-tiossa, että kriteerien avulla kehitettävässä arviointimallissa peruskoulun päättöarvosanat perustuivat tavoiteperusteiseen arviointiin eikä luokan tai koulun sisäiseen suhteelliseen arviointiin (Opetushallitus 1996, 8). Tavoiteperusteinen oppilaan arviointi määriteltiin siten, että ”oppilaan tiedot ja taidot suhteutetaan asetettui-hin tavoitteisiin. Niitä ei verrata muiden oppilaiden saavutuksiin. Peruskoulun

41 Lainaus löytyy Opetushallituksen oppilaan arviointia käsittelevästä artikkelista ”Arviointikäytännöt eri vuosikymmeninä”-otsikon alta osoitteesta: http://www.oph.fi/saadokset_ja_ohjeet/ohjeita_koulutuksen_jarjestamiseen/perusopetuksen_jarjestaminen/perusopetuksen_oppilaan_arviointi/oppilaan_arviointi_menneina_vuosikymmenina

päätyessä arviointi suhteutettaisiin peruskoulun kokonaistavoitteisiin. Peruskoulun päättövaiheen lähestyessä oppilas ottaisi tavoitteekseen tiettyjen arvosanojen saavuttamisen ja näin kouluaikainen arviointi ja tavoiteopiskelu linkittyisivät päästötodistukseen”. (Opetushallitus 1996, 3.) Tavoiteperusteisen arvioinnin määritelmä jäi pois Peruskoulun päättöarvioinnin kriteereistä (1999), jossa oli enää maininta siitä, että ”Päättöarvioinnin on annettava monipuolinen ja luotettava kuva oppilaan osaamisesta suhteutettuna perusopetuksen tavoitteisiin. Arvioinnin on myös annettava tietoa oppilaan menestymisestä suhteessa oppiaineiden tavoitteisiin.” (Perusopetuksen päättöarvioinnin kriteerit 1999, 5.)

Koska päättöarvioinnin tehtävänä oli kuvata monipuolisesti ja luotettavasti oppilaan osaamista, päättöarvioinnin kriteerit suhteutettiin kunkin oppiaineen tavoitteisiin ja keskeisiin sisältöihin koko perusopetuksen oppimäärän osalta (Perusopetuksen päättöarvioinnin kriteerit 1999, 5). Käytännössä Perusopetuksen päättöarvioinnin kriteerit johdettiin Peruskoulun opetussuunnitelman perusteissa (1994) olevista oppiaineiden tavoitteista ja keskeisistä sisällöistä, jotka suhteutettiin hyvään osaamiseen eli arvosanaan 8. Oppilas sai arvosanan 10, mikäli hän ylitti lähes kaikkien oppiaineen kriteereiden edellyttämän osaamistason ja osoitti erityistä harrastuneisuutta oppiainetta kohtaan. Mikäli oppilas ylitti useimpien oppiaineen kriteereiden edellyttämän osaamisen tason, hän sai arvosanan 9. Arvosana 8 edellytti keskimääräistä oppiaineen kriteereiden edellyttämää osaamista. Joidenkin kriteerien saavuttamatta jättämisen saattoi kompensoida muiden kriteereiden ylittäminen. Arvosanassa 7 oppilaan oletettiin osoittavan osaamista, jota useimmat oppiaineen kriteerit edellyttivät ja arvosanassa 6 oppilas osoitti joidenkin oppiaineen kriteereiden edellyttämää osaamista. Mikäli oppilas yritti joidenkin tavoitteiden saavuttamista, mutta hän pystyi osoittamaan vain jossain määrin kriteereiden edellyttämää osaamista, hän sai arvosanan 5. (Perusopetuksen päättöarvioinnin kriteerit 1999, 6.)

Vuoden 1999 päättöarvioinnin kriteerit olivat suosituksia, ja niiden käytöstä päätettiin paikallisesti, jolloin kriteereihin perustuva järjestelmä ei ratkaissut päättöarvosanojen valtakunnallisen vertailtavuuden ongelmaa. Opetushallitus alkoi kehittää oppilasarvioinnin yhtenäistämiseksi valtakunnallista koepankkia, josta kunnat ja koulut olisivat voineet tilata eri oppiaineiden tehtäväsarjoja oppilaidensa osaamistason määrittelyn helpottamiseksi ja päättöarvosanojen antamisen tueksi. (Opetushallitus 1996, 15–16, 19; Opetushallitus 1997; Koulutuksen tuloksellisuuden arviointimalli 1998, 47.)

Vuonna 1999 voimaan tulleilla perusopetuslailla (628/1998) ja -asetuksella (852/1998) sekä Perusopetuksen oppilaan arvioinnin perusteilla (1999) ja Perusopetuksen päättöarvioinnin kriteereillä (1999) luotiin pohja nykyiselle oppilaan arviointia ohjaavalle järjestelmälle, jossa yhdistyvät formaali oppilaan arviointi ja pedagoginen arviointimalli. Formaalia oppilaan arviointia edustavat desentralisoitu yhtenäinen lainsäädäntöpohja ja kansalliset normit, jotka on määritelty Perusopetuksen opetussuunnitelman perusteissa (2004), mutta myös oppiainekohtaisesti

laaditut hyvän osaamisen kuvaukset arvosanalle kahdeksan. Pedagoginen arviointi ilmenee perustetekstin rakenteessa, jossa arviointi jaetaan opintojen aikaiseen arviointiin ja päättöarviointiin ja joita opettajat tulkitsevat arviointityössään itsenäisesti.

Suomalaiseen desentralisoituun järjestelmään kuuluu, että Suomessa ei ole ollut enää yli kahteenkymmeneen vuoteen tarkastusjärjestelmää, joka kontrolloisi kouluja tai opetuksenjärjestäjiä. Kouluja koskeva päätöksenteko on hajautettu kunnille ja kouluille, joilla on vastuu ja valta esimerkiksi oppilaan arvioinnissa.

3.1 Oppilaan arviointi perusopetuslaissa ja -asetuksessa

Perusopetuksen oppilaan arviointia ohjataan seuraavilla kolmella, toistaiseksi voimassa olevalla säädöksellä: perusopetuslailla (628/1998), perusopetusasetuksella (852/1998) sekä Perusopetuksen opetussuunnitelman perusteilla. Lisäksi oppilaan suorituksen arviointiin sovelletaan, mitä Opetushallitus lakien nojalla määrää. Arviointia koskevissa säädöksissä korostuvat mahdollisuus erilaisiin joustaviin arviointikäytäntöihin ja pyrkimys arvioinnin yhdenmukaisuuteen sekä oppilaiden ja opiskelijoiden yhdenvertaiseen kohteluun (Lankinen & Lahtinen 2009, 236). Kun perusopetusta ohjaavaa lainsäädäntöä valmisteltiin 1990-luvun lopulla, käytiin siinä yhteydessä vilkasta keskustelua oppilaan suoritusten arvioinnista. Vaikka opetusministeriön työryhmä esitti viisiportaiseen arvosteluasteikkoon siirtymistä, sitä ei otettu käyttöön eduskunnan sivistysvaliokunnan mietinnössä esitettyjen kannanottojen takia ja näin arviointiasteikko säilyi entisen kaltaisena eli numeroarvosanoina neljästä kymmeneen. (Lankinen & Lahtinen 2009, 237.)

Oppilaan arvioinnista säädetään perusopetuslain 22 §:ssä, jossa todetaan, että arvioinnilla pyritään ohjaamaan ja kannustamaan opiskelua sekä kehittämään oppilaan edellytyksiä itsearviointiin (perusopetuslaki 628/1998). Samalla korostetaan oppimisen, työskentelyn ja käyttäytymisen monipuolista arviointia. Atjosen mukaan perusopetuslain 22 pykälä on keskeinen, sillä siinä korostetaan oppimisen ja ohjaamisen kannustamista, kehittämistä, itsearviointia, ohjaamista ja monipuolisuutta (Atjonen 2007, 61). Arvioinnin monipuolisuus on Atjosen mielestä eettisesti hyvää, sillä se antaa oppilaille mahdollisuuden osoittaa osaamistaan yksilöllisesti tai pienryhmissä eri tavoilla kuten kirjallisesti, sanallisesti, portfolioilla tai toiminnallisesti (Atjonen 2007, 63–65). Lankinen ja Lahtinen tulkitsevat pykälässä mainittua arvioinnin monipuolisuutta niin, ettei arviointi saa perustua ainoastaan kokeissa osoitettuun osaamiseen (Lankinen & Lahtinen 2009, 237).

Koesuoritukset saattavat luoda virheellisen mielikuvan objektiivisuudesta ja siitä, mikä on oppilaan osaamistaso oppiaineessa. Kokeet ovat kuitenkin osa tietoon kohdistuvaa arviointia. On tärkeää voida arvioida sitä, miten oppimiselle asetettuja tavoitteita saavutetaan sekä sitä, miten oppimistulokset vastaavat tutkinnoille asetettuja tavoitteita. Suulliset ja kirjalliset kokeet ovat yhä yleinen tapa arvioida oppilaan osaamista. Tiedon ja arvioinnin suhdetta tutkinut Kvale (1995) on tutkimuksissaan päätyynyt siihen, että kokeet ovat yhteiskunnallisesti ja kulttuurisesti merkityksellisiä, koska niiden avulla määritellään ja tulkitaan sitä, mitä pidetään kunkin oppiaineen yhteisenä tietopääomana (Kvale 1995 Virran 1999, 6 mukaan). Kokeiden ja muiden arviointien kautta valitaan ja jäsennetään tietoa, jolloin niillä on myös tietoa konstruoiva ja tiedon omaksumista motivoiva vaikutus (Virta 1999, 6–7). Kokeissa oppilaalta edellytetään usein täsmällistä tietoa (Virta 1999, 14). Atjonon esittää, että kirjallisia kokeita tai tenttejä voitaisiin kehittää niin, että ne antaisivat arviointityöhön monipuolisia ja erimuotoisia arvioinnin välineitä (Atjonon 2007, 63). Kokeet voivat toisinaan olla esimerkiksi aktiivisessa ryhmässä olevalle hiljaiselle tai vetäytyvälle oppilaalle ainoa keino päästä osoittamaan omaa osaamistaan.

Opintosuoritusten arvioinnissa ja opinnoissa etenemisestä on lisäksi voimassa, mitä perusopetusasetuksessa (852/1998) säädetään ja Opetushallitus määrää. Arvioinnista säädetään tarkemmin perusopetusasetuksen pykälissä 10–14, joissa arviointi on jaettu opintojen aikaiseen arviointiin ja opinnoissa etenemisedellytysten määrittelyyn sekä päättöarviointiin. Arvosanan antamisen keskeinen tehtävä on välittää tietoa oppilaan suoriutumisesta esimerkiksi huoltajille. Suomessa siitä säädetään perusopetusasetuksen (852/1998) pykälässä 10. Opintojen aikaisessa arvioinnissa lähtökohtana on, että oppilaan ja hänen huoltajansa tulee saada riittävän usein tietoa oppilaan opintojen edistymisestä, työskentelystä ja käyttäytymisestä. Opetussuunnitelmassa määrätään tietojen antamisen ajankohdat ja käytänteet. Suomessa kunnille ja kouluille on lainsäädännössä siirretty itsenäistä valtaa, ja ne voivat itse päättää esimerkiksi siitä, milloin oppilaita arvioidaan kouluvuoden aikana. Ainoastaan lukuvuositodistuksesta määrätään asetuksessa siten, että perusopetuksen oppilaiden tulee saada lukuvuositodistus aina kouluvuoden päättyessä. Todistuksessa arviot annetaan numeroin tai sanallisesti kuten asetuksen (852/1998) 10 pykälän 2 momentissa määrätään. Suomessa todistukseen merkitään oppilaan opinto-ohjelma ja oppiaineittain tai aineryhmittäin arvio siitä, miten oppilas on saavuttanut asetetut tavoitteet, sekä arvio oppilaan käyttäytymisestä. Numeroarvostelussa käytetään asteikkoa 4–10. Arvosana 5 osoittaa välttäviä, 6 kohtalaisia, 7 tyydyttäviä, 8 hyviä, 9 kiitettäviä ja 10 erinomaisia tietoja ja taitoja. Arvosana 4 vastaa hylättyä suoritusta. Sanallista arviota voidaan käyttää, jollei Opetushallitus toisin määrää: yleensä arvioinnissa vuosiluokilla 1–7, valinnaisen aineen tai aineryhmän arvioinnissa, perusopetuslain 25 §:n 2 momentissa tarkoitettujen pidennettyjen oppivelvollisuuden piirissä olevien oppilaiden arvioinnissa, niiden oppilaiden arvioinnissa, joiden äidinkieli on muu kuin suomi tai ruotsi tai muu kuin opetuksessa käytettävä kieli ja aina numeroarvostelun lisäksi.

Koulutusjärjestelmästä ja arvosana-asteikoista riippumatta arvosanan keskeinen tehtävä on kertoa oppilaalle ja hänen sidosryhmilleen opinnoille asetettujen tavoitteiden saavuttamisesta hyväksytysti ja siitä, että oppilas voi siirtyä opinnoissaan eteenpäin (Guskey 2011, 86). Näin on myös suomalaisessa koulutusjärjestelmässä. Perusopetusasetuksen (852/1998) pykälässä 11 säädetään, että oppilas, joka on saavuttanut vuosiluokan oppimäärään sisältyvissä eri oppiaineissa vähintään välttäviä tietoja tai taitoja osoittavan numeron tai vastaavan sanallisen arvion, siirretään seuraavalle vuosiluokalle. Pykälässä 12 säädetään perusopetuksen suorittamisesta ja päättötodistuksesta siten, että perusopetuksen koko oppimäärän hyväksytysti suorittanut oppilas saa päättötodistuksen.

Suomalaisen koulutusjärjestelmän ominaispiirre, opettajien itsenäinen asema oppilaan arvioinnissa, perustuu lainsäädäntöön. Perusopetusasetuksen (852/1998) 13 pykälässä todetaan, että oppilaan arvioinnista päättää kunkin oppiaineen tai ainekokonaisuuden osalta oppilaan oma opettaja tai, jos heitä on useita, opettajat yhdessä. Samoin menetellään käyttäytymisen arvioinnissa. Oppilaalla ja hänen huoltajallaan on 13 §:n 2 momentin mukaan oikeus saada tietoa koulun käyttämistä arviointiperusteista sekä niiden soveltamisesta oppilaan arvioinnissa. (Lankinen & Lahtinen 2009, 239.)

3.2 Oppilaan arviointi Perusopetuksen opetussuunnitelman perusteissa (2004)

Perusopetuksen opetussuunnitelman perusteissa (2004) on kiinnitetty erityistä huomiota arvosanojen valtakunnalliseen vertailukelpoisuuteen ja oppilaiden tasavertaiseen kohteluun. Perusteisiin on sisällytetty kuvaukset oppilaan hyvästä osaamisesta eri oppiaineiden tuntijaon määrittelemissä nivelkohdissa. Numeroarviointia käytettäessä hyvän osaamisen kuvaus määrittelee tason arvosanalle 8. Sanallisessa arvioinnissa kuvaus oppilaan hyvästä osaamisesta tukee opettajaa hänen arvioidessaan oppilaan edistymistä, ja se on arvioinnin perusta kuvattaessa, miten oppilas on saavuttanut tavoitteet. Nykyisten opetussuunnitelman perusteiden (2004) mukaan valtakunnallisesti käyttöön otetut kriteerit ovat oikeudellisesti sitovia määräyksiä. Oppiainekohtaisilla päättöarvioinnin kriteereillä määritellään kansallisesti se tietojen ja taitotasoa, jonka oppilaan oletetaan saavuttaneen perusopetuksen päättövaiheessa. Oppiainekohtaisesti määritellyt päättöarvosanan kriteerit ovat oppilaan arvioinnin pohjana. (Lankinen & Lahtinen 2009, 240.)

Kriteerien käyttöönotto toi uuden elementin suomalaisen arviointikäytäntöeseen 2000-luvulla (Virta 1999, 96). Opetushallitus on määritellyt kriteerit mittariksi tai apuvälineeksi, jolla osaamisen tasoa on pyritty kuvaamaan (Opetushallitus 1996, 11). Kullekin oppiaineelle määriteltyjen hyvän osaamisen vaatimustasoa kuvaavien kriteerien tarkoituksena on ollut myös opettajan arviointityön tukeminen, sen lisäksi, että oppiainekohtaisilla päättöarvioinnin kriteereillä on haluttu parantaa päättöarvosanojen vertailukelpoisuutta ja tätä kautta oppilaiden yhdenvertaisuutta, kun he hakeutuvat jatko-opintoihin. Opetushallituksen päättöarvioinnin kriteereitä valmistelleen työryhmän mukaan kriteerien ensisijainen tehtävä on osaamisen arviointi yksilötasolla. Sen lisäksi oppiainekohtaisten kriteerien tulee vastata kysymyksiin, millä tai miten osaamista arvioidaan. (Opetushallitus 1996, 11.) Kriteereillä käsitetään tässä yleisiä oppiaineitten oppimäärän hallintaa koskevia tavoitteita, jotka määritellään Perusopetuksen opetussuunnitelman perusteissa (2004). Edellä kuvattuun perustuen suomalaista oppilaan arviointijärjestelmää voidaan määritellä toisaalta kriteereihin perustuvaksi, toisaalta opetussuunnitelmaan perustuvaksi.

Perusopetuksen opetussuunnitelman perusteissa (2004) oppilaan arviointi jaetaan kahteen tehtäviltään erilaiseen osaan: opintojen aikaiseen arviointiin ja päättöarviointiin. Opintojen aikaisen arvioinnin tehtävänä on ohjata ja kannustaa opiskelua ja kuvata, miten hyvin oppilas on saavuttanut henkilökohtaiselle kasvulle ja oppimiselle asetetut tavoitteet. Päättöarvioinnin tehtävänä on puolestaan määritellä, miten oppilas on perusopetuksen päättyessä saavuttanut eri oppiaineissa asetetut oppimäärien tavoitteet. (Numminen & Ouakrim-Soivio 2009, 52.)

Taulukkoon 2 on kerätty keskeisimmistä oppilaan arviointia ohjaavista säädöksistä ne kohdat, jossa määritellään oppilaan opintojen aikaista ja päättöarviointia koskevat tehtävät ja periaatteet.

TAULUKKO 2. Oppilaan arviointia ohjaavat keskeisimmät säädökset sisältöalueittain.

Lähde: Numminen & Ouakrim-Soivio 2009, 53.

Oppilaan arvioinnin tehtävät	
<p>Arviointi opintojen aikana</p> <ul style="list-style-type: none"> ohjata ja kannustaa opiskelua. kuvata, miten hyvin oppilas on saavuttanut kasvu- ja oppimiselle asetetut tavoitteet 	<p>Päätösarviointi</p> <ul style="list-style-type: none"> määritellä, miten oppilas on opiskelun päättyessä saavuttanut perusopetuksen oppimäärän tavoitteet eri oppiaineissa.
Oppilaan arvioinnin periaatteet	
<p>Arviointi opintojen aikana</p> <ul style="list-style-type: none"> pyrkii olemaan totuudenmukaista ja perustuu monipuoliseen näyttöön kohdistuu oppilaan oppimiseen ja edistymiseen oppimisen eri osa-alueilla ottaa huomioon arvioinnin merkityksen oppimisprosessissa muodostaa kokonaisuuden, jossa on tärkeää opettajan antama jatkuva palaute on keino, jolla opettaja ohjaa oppilasta tiedostamaan tämän omaa ajattelua ja toimintaansa sekä auttaa oppilasta ymmärtämään oppimistaan ottaa huomioon oppilaan edistymisen, työskentelyn ja käyttäytymisen suhteessa opetussuunnitelman tavoitteisiin ja kuvauksiin oppilaan hyvästä osaamisesta ottaa huomioon opetussuunnitelmassa määritellyt yleiset ja oppiainekohtaiset arvioinnin periaatteet. lisäksi oppilaalle ja hänen huoltajalleen tulee antaa etukäteen tietoa arvioinnin perusteista ja jälkikäteen pyydettäessä on selvitetävää, miten niitä on arvioinnissa sovellettu. 	<p>Päätösarviointi</p> <ul style="list-style-type: none"> on valtakunnallisesti vertailukelpoista ja kohtelee oppilaita tasavertaisesti perustuu kussakin yhteisessä oppiaineessa oppilaan osaamiseen päättövaiheessa vuosiluokilla 8 ja 9 perustuu kaikkiin yhteisiin oppiaineisiin laadittuihin kriteereihin perustuu oppilaan arviointiin perusopetuksen päätösarvioinnin kriteereiden pohjalta ja monipuoliseen näyttöön ottaa huomioon työskentelyn arvioinnin, joka sisältyy oppiaineen arvosanaan.
<p>Opintojen aikana numeroarviointia koskevat periaatteet</p> <ul style="list-style-type: none"> numeroarvosana kuvaa osaamisen tasoa. hyvän osaamisen kuvaus määrittelee tason arvosanalle kahdeksan (8). oppiaineet, aineryhmät ja käyttäytyminen voidaan arvioida numeroilla. yhteisten aineiden arvioinnissa tulee todistuksissa käyttää numeroarvostelua viimeistään kahdeksannella vuosiluokalla. 	<p>Päätösarvioinnissa</p> <ul style="list-style-type: none"> numeroin arvioidaan yhteiset oppiaineet (äidinkieli ja kirjallisuus, toinen kotimainen kieli, ensimmäinen vieras kieli, matematiikka, fysiikka, kemia, biologia, maantieto, terveystieto, uskonto tai elämäntutkimus, historia, yhteiskuntaoppi, musiikki, kuvataide, käsityö, liikunta sekä kotitalous) sekä ne valinnaiset aineet, jotka muodostavat yhtenäisen, vähintään kahden vuosiviikkotunnin oppimäärän.
Sanallinen arviointi	
<p>Opintojen aikana arvioitaessa sanallista arviointia voidaan käyttää</p> <ul style="list-style-type: none"> aina numeroarvioinnin lisäksi kaikissa yhteisissä oppiaineissa vuosiluokilla 1–7 valinnaisena aineena opetettavan aineen tai aineryhmän arvioinnissa perusopetuslain 25 §:n 2 momentissa tarkoitetun pidennetyn oppivelvollisuuden piirissä olevien oppilaiden arvioinnissa niiden oppilaiden arvioinnissa, joiden äidinkieli on muu kuin suomen ja ruotsin kieli. Lukuvuositodistuksesta tulee ilmetä, onko oppilas saavuttanut vuosiluokan tavoitteet hyväksytysti. kuvaus oppilaan hyvästä osaamisesta tukee opettajaa hänen arvioidessaan oppilaan edistymistä, ja se on arvioinnin perusta kuvattaessa, miten oppilas on saavuttanut tavoitteet. sanallisella arvioinnilla kuvataan myös oppilaan edistymistä ja oppimisprosessia. 	<p>Päätösarvioinnissa</p> <ul style="list-style-type: none"> sanallisesti voidaan arvioida oppimäärältään alle kahden vuosiviikkotunnin käsittävät valinnaiset aineet.
<p>Muita arvioinnissa huomioon otettavia asioita</p> <ul style="list-style-type: none"> arviointipalautetta tulee antaa oppilaalle ja hänen huoltajalleen riittävästi ja monipuolisesti. tietoa tulee antaa oppilaan edistymisestä, vahvuuksista sekä niistä osa-alueista, joita on kehitettävä. arviointipalautetta voidaan antaa välitodistuksin, tiedottein ja arviointikeskusteluissa tai muilla tavoin 	

3.2.1 Arviointi opintojen aikana

Oppilaan arviointi on kontekstisidonnaisia ja se liittyy olennaisesti kulloiseenkin koulutusjärjestelmään, sen toimintaympäristöön ja oppimiskäsitykseen (Jakku-Sihvonen & Heinonen 2001, 78–80; Klapp Lekholm 2008, 12–13). Näin myös Opetushallitus toteaa oppilaan arvioinnista verkkosivuillaan.⁴² Arviointia ja arvosanan antamista ohjaavien asiakirjojen tulisi ajasta ja paikasta riippumatta vastata ainakin arvioinnin peruskysymyksiin eli miksi, mitä ja miten arvioidaan ja arvosanoja annetaan (Kuusela 2001; Robson 2001; Atjonen 2007, 21–22; Klapp Lekholm 2008, 11–13). Calfee & Masuda (1997, 82–83) sekä Pickford & Brown (2006, 21) liittävät edellisiin kysymyksiin vielä kysymykset keitä varten arvioidaan, milloin arvioidaan ja kuka arvioi.

Opetushallitus lähestyy arvioinnin tehtäviä mitä ja miten -kysymysten avulla eli mitä arvioidaan ja miten arvioidaan.⁴³ Opetushallituksen mukaan arvioinnilla on kaksi tehtävää: oppimisen tukeminen ohjaavan ja kannustavan palautteen avulla ja osaamisen tason määrittely ja sen ilmaisemiseen. Osaamisen tason ilmaiseminen korostuu päättöarvioinnissa, kun oppilaat saavat päättöarvosanat, joiden perusteella heidät valitaan jatko-opintoihin.⁴⁴ Kuten taulukossa 3 esitettiin, voimassa oleva Perusopetuksen opetussuunnitelman perusteet (2004) jakaa oppilaan arvioinnin kahteen osaan, arviointiin opintojen aikana (POPS 2004, 260–264) ja päättöarviointiin (POPS 2004, 264–266). Seuraavassa tarkastellaan sitä, miten opetussuunnitelman perusteet ohjaavat opintojen aikana suoritettavaa arviointia ja miten päättöarviointia opetussuunnitelmien perusteissa ohjataan.

Kysymykseen miksi oppilasta arvioidaan opintojen aikana Perusopetuksen opetussuunnitelman perusteet (2004) vastaavat, että sillä ”ohjataan ja kannustetaan opiskelua, kuvataan miten hyvin oppilas on saavuttanut kasvu- ja oppimiselle asetetut tavoitteet”. Arviointi ”auttaa oppilasta muodostamaan realistisen kuvan oppimisestaan ja kehittymisestään” ja ”tukee oppilaan persoonallisuuden kasvua”. Arvioinnin tehtävänä on auttaa opettajaa ”ohjaamaan oppilasta tiedostamaan omaa ajatteluaan ja toimintaansa sekä auttaa oppilasta ymmärtämään oppimistaan”. (POPS 2004, 260.) Perusopetuksen opetussuunnitelmien perusteissa oppilaan opintojen aikaisen arvioinnin keskeisimpiä tehtäviä on yksilön persoonallisen kasvu- ja kehitysprosessin tukeminen (Jakku-Sihvonen & Heinonen 2001, 80; Atjonen 2005, 144). Oppilaan ja oppimisen arvioinnin lisäksi arvioinnin tehtäviin kuuluu myös opettajan ammatillisen kehittymisen tukeminen (Atjonen 2005, 144).

42 Oppilaan arvioinnin tehtävää käsittelevä artikkeli löytyy Opetushallituksen sivustolta: http://www.oph.fi/saadokset_ja_ohjeet/ohjeita_koulutuksen_jarjestamiseen/perusopetuksen_jarjestaminen/perusopetuksen_oppilaan_arviointi/oppilaan_arviointi_menneina_vuosikymmenina.

43 Oppilaan arvioinnin tehtävä on määritelty Opetushallituksen sivustolla olevassa luvussa ”Oppilaan arvioinnin tehtävä, joka löytyy osoitteesta: http://www.oph.fi/saadokset_ja_ohjeet/ohjeita_koulutuksen_jarjestamiseen/perusopetuksen_jarjestaminen/perusopetuksen_oppilaan_arviointi/oppilaan_arviointi_menneina_vuosikymmenina.

44 Oppilaan arvioinnin tehtävien määrittely löytyy oppilaan arviointia käsittelevältä Opetushallituksen sivustolta: http://www.oph.fi/saadokset_ja_ohjeet/ohjeita_koulutuksen_jarjestamiseen/perusopetuksen_jarjestaminen/perusopetuksen_oppilaan_arviointi/oppilaan_arviointi_menneina_vuosikymmenina.

Oppilaan arvioinnin tehtäviä sekä Suomessa että muissa koulutusjärjestelmissä on kertoa oppilaalle, hänen huoltajilleen ja muille sidosryhmille siitä, miten opinnoissa on menestytty (Calfee & Masuda 1997, 78–79; Loyd & Loyd 1997, 481) sekä vahvistaa oppilaan tai opiskelijan siirtyminen seuraavalle luokalle tai seuraavaan koulutusmuotoon (Broadfoot 1996, 25; Klapp Lekholm 2008, 21). Perusopetuksen opetussuunnitelman perusteissa opinnoissa etenemisestä vuosiluokittain on pääperiaatteena, että oppilas siirtyy seuraavalle vuosiluokalle, mikäli hän on suorittanut hyväksytysti kaikki opetussuunnitelmassa määritellyt vuosiluokan oppimäärään kuuluvat eri oppiaineiden tai aineryhmien opinnot (POPS 2004, 261). Tietyissä poikkeustilanteissa on mahdollista, että oppilas voidaan siirtää seuraavalle vuosiluokalle, vaikka kaikkia opintoja ei olisikaan suoritettu hyväksytysti. On myös mahdollista, että oppilasta ei siirretä eteenpäin, vaikka oppilaan suoritukset olisivat hyväksytyjä. (POPS 2004, 261–262.) Opintojen aikaisen arvioinnin keskeisenä tehtävänä on antaa arviointipalautetta oppilaalle ja hänen huoltajalleen riittävästi ja monipuolisesti. Tietoa tulee antaa oppilaan edistymisestä, vahvuuksista ja kehittämisaalueista. (POPS 2004, 261.)

Kysymykseen miten arvioidaan, perusteet antavat vastauksena: totuudenmukaisesti ja monipuoliseen näyttöön perustuen. Vuoden 2004 jälkeen perusopetuksen arviointi ja arvosanojen antaminen on perustunut kuvaukseen oppilaan hyvästä osaamisesta, jolla määritellään se tieto- ja taitopohja, johon oppilaan arviointi perustuu (POPS 2004, 260; Lankinen & Lahtinen 2009, 240). Suomessa arvioidaan numeroin, sanallisesti tai niitä yhdistellen (Perusopetusasetus 852/1998; POPS 2004, 260). Numeroin voidaan arvioida oppiaineet, aineryhmät ja käyttäytyminen. Numeroarvosanalla kuvataan osaamisen tasoa. Mikäli oppilaalle annetaan arvosana numerolla, hyvän osaamisen kuvaus määrittelee tason arvosanalle 8, jos taas oppilas arvioidaan sanallisesti, kuvaus oppilaan hyvästä osaamisesta kohdentuu oppilaan edistymiseen ja asetettujen tavoitteiden saavuttamiseen. Kuvaukset oppilaan hyvästä osaamisesta on laadittu jokaisen oppiaineen osion päätteeksi tuntijaon nivelkohtaan, ja ne siis vaihtelevat oppiaineittain ja vuosiluokittain. Sanallisella arvioinnilla voidaan kuvata myös oppilaan edistymistä ja oppimisprosessia. Yhteisten aineiden arvioinnissa tulee todistuksissa käyttää numeroarviointia viimeistään kahdeksannelle vuosiluokalle. Jos sanallista arviointia käytetään lukuvuositolustuksessa, siitä tulee ilmetä, onko oppilas saavuttanut vuosiluokan tavoitteet hyväksytysti. (POPS 2004, 260.)

Atjosen mukaan arvioinnin tulee aina kiinnittyä asetettuihin tavoitteisiin, jolloin arviointi palvelee myös seuraavien tavoitteiden asettamista (Atjonen 2005, 144; Atjonen 2007, 76). Myös Kansanen (1998) liittyy arvioinnin ydinolemukseen tavoitteiden vertaamisen tuloksiin tai päinvastoin (Kansanen 1998, 429). Opetussuunnitelman perusteiden mukaan arviointi kohdistuu oppilaan oppimiseen ja edistymiseen oppimisen eri osa-alueilla. Perusopetuksen opetussuunnitelman perusteissa (2004) määritellään arvioinnin kohteiksi oppilaan edistyminen, työskentely ja käyttäytyminen suhteessa opetussuunnitelman tavoitteisiin ja kuvauksiin

oppilaan hyvästä osaamisesta. (POPS 2004, 260.) Perusopetuksessa kaikki oppiaineet arvioidaan ympäristö- ja luonnontieto -aineryhmää lukuun ottamatta toisistaan erillisinä. Käytännössä tämä tarkoittaa sitä, että vuosiluokilla 1–4 arvioidaan yhtenä kokonaisuutena ympäristö- ja luonnontieto ja vuosiluokilla 5–6 yhtenä kokonaisuutena arvioidaan biologia ja maantieto, samoin fysiikka ja kemia. Kun tullaan perusopetuksen vuosiluokille 7–9, arvioidaan biologia, maantieto, fysiikka, kemia ja terveystieto kaikki erillisinä oppiaineina. (POPS 2004, 262.)

Työskentelyn arviointi on osa oppilaan oppimistaitojen arviointia, ja sen pohjana ovat tavoitteet, jotka työskentelylle eri oppiaineissa on asetettu. Työskentelyn arviointi kohdistuu oppilaan taitoon suunnitella, säädellä, toteuttaa ja arvioida omaa työtään sekä siihen, miten vastuullisesti oppilas työskentelee ja miten hän toimii yhteistyössä toisten kanssa. Työskentelyn arviointi on yleensä osa oppiaineen arviointia, vaikka sitä voidaan arvioida myös oppiaineesta erillisenä. Käyttätymisen arviointi kohdistuu puolestaan siihen, miten oppilas ottaa huomioon muut ihmiset ja ympäristön sekä noudattaa sääntöjä. Oppilaan käyttäytymistä arvioivat kaikki oppilasta opettavat opettajat. Opetussuunnitelmassa tulee asettaa tavoitteet oppilaan käyttäytymiselle. Käyttätymisen tavoitteita määriteltäessä otetaan huomioon koulun kasvatukselle asetetut tavoitteet. (POPS 2004, 262.) Käyttätyminen arvioidaan itsenäisesti eikä osana oppiaineen arviointia.

3.2.2 Päätösarviointi

Päätösarvioinnin tehtävänä on määritellä, miten hyvin oppilas on opiskelun päättyessä saavuttanut kyseisessä oppiaineessa perusopetuksen oppimäärän tavoitteet (POPS 2004, 264). Päätösarvioinnin tulee olla valtakunnallisesti vertailukelpoista ja sen tulee kohdella oppilaita tasavertaisesti (POPS 2004, 264). Lankisen ja Lahtisen mukaan perusopetusta ohjaavassa lainsäädännössä ja normeissa korostuvat pyrkimys arvioinnin yhdenmukaisuuteen sekä oppilaiden ja opiskelijoiden yhdenvertaiseen kohteluun (Lankinen & Lahtinen 2009, 236, 240). Päätösarvioinnin käyttäminen jatkokoulutukseen valikoinnin perusteena asettaa arvioinnille omat haasteensa, joista tärkeimpiä ovat oppilaiden oikeusturva ja arvosanojen vertailukelpoisuus (Jakku-Sihvonen 2001, 85). Virta huomauttaa, että niin kauan kuin koulutukseen valikoidutaan kouluarvosanojen perusteella, oppilaiden kasvun tukeminen arvioinnin avulla on vaikeaa (Virta 1999, 39).

Perusopetuksen opetussuunnitelman perusteet ohjaavat arvioimaan tietojen lisäksi myös taitoja sekä opintojen aikana että päätösarvioinnissa (POPS 2004, 260, 264). Perusteissa todetaan, että ”päättöarvosanan tulee perustua kussakin yhteisessä oppiaineessa oppilaan osaamiseen perusopetuksen päättövaiheessa vuosiluokilla 8–9” (POPS 2004, 264). Oppilaan osaaminen arvioidaan perusopetuksen päätösarvioinnin kriteereiden pohjalta ja monipuoliseen näyttöön perustuen. Päätösarvioinnin kriteerit on laadittu siten, että oppilas saa arvosanan 8, mikäli hän osoittaa keskimäärin oppiaineen kriteereissä edellytettyä osaamista. Mikäli oppilas ei ole

saavuttanut joitain kriteerejä, hän voi kompensoida puuttuvia osa-alueita ylittämällä muissa kriteereissä niiden edellyttämän tason. Oppilas saa arvosanan 5 eli hän on saavuttanut välttävasti perusopetuksessa vaadittavat tiedot ja taidot, mikäli hän pystyy osoittamaan jossakin määrin kriteerien edellyttämää osaamista. (POPS 2004, 264.)

Perusopetuksen päättövaiheessa numeroin arvostellaan kaikki yhteiset oppiaineet.⁴⁵ Valinnaisaineiden arvioinnista opetussuunnitelman perusteissa määrätään siten, että ”Myös ne valinnaiset aineet, jotka muodostavat yhtenäisen, vähintään kahden vuosiviikkotunnin oppimäärän, arvioidaan numeroin. Sen sijaan sanallisesti arvioidaan oppimäärältään alle kahden vuosiviikkotunnin käsittävät valinnaiset aineet ja tällaisista oppimääristä koostuvat kokonaisuudet. Mikäli sanallisesti arvioitu valinnainen aine katsotaan jonkin yhteisen aineen syventäviksi opinnoiksi, sen suoritus voi korottaa kyseisen oppiaineen arvosanaa.” (POPS 2004, 263.)

3.3 Yhteiskuntaopin ainekohtainen arviointi

Tämän tutkimuksen pääpaino on yhteiskuntaopin osaamisen arvioinnissa, jota tarkastellaan seuranta-arvioinnissa saavutetun osaamistason ja oppilaiden saamien yhteiskuntaopin arvosanojen avulla. Historiaa ei käsitellä tässä yhteydessä sen tiedonluonteen enempää kuin perusopetuksessa opetettavan oppiaineen näkökulmista, vaan sitä käytetään yhteiskuntaopissa saatujen empiiristen tulosten vertailemiseen. Yhteiskuntaopin valitseminen oppiaineeksi, jonka oppimistuloksia ja oppilaiden saamia arvosanoja tutkitaan, on tietoista, ja siihen ovat vaikuttaneet useat seikat.

Suomessa yhteiskuntaoppia opetetaan perusopetuksessa yleensä yhdeksännellä luokalla.⁴⁶ Kyseessä on siis oppiaine, joka opiskellaan yleensä yhden lukuvuoden aikana, yleensä saman opettajan johdolla ja yleensä samassa ryhmässä. Lukuvuoden päättyessä oppilaille annetaan yhteiskuntaopin päättöarvosana. Yhteiskuntaoppi on kiinnostava oppiaine myös siksi, että vaikka sitä on opetettu Suomessa jo 1800-luvun lopulta lähtien, vasta vuonna 2004 vahvistetussa Perusopetuksen opetussuunnitelman perusteissa yhteiskuntaopille kirjoitettiin selkeästi omat perustetekstit, jossa oppiaineen ja oppilaan opiskelun tavoitteet, sisällöt ja arviointikriteerit määriteltiin ja oppiaine erotettiin historian oppiaineen yhteydestä. Oppilasarvioinnissa historia ja yhteiskuntaoppi on arvioitu kaikissa kouluissa viimeistään keväästä 2007 lähtien erillisinä oppiaineina, ja molemmista oppiaineista on merkitty päättötodistukseen oma arvosanansa.

45 Numerolla arvioitavat oppiaineet ovat äidinkieli ja kirjallisuus, toinen kotimainen kieli, ensimmäinen vieras kieli, matematiikka, fysiikka, kemia, biologia, maantieto, terveystieto, uskonto tai elämäntutkimustieto, historia, yhteiskuntaoppi, musiikki, kuvataide, käsityö, liikunta sekä kotitalous (POPS 2004, 263).

46 Valtioneuvoston asetus 423/2012 muuttaa toistaiseksi voimassa olevaa tuntijakoa siten, että uuden päätöksen astuessa käytännössä voimaan yhteiskuntaoppia opetetaan peruskoulun neljänneltä vuosiluokalta kuudennelle vuosiluokalle yhteensä kaksi vuosiviikkotuntia ja perusopetuksen yhdeksännellä luokalla yhteensä kolme vuosiviikkotuntia.

Kolmas yhteiskuntaopin osaamiseen ja sen arviointiin liittyvä kiinnostava seikka on sitä opettavien opettajien kelpoisuus. Yleensä perusopetuksen yhdeksännellä luokalla yhteiskuntaoppia opettavat aineenopettajat ovat suorittaneet pääaineopintonsa historiassa. Näin oli myös seuranta-arviointiin osallistuneiden koulujen opettajien kelpoisuuden osalta. Opettajakyselyyn vastanneista opettajista 84 % oli aineenopettajia, ja suurimmalla osalla heistä oli pääaineen opinnot historiassa: historian laudatur-arvosanan tai sitä vastaavan määrän, 55 opintoviikkoa tai 120 opintopistettä oli suorittanut 75 % vastanneista, 60 opintopistettä tai 35 opintoviikkoa eli cum laude -tasoisesti historiaa oli opiskellut 12 % vastanneista ja approbatur-tasoisesti eli 15 opintoviikkoa tai 25 opintopistettä historiaa oli opiskellut 2 % vastanneista. Luokanopettajan monialaiset opinnot oli suorittanut yhteensä 6 % vastanneista. Sen sijaan yhteiskuntaoppiin kuuluvia yhteiskuntatieteitä oli pääaineenaan eli laudatur-tasoisesti tai 55 opintoviikkoa tai 120 opintopistettä opiskellut 11 % opettajista. Suurimmalla osalla vastanneista (67 %) oli yhteiskuntatieteistä niin sanottu sivuaineopintokokonaisuus eli 60 opintopistettä, joka vastaa 35 opintoviikkoa tai cum laude -tasoisia opintoja. (Ouakrim-Soivio & Kuusela 2012, 32–33.)

Peruskoulun opetussuunnitelmien perustetekstit ovat säilyneet rakenteeltaan vuosikymmenten ajan lähestulkoon samanlaisina, sillä niistä löytyvät opetustuntien määrät, opetuksen tavoitteet ja sisällöt, taitojen opettaminen, työtavat, oppiainekohtainen arviointi sekä aihekokonaisuudet. Suurimmat erot opetussuunnitelmien rakenteessa liittyvät siihen, kuinka yksityiskohtaisesti ja missä laajuudessa perustetekstejä on laadittu. Ääripäät lienevät vuoden 1970 Komiteanmietintö, joka oli yksityiskohtainen jo siksi, että sen avulla linjattiin ja ohjattiin tuolloin kaikille uutta peruskoulujärjestelmää. Toisen ääripään muodostaa vuoden 1994 peruskoulun opetussuunnitelma, joka puolestaan on hyvin toteava ja yleisluonteinen.

Vuosien 1970–1994 peruskoulun opetussuunnitelmien perusteissa historia ja yhteiskuntaoppi muodostivat yhden yksittäisen oppiainekokonaisuuden, jossa oppiaineen tavoitteet, sisällöt ja arviointiperusteet kirjattiin pääsääntöisesti opetussuunnitelmien perusteteksteihin yhteisesti. Vuoden 1970 Komiteanmietintö muodostaa kuitenkin sikäli poikkeuksen, että historian ja yhteiskuntaopin perustetekstit kirjoitettiin siinä toisistaan erillisinä ikään kuin kahden eri oppiaineen lähtökohdista käsin (Komiteanmietintö 1970: A5, 215). Seuraavan kerran näiden kahden oppiaineen, historian ja yhteiskuntaopin, erillisuus näkyy vasta vuoden 1999 Perusopetuksen päättöarvioinnin kriteereissä ja vuonna 2004 vahvistetussa Perusopetuksen opetussuunnitelman perusteissa. Vuoden 1970 Komiteanmietinnön perusteteksteissä yhteiskuntatiedollisten tietojen ja taitojen opettaminen ja niiden soveltaminen kytkettiin tiiviisti kansalaistaidon opetukseen. Perustetekstissä korostettiin, että opettajalla oli mahdollisuus täydentää yhteiskuntaopin ja taloustiedon opetusta käsittelemällä kyseistä kokonaisuutta kansalaistaidon viikoittaisella tunnilla. (Komiteanmietintö 1970: A5, 218.) Kansalaistaito oli yhteiskuntaopin ”apuaine”, sillä kansalaistaidon tehtävänä oli auttaa oppilasta muodostamaan kokonaiskuva siitä, miten hän toimii tulevaisuudessa yhteiskunnan jäsenenä. Vuoden 2001 tuntijaossa

ja myöhemmin vuoden 2004 Perusopetuksen opetussuunnitelman perusteissa kansalaistaito poistui opettavien oppiaineiden joukosta ja osa kyseisen oppiaineen sisällöistä siirrettiin opettavaksi osana joko aihekokonaisuuksia tai osana muiden oppiaineiden, kuten yhteiskuntaopin, terveystiedon, kotitalouden tai ympäristö- ja luonnontiedon sisältöjä.

Vaikka Heinonen (2005) on väitöstudiumissaan osoittanut, että oppikirjat ohjaavat opettajien opetustyötä yhä edelleen ja oppikirjojen sisältöjä saatetaan hyödyntää erityisesti koulukohtaisia opetussuunnitelmia laadittaessa (Heinonen 2005, 57, 299), luo opetussuunnitelman perusteet kuitenkin yhtenäiset, valtakunnalliset raamit sille, mitä ja minkä tavoitteiden mukaisesti eri oppiaineita perusopetuksessa opetetaan ja arvioidaan. Koska yhteiskuntaopin oppimistulosten seuranta-arvioinnit ja oppilaille annettava arvosanat perustuvat samoihin ohjausasiakirjoihin eli voimassa olevaan valtioneuvoston asetukseen perusopetuksen tuntijaosta sekä Perusopetuksen opetussuunnitelman perusteisiin (2004), on perustelua tarkastella sitä, mitä tavoitteita yhteiskuntaopille on oppiaineena asetettu, miten niitä arvioidaan sekä kuinka suurella tuntimäärällä yhteiskuntaoppia Suomessa opetetaan.

Kvale (1995) on esittänyt, että arvioitaessa oppilaiden tietoja ja niistä muodostettuja konstruktioita, kartoitetaan samalla oppiaineen kollektiivista tietopääomaa (Kvale 1995 Virran 1999, 6 mukaan). Opetussuunnitelma, johon arviointi perustuu, on Lawtonin (1975) mukaan valikoima kulttuurin sisällöistä, jotka on tarkoitettu koko ikäluokalle opettavaksi (Lawton 1975 Virran 1999, 6 mukaan). Näin ollen esimerkiksi yhteiskunnallisten aineiden seuranta-arviointien tehtävillä on sikäli merkitystä, että niiden laatijat määrittelevät ja tulkitsevat sitä, mitä yhteiskunnassa pidetään kunkin oppiaineen tietopääomana. Lisäksi seuranta-arvioinnit ja niiden tehtävien asettelu heijastavat koululaitoksessa ja yhteiskunnassa vallalla olevia käsityksiä pätevydestä ja siitä, mitä missäkin oppiaineessa tulisi tietää.

Löfström (2001) käyttää yhteiskuntatieto-käsitettä harkittuna nimityksenä ja synonyminä yhteiskuntaopille. Hänen mukaansa yhteiskuntatiedossa korostuvat tieto ja tietäminen aktiivisena toimintana. Yhteiskuntaoppi-käsitteeseen sisältyvä oppi on merkitykseltään normatiivinen ja staattinen. (Löfström 2001a, 9.) Yhteiskuntatiedolla voidaan tarkoittaa myös kaikissa toisen asteen koulutusmuodoissa ja aikuis-koulutuksessa eri nimikkeillä opettavia yhteiskunnallisia aineita historiaa lukuun ottamatta (Löfström 2001a, 9; Löfström 2002, 89–90; Virta 2002, 37).

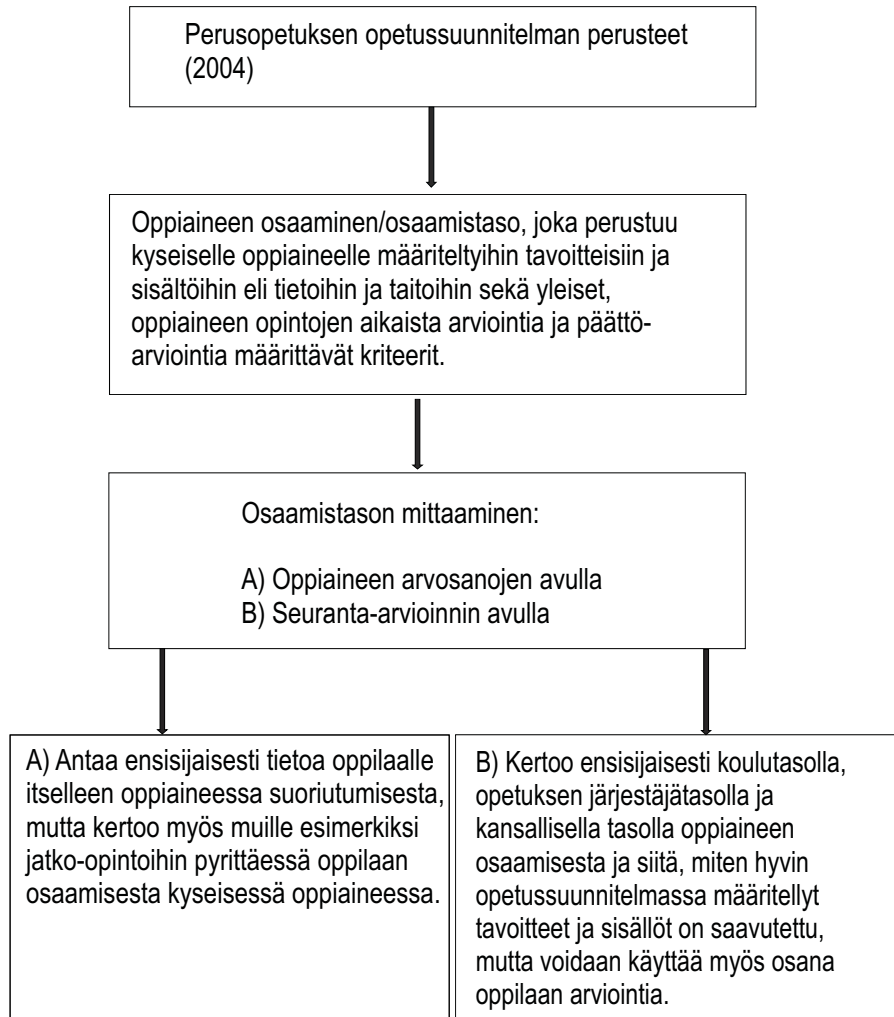
Van den Berg ja Löfström (2011) vertaavat toisiinsa ihmisten käsityksiä historiasta ja yhteiskunnasta, sillä samalla tavalla kuin nuorten käsitykset ihmiskunnan historiasta ovat paljolti muualta kuin koulun historian tunneilta omaksuttuja, myös nuorten yhteiskunnalliset tiedot ja tulkinnat ovat todennäköisesti suurelta osin muualta peräisin kuin kouluopetuksesta (van den Berg & Löfström 2011, 82). Vuonna 1999 tehdyssä CIVED-tutkimuksessa esitettiin sama lähtöoletus siitä, että nuorten yhteiskunnallinen ajattelu määräytyy yhä enenevässä määrin koulun ulkopuolelta.

Vaikuttavina tekijöinä tutkimuksessa pidettiin muun muassa kyseisen maan poliittista kulttuuria, mediaa, kotia, ystäväpiiriä ja nuorten omia kokemuksia yhteiskunnallisesta toiminnasta. (Suutarinen ym. 2001, 3.) Van den Bergin teemahaastattelututkimus vuodelta 2010 osoittaa puolestaan, että koulun rooli yhteiskuntaa ja sen kehitystä analysoivana tiedonlähteenä oli selvästi vähentynyt, sillä nuoret seuraavat yhä enemmän median tarjoamia tietoja ja ovat alttiita esimerkiksi television esittämille tulkinnoille ympäröivästä yhteiskunnasta (van den Berg 2010, 11). Löfström, Virta ja van den Berg lähestyvät yhteiskunnallista osaamista aktiiviseen kansalaisuuteen kuuluvien taitojen kuten poliittisen lukutaidon, sosialisointitaitojen, elinikäisen oppimisen taitojen sekä demokratiataitojen näkökulmasta (Löfström, Virta & van den Berg 2010, 6). Suuri haaste yhteiskuntatiedollisen osaamisen määrittelyssä on se, missä määrin oppilaan tulee osata olla kriittinen, missä määrin taas yhteiskuntatiedolliseen osaamiseen kuuluu sosialisointin ulottuvuus (Löfström, Virta & van den Berg 2010, 7).

Yhteiskuntaopin osaamisella ei kuitenkaan tässä yhteydessä tarkoiteta sitä oppilaan laajempaa yhteiskunnallista toimintatapaa, arvoja tai asenteita, joita hän oppii, omaksuu ja soveltaa koulussa tai sen ulkopuolella toimiessaan. Tässä tutkimuksessa yhteiskuntaopin osaamisella tarkoitetaan sitä Perusopetuksen opetussuunnitelman perusteissa (2004) määriteltyä osaamista, joka perustuu Opetushallituksen määrittelemiin ja määräämiin oppiaineiden tavoitteisiin sekä niihin keskeisiin tietoihin ja taitoihin, jotka oppilaalla tulee olla perusopetuksen päättövaiheessa. Oppiaineiden osaamisen arviointi perustuu Perusopetuksen opetussuunnitelman perusteiden (2004) ohjeisiin opintojen aikaisesta arvioinnista ja päättöarvioinnista, kuten jo aiemmin taulukossa 2 on kootusti esitetty sekä samaan Opetushallituksen ohjausasiakirjaan perustuvaan seuranta-arviointiin. Tässä tutkimuksessa ei kuitenkaan enää kuvata yhteiskuntaopin ja historian seuranta-arvioinnin osaamisetasoa, sillä se on raportoitu perusteellisesti Opetushallituksen julkaisussa Historian ja yhteiskuntaopin oppimistulokset perusopetuksen päättövaiheessa 2011 yhteiskuntaopissa (Ouakrim-Soivio & Kuusela 2012, 60–75) ja historiassa (Ouakrim-Soivio & Kuusela 2012, 49–60).

Yhteiskuntaopin osaaminen ja sen mittaaminen kahdella rinnakkaisella mittarilla, arvosanoilla ja seuranta-arvioinnilla, perustuu Perusopetuksen opetussuunnitelman perusteisiin (2004), sillä muuta yhteistä ja yhteismitallista osaamistason määritelmää tai määrittelijää ei tällä hetkellä suomalaisessa koulutusjärjestelmässä ole. Perusopetuksen opetussuunnitelman perusteet (2004) on se ohjausasiakirja, jolla suomalaisessa koulutusjärjestelmässä pyritään turvaamaan opetuksen tasa-arvo, tasalaatuisuus ja saavutettavuus. Opetussuunnitelman perusteissa määritellyt oppilaan arvioinnin kriteerit pyrkivät turvaamaan oppilaiden yhdenvertaisen ja yhteismitallisen kohtelun oppilasarvioinnissa, kuten tämän tutkimuksen luvussa 3.2 on yksityiskohtaisemmin jo kuvattu.

Kuviossa 2 on esitetty, kuinka Perusopetuksen opetussuunnitelman perusteissa (2004) määritelty oppiaineen osaaminen voidaan arvioida joko arvosanoilla tai seuranta-arvioinnilla. Kummankin rinnakkaisen tavan mitata yhteiskuntaopin osaamista tulisi tuottaa kutakuinkin samanlaisia tuloksia, koska nämä rinnakkaiset mittaamistavat perustuvat samoihin opetussuunnitelman perusteissa määriteltyihin tavoitteisiin, sisältöihin ja kriteereihin.



KUVIO 2. Oppiaineen osaaminen ja osaamistason mittaamismahdollisuudet suomalaisessa koulutusjärjestelmässä.

Yhteiskuntaopin osaamistason määrittelemiseksi seuraavassa tarkastellaan yhteiskuntaopin opetukselle asetettuja tiedollisia ja taidollisia tavoitteita, oppiaineen sisältöjä sekä resursseja, joihin yhteiskuntaopin opetus ja sille määritellyt hyvän osaamisen kriteerit perustuvat.

3.3.1 Yhteiskuntaopin opetuksen keskeiset painopisteet, resurssit ja tavoitteet

Tämän alaluvun tarkoituksena on selvittää sitä, kuinka eri vuosikymmeninä opetussuunnitelman perusteteksteissä on määritelty yhteiskuntaopin oppiaineelle asetetut keskeiset painopisteet, tuntiresurssit ja tavoitteet, joiden perusteella oppilaan arviointia, samoin kuin oppiaineen oppimistuloksia määritellään.

Peruskoulu-uudistus 1970-luvun alussa teki yhteiskuntaopista kaikille yhteisen oppiaineen, jota opetettiin vuoteen 2004 asti historian yhteydessä. Arola toteaa, että yhteiskunnallinen opetus uudistui peruskoulun myötä, mutta se oli edelleen luonteeltaan maltillista eikä sitä haluttu sitoa yhteiskunnallisiin ja poliittisiin kiistoihin, vaan niitä pyrittiin pikemminkin välttämään. 1970-luvulta aina 1990-luvulle yhteiskuntaopin peruslinjaukset säilyivät lähes muuttumattomina. (Arola 2001, 5.) Yhteiskunnallisten kysymysten painoarvo kasvoi 2000-luvun alussa, kun perusopetuksen sekä lukion opetussuunnitelmien painopiste muuttui yhteiskunnallisemmaksi. Tämä näkyi esimerkiksi siinä, että talouteen liittyvien sisältöjen määrä ja merkitys kasvoivat (Arola 2005, 32). Vaikka yhteiskunnallisesti tarpeellisista tiedoista ja taidoista sekä kansalaisyhteiskunnan vahvistamisesta alettiin keskustella, muodostui Arolan mukaan 2000-luvun haasteeksi se, että yhteiskuntatiedosta puhutaan ja kirjoitetaan ilman yhtenäistä kokonaiskuvaa (Arola 2001, 7). Yhteiskunnalliset hankkeet ovat 2000-luvun alussa olleet näkyviä ja ne ovat myös herättäneet keskustelua. Hankkeiden avulla on haluttu muun muassa tukea aktiivisen kansalaisuuden ja kansalaisvaikuttamisen juurtumista osaksi yhteiskuntatiedollista opetusta. Hankkeiden käynnistymisen taustalla olivat esimerkiksi kansainvälisen IEA:n CIVED-tutkimus (1999) ja ICCS-tutkimus (2009), jotka osoittivat, että suomalaisnuoret olivat vain vähän kiinnostuneita politiikasta, jolloin esimerkiksi poliittiset päättäjät huolestuivat nuorten alhaisesta äänestysaktiivisuudesta ja kansalaisten passivoitumisesta. Päättäjien huoli oli osaltaan vaikuttamassa siihen, että valtioneuvosto perusti kansalaisvaikuttamisen politiikkaohjelman (2003–2007), jonka tarkoituksena oli tukea kansalaisten kiinnostusta edustukselliseen demokratiaan ja lisätä oppilaiden osallisuutta koulujen arjessa esimerkiksi oppilaskuntatoiminnan avulla (Nousiainen & Piekkari 2007, 7–8).

Kun valtioneuvosto teki tuntijakopäätöksen 21.8.1969, historiaa ja yhteiskuntaoppia opetettiin tällä oppiainanimikkeellä peruskoulussa yhteensä kymmenen vuosiviikkotuntia. Tuntimäärät jakautuivat tasaisesti vuosiluokille 5–9 siten, että historiaa opetettiin kaksi vuosiviikkotuntia luokilla 5–8 ja yhteiskuntaoppia yhdeksännellä vuosiluokalla niin ikään kaksi vuosiviikkotuntia. Tämän lisäksi 9. vuosiluokan kansalaistaidon yksi vuosiviikkotunti oli tarkoitettu yhteiskunnallisen sisältöjen opettamiseen. Käytännössä yhteiskuntaoppia opetettiin peruskoulussa kolme vuosiviikkotuntia peruskoulun viimeisellä vuosiluokalla. (Komiteanmietintö 1970: A4, 24, 170.) Vuonna 1972 silloinen Kouluhallitus vahvisti peruskoulun opetussuunnitelman kaavan ja määräsi sisällöksi vuoden 1970 komiteanmietinnön toisessa osassa olevat oppimäärät, jotka perustuivat vuoden 1969 tuntijakopäätökseen. Historiaa

ja yhteiskuntaoppia sekä kansalaistaidon yhteiskunnallisia sisältöjä opetettiin edellä kuvatun mukaisesti vuoteen 1985 asti, jolloin seuraavan kerran peruskoulun opetussuunnitelman perusteet määriteltiin uudelleen.

Vuoden 1985 Perusopetuksen opetussuunnitelman perusteita edelsi peruskouluasetuksen muutos (714/1984), jossa vuosiluokilla 7–9 määrättiin valtioneuvoston tuntijakopäätöksen mukaisesti opetettavaksi historiaa ja yhteiskuntaoppia yhteensä kuusi vuosiviikkotuntia. Historiaa opetettiin kaksi tuntia viikossa vuosiluokilla 7–8 ja 9. vuosiluokalla opetettiin kaksi tuntia yhteiskuntaoppia. Oppiaineen nimi säilyi ennallaan. Edellä kuvattu tuntimäärä ja oppiaineen nimi siirtyi myös vuoden 1993 valtioneuvoston päätökseen peruskoulun tuntijaosta (839/1993). Käytännössä tämä on tarkoittanut sitä, että vuodesta 1985 vuoteen 2004 historia ja yhteiskuntaoppi -nimistä oppiainetta on opetettu peruskoulussa lähes 20 vuotta samalla tuntimäärällä ja samalla oppiainenimikkeellä kaikille pakollisena oppiaineena.

Vuoden 2001 valtioneuvoston asetuksessa (1435/2001) vahvistettiin historian ja yhteiskuntaopin opetukseen käytettäväksi yhteensä kymmenen vuosiviikkotuntia. Asetuksessa yhdeksännelle vuosiluokalle määrätty kolmas yhteiskuntaopin tunti on tarkoitettu taloustiedon opettamista varten. Todettakoon kuitenkin, että kansalaistaito oppiaineena hävisi vuoden 2001 tuntijakopäätöksen myötä. Vuoden 2004 opetussuunnitelman perusteet eivät rajaa historiaa opetettavaksi ainoastaan luokilla 7 ja 8 ja yhteiskuntaoppia perusopetuksen 9. vuosiluokalla. Päätös tehdään kunnissa tai kouluissa. Yhteiskuntaopin perustetekstissä taloustiedon tavoitteita ja sisältöjä on lisätty, ja siksi lisätunnin siirtämistä juuri yhteiskuntaoppiin pidettiin perusteltuna. Käytännössä suurin osa kouluista opettaa historiaa vuosiluokilla 7–8 ja yhteiskuntaopin oppimäärän vuosiluokalla 9.

Löfström, Virta ja van den Berg pitävät historian ja yhteiskuntaopin oppiaineiden erottamista hyödyllisenä siksi, että yhteiskuntaopilla on ollut mahdollisuus profiloitua aiempaa selvemmin itsenäisenä oppiaineena (Löfström, Virta & van den Berg 2010,6). Kahden erillisen oppiaineen mallia on myös kritisoitu. Esimerkiksi Suutarinen (2006b, 112, 115) on esittänyt, että kahden erillisen oppiaineen malli saattaa johtaa ainejakoisuuden syvenemiseen ja opetuksen tavoitteiden tietosisältöjen painottumiseen.

OECD:n julkaisun mukaan vuonna 2008 OECD-maissa käytettiin 12–14-vuotiaiden oppilaiden yhteiskuntaopin opetukseen (*social studies*) keskimäärin 12 tuntia viikossa (Education at a Glance 2010, 376). Van den Berg ja Löfström huomauttavat (2011), että tarkasteltaessa vuoden 2001 tuntijakopäätöstä on Suomessa luokilla 3–5 yhteiskunnallista opetusta vain noin neljäsosa siitä määrästä, jota OECD- ja EU-maissa keskimäärin vastaavilla luokka-asteilla tarjotaan (Education at a Glance 2007, 370). Van den Bergin ja Löfströmin mukaan yhteiskunnallisiin asioihin liittyvien aiheiden voidaan sanoa kadonneen alakoulujen kaikille yhteisestä opetuksesta (van den Berg & Löfström 2011, 79–80), kun kansalaistaito poistui oppiaineena vuoden 2001 tuntijakopäätöksessä.

Kartovaara (2004) pitää opetussuunnitelman tavoitteita ja niiden määrittelyä keskeisimpänä koulutusalan ohjauskeinona. Opetussuunnitelman perusteisiin kirjatujen tavoitteiden avulla poliittiset päättäjät ja toimivaltaiset viranomaiset ilmaisevat linjaukset koulutuksen päämääristä. (Kartovaara 2004, 66.) Perusopetuksen opetussuunnitelman perusteisiin (2004) kirjatut, yhteiskuntaopin opetukselle asetetut tavoitteet ovat sekä oppilaan arvioinnin että yhteiskunnallisten aineiden seuranta-arvioinnin kannalta keskeisiä, koska kummassakin tapauksessa ne toimivat arvioinnin kiinnekohtina oppiaineelle erikseen määriteltyjen kriteereiden lisäksi. Oppimistulosten arviointiin laaditut tehtävät perustuvat opetussuunnitelmassa oleviin valtakunnallisiin tavoitteisiin ja siinä määriteltyjen tietojen ja taitojen arviointiin. Oppilaan opintojen aikainen arviointi ja päättöarvioinnin kriteerit perustuvat samoihin oppiaineen tavoitteisiin ja oppiaineelle määriteltyihin kriteereihin kuin oppimistulosten arviointikin (Perusopetuksen päättöarvioinnin kriteerit 1999, 6; POPS 2004, 226–227).

Yhteiskuntaopin tavoitteiden määrittely on muuttunut vuosikymmenten aikana niiltä osin kuin oppiaineelle asetetut tavoitteet on kirjoitettu, milloin historian oppiaineen yhteyteen, milloin taas erikseen. Peruskoulun ensimmäisten vuosikymmenten perustetekstit laadittiin yksityiskohtaisiksi ja laajoiksi ja se näkyi myös yhteiskuntaopin tavoitteiden määrittelyssä. Vuoden 1985 perusteissa määriteltiin ensin oppiaineen kasvatukselliset tavoitteet ja sen jälkeen oppiaineen varsinaiset tavoitteet (POPS 1985, 131). Vuoden 1994 Peruskoulun opetussuunnitelman perusteet poikkesivat paitsi tyyliltään ja rakenteeltaan edellisistä opetussuunnitelmien perusteista. Oppiaineille laadittiin ensin opiskeluun liittyvät tavoitteet ja tämän jälkeen oppiaineen omat, yksityiskohtaisemmat tavoitteet (POPS 1994, 97). Vuoden 1994 perusteissa laadittiin oppiainekohtaiset, yleiset tavoitetekstit, eräänlaiset ingressit, joilla haluttiin korostaa kyseisen oppiaineen ja sen tiedonluonteen erityispiirteitä. Yhteiskuntaopin tiedonalalähtöinen ingressitekstin käyttö siirtyi sittemmin vuoden 2004 perusteisiin. Vuoden 1994 perusteissa yleisiksi opetuksellisiksi tavoitteiksi historiassa ja yhteiskuntaopissa määriteltiin oppilaan ohjaaminen niin, että hän ymmärtää ja jäsentää nykyaikaa, oppilaan oman identiteetin vahvistaminen, oppilaan ohjaaminen kunnioittamaan ihmistä ja ihmisen työtä sekä oppilaan auttaminen, jotta hän kasvais aktiiviseksi, kriittiseksi ja yhteiskuntavastuuta tuntevaksi yhteiskunnan jäseneksi. Historian ja yhteiskuntaopin opetuksen keskeisiä tarkastelukohteita olivat aika, ihminen ja ympäristö. (POPS 1994, 97.) Tavoitteisiin kirjattiin myös, että ”oppilas saa sellaisia tietoja ja taitoja, joita hän tarvitsee toimiessaan ja asioidessaan yhteiskunnassa, ja osaa käyttää historiallista ja yhteiskunnallista tietoa sekä käsitellä historiallisia ja ajankohtaisia ongelmia” (POPS 1994, 97).

Eri vuosikymmenten rakenteellisista eroista huolimatta peruskoulun opetussuunnitelman perusteista vuosilta 1970, 1985 ja 1994 löytyy useita yhteisiä ja vuosikymmenestä toiseen kirjattuja samoja tavoitteita, joita olivat: osaamisen tavoitteet, kuten yhteiskunnallisten instituutioiden tunteminen ja niissä toimiminen, aktiivisuus yhteiskunnallisena toimijana, kansainvälisyys ja taloudellisten peruskäsitteiden sekä talouselämän ilmiöiden ja yksilön talouteen liittyvien asioiden ymmärtäminen

(Komiteanmietintö 1970: A5, 215, 227–228; POPS 1985, 143–145; POPS 1994, 97–98). Tavoitteiden kirjaamistapa on vuosikymmenten aikana muuttunut. Esimerkiksi vuoden 1970 komiteanmietinnössä aktiivisuus ja yhteiskunnassa toimiminen kirjattiin perusteisiin, samoin vastuullinen suhtautuminen yhteiskuntaan (Komiteanmietintö 1970: A5, 215). Vuoden 1985 perusteet laajensivat kansalaisena toimimisen tavoitteita koskemaan yhteiskuntaeettisiä asenteita, kuten tasa-arvoa ja ihmisoikeuksia (POPS 1985, 131). Yhteiskunnan aktiivisen toimijan roolia vuoden 1985 perusteissa laajennettiin käsittämään jopa maailmanlaajuisista yhteisvastuuta, jossa pyrittiin kasvattamaan oppilaita maailmanlaajuisen yhteisvastuun ymmärtäjiksi ja aikanaan sen aktiivisiksi, vastuullisiksi kantajiksi (POPS 1985, 132). Suomalaisen yhteiskunnan kansainvälistyminen ja monikulttuuristuminen on puolestaan näkynyt peruskoulun nelikymmenvuotisessa historiassa siinä, että historian ja yhteiskuntaopin perusteisiin on kirjattu yhteiskunnallisten ilmiöiden tarkastelu milloin kansalliselta, milloin taas kansainväliseltä kannalta (Komiteanmietintö 1970: A5, 217) tai kulttuurien erilaisuuden kunnioittamisen ja ymmärtämisen kannalta (POPS 1985, 131; POPS 1994, 97).

Talousasioiden, talouden peruskäsitteiden ja taloudellisten ratkaisujen ymmärtäminen arkielämässä on nostettu kaikkina vuosikymmeninä yhdeksi keskeisimmistä yhteiskuntatiedollisen osaamisen tavoitteista. Vuoden 1970 perusteissa talouteen liittyvät tavoitteet kirjattiin monisanaisesti paitsi historiaan ja yhteiskuntaoppiin myös kansalaistaitoon ja siellä erityisesti osaksi taloudellisuuskasvatusta, jolla nimellä taulouskasvatus tuolloin tunnettiin. Yhteiskuntaopissa tavoitteena oli muun muassa, että oppilaat hallitsivat tärkeimpiä taloudellisia ilmiöitä koskevat peruskäsitteet, tutustuivat talouselämän toiminnallisiin kokonaisuuksiin ja taloudellisiin ratkaisuihin sekä ymmärsivät kansantaloudellisten tapahtumien keskinäistä vuorovaikutusta. (Komiteanmietintö 1970: A5, 215, 227.) Vuoden 1985 perusteissa yhteiskuntaopin osaamisen tavoitteina oli muun muassa se, että oppilas perehdytettiin yhteiskuntaan ja talouselämään ja ohjattiin ymmärtämään yksilökohtaisia poliittisia ja taloudellisia ratkaisuja sekä valtiollisten ja kansantaloudellisten tapahtumien keskinäistä vuorovaikutusta (POPS 1985, 131–132). Vuoden 1994 perusteissa sama asiasältö määriteltiin hieman toisin sanoin, kun oppiaineen tavoitteeksi asetettiin, että oppilas ymmärsi olevansa kansalainen, kuluttaja ja vaikuttaja, kun tehdään yhteiskunnallisia ratkaisuja (POPS 1994, 98).

Opetussuunnitelmien perusteet heijastavat omaa aikaansa sekä yhteiskunnallisia ja koulutuspoliittisia muutostarpeita. Vuoden 1970 perusteissa osa yhteiskuntaopin tavoitteista kirjattiin yksityiskohtaisesti kansalaistaidon opetussuunnitelmaan (Komiteanmietintö 1970: A5, 170, 180, 185). Jos 1970- ja 1980-lukujen peruskoulun opetussuunnitelmien perusteet olivat yksityiskohtaiset, niin vuoden 1994 perusteet olivat yleisluontoiset ja väljästi määritellyt. Esimerkiksi vuoden 1994 perustekstistä historian ja yhteiskuntaopin oppiaineen opiskelun tavoitteissa mainittiin muun muassa historian ja yhteiskunnallisten kysymysten harrastuneisuuden tukeminen, myönteisten yhteiskuntaeettisten asenteiden, sosiaalisen vastuuntunnon,

tasa-arvon, ihmisoikeuksien ja rauhan arvostaminen. Oppiainekohtaiset tavoitteet liittyivät paitsi mielenkiinnon virittämiseen historiallisissa ja yhteiskunnallisissa kysymyksissä, myös jatkuvan muutoksen ymmärtämiseen ja siihen sopeutuminen sekä oman aseman ja kotiseudun ymmärtämiseen osana nyky-yhteiskuntaa (POPS 1985, 134; POPS 1994, 97–98).⁴⁷ Vuoden 2004 perusopetuksen opetussuunnitelman tavoitteissa määritellään yhteiskuntaopin oppiaineen luonteesta lähtevä opetuksellinen tehtävä, jonka mukaan oppilasta ohjataan kasvamaan yhteiskunnan aktiiviseksi ja vastuulliseksi toimijaksi. Perusopetuksen vuosiluokkien 7–9 yhteiskuntaopin opetuksen tulee antaa perustiedot ja -taidot yhteiskunnan rakenteesta ja toiminnasta sekä kansalaisen vaikutusmahdollisuuksista. Opetuksen tarkoituksena on tukea oppilaan kasvua suvaitsevaiseksi ja demokraattiseksi kansalaiseksi ja antaa kokemuksia yhteiskunnallisesta toimimisesta ja demokraattisesta vaikuttamisesta. (POPS 2004, 226.)

Perustetekstissä on määritelty tarkemmin oppiaineen tavoitteet vuosiluokille 7–9, jossa yhteiskuntaopin tavoitteiksi määriteltiin, että oppilas

- saa käsityksen yhteiskunnallisen tiedon luonteesta
- oppii hankkimaan ja soveltamaan yhteiskuntaa ja talouselämää käsittelevää informaatiota kriittisesti ja toimimaan aktiivisena vaikuttajana
- oppii tuntemaan julkiset palvelut
- saa valmiuksia työnteon kunnioittamiseen
- oppii yrittäjyyden perusteet ja ymmärtää yrittäjyyden merkityksen yhteiskunnan hyvinvoinnin tekijänä
- oppii ymmärtämään yhteiskunnallisten päätösten vaikutuksia kansalaisten elämään
- kiinnostuu yhteiskunnallisesta osallistumisesta ja vaikuttamisesta
- oppii tarkastelemaan ja kehittämään osaamistaan vastuullisena kuluttajana ja yhteiskunnallisena toimijana ja tuntee tekojensa oikeudelliset seuraamukset. (POPS 2004, 226.)

Löfström kiteyttää peruskoulun opetussuunnitelman perusteissa (1994) kirjattujen yhteiskuntaopin opetuksen tavoitteiden ongelman siten, että perusteteksteissä usein käytetyt ilmaisut ”oppilas ymmärtää, oppilas osaa arvioida, vertailla jne.” ilmentävät ongelmallista suhtautumista yhteiskunnalliseen tiedonmuodostukseen ja yhteiskunnalliseen ajatteluun, joka oppikirjoissa esitetään usein staattisena ja muuttumattomana. Löfströmin mukaan yhteiskuntaopin opetuksen ongelmana on usein se, että yhteiskunnan ilmiöitä ei problematisoida tai analysoida eikä niiden erilaisia tulkintoja pohdita kriittisesti, vaan yhteiskunnan instituutiot esitetään ja kuvataan, kuten ne muodollisesti ja virallisesti ovat (Löfström 2001a, 12) ja myös yhteiskuntaopin oppikirjat tukevat tätä suhtautumistapaa. Vaikka yhteiskuntaopin

47 Vuoden 1994 Peruskoulun opetussuunnitelman perusteissa erityisenä yhteiskuntaopin tavoitteena on kirjattu, että ”oppilas saa sellaisia tietoja ja taitoja, joita hän tarvitsee toimiessaan ja asioidessaan yhteiskunnassa” (POPS 1994, 97).

tavoitteisiin on kirjattu jo vuosikymmenten ajan rakentava kriittisyys ja aktiivinen kansalaisuus (POPS 1985, 134; POPS 1994, 98; POPS 2004, 226), näyttää Löfströmistä siltä, että nuorten yhteiskunnallisen kritiikin ja yhteiskunnan nykyisen käytännön vaihtoehtoja ei ole pohdittu riittävästi (Löfström 2001a, 13). Löfströmin mukaan yhteiskuntatiedollisen opetuksen tehtävinä ja tavoitteina ovat aina olleet nuorten sosiaalistaminen yhteiskuntaan, perustiedon jakaminen kansalaiselle ja kunnan kansalaisen kasvattaminen (Löfström 2001a, 12). Virta toteaa kuten Löfström edellä, että opetussuunnitelmaan kirjaus, ”nykysuomalaista yhteiskuntaa tulisi tarkastella toiminnallisissa yhteyksissä” (POPS 1994, 97), viittaa ajatukseen siitä, että instituutioiden ja järjestelmän tarkastelun tulisi olla oppilaan näkökulmasta elävää ja konkreettista (Virta 2000, 58).

Yhteiskuntaopin tavoitteita ja niihin kytkeytyviä kriteereitä ei voi pitää opettajille itsestään selvinä oppiaineen opetuksen ja oppilasarvioinnin ohjenuorina. Yhteiskunnallisten aineiden seuranta-arvioinnissa opettajia pyydettiin nimeämään kolme sellaista yhteiskuntaopin tavoitetta tai kriteeriä, joita he ovat painottaneet opetuksessaan. Kriteerit ja tavoitteet on luokiteltu seuraavasti opettajien mainintojen perusteella: yhteiskunnallisen tiedon hankkiminen ja soveltaminen käytännössä, oppilaan kiinnostuksen herättäminen yhteiskunnalliseen vaikuttamiseen ja osallisuuteen sekä aktiivinen kansalaisuus ja vastuullinen toiminta lähiympäristössä ja yhteiskunnassa. Yhteiskuntaopissa 32 % opettajista jätti kokonaan vastaamatta yhteiskuntaopin tavoitteita ja kriteerejä koskevaan kysymykseen, he nimesivät yhteiskuntaopin sisältöjä tai vastauksissa oli selvästi ilmaistu, etteivät he tiedä perusteissa määriteltyjä tavoitteita tai kriteerejä. (Ouakrim-Soivio & Kuusela 2012, 35.)

3.3.2 Yhteiskuntaopin taidot ja tiedot

Historian ja yhteiskuntaopin opetussuunnitelmien perusteissa tiedot ovat olleet vuosikymmenten ajan keskeisellä sijalla aina vuoden 1994 perusteisiin asti. Tuoloin perustetekstissä tietojen rinnalle nostettiin myös taidot. Virran mukaan vuoden 1994 opetussuunnitelman tavoitteissa korostettiin yhteiskunnassa tarvittavien tietojen lisäksi taitoja sekä yhteiskuntaeettisten arvojen omaksumista. (Virta 2000, 58.) Perusteisiin kirjattiin historian ja yhteiskuntaopin tavoitteiksi, että oppilaan tuli saada sellaisia tietoja ja taitoja, joita hän tarvitsisi toimiessaan ja asioidessaan yhteiskunnassa (POPS 1994, 97).

Vuoden 2004 opetussuunnitelman perusteissa taitojen tulo tietojen rinnalle näkyy entistä yksityiskohtaisemmin määritellyissä oppiaineen tavoitteissa ja päättöarvioinnin kriteereissä. Opetussuunnitelman perusteissa taidot kirjattiin ns. johdantotekstiin siten, että vuosiluokkien 7–9 yhteiskuntaopin opetuksen tulee antaa perustiedot ja -taidot yhteiskunnan rakenteesta ja toiminnasta sekä kansalaisen vaikutusmahdollisuuksista. Oppiainekohtaisia taitoja opetussuunnitelman perusteissa on määritely muun muassa seuraavasti: taito oppia hankkimaan ja soveltamaan yhteiskuntaa

ja talouselämää käsittelevää informaatiota kriittisesti ja toimimaan aktiivisena vaikuttajana, taito hyödyntää julkisia palveluita, taito ymmärtää yrittäjyyden perusteet ja yrittäjyyden merkitys yhteiskunnassa, yhteiskunnallisten päätösten ja niiden vaikutusten ymmärtäminen tavallisten ihmisten arjessa, osallistumisen ja vaikuttamisen taidot sekä taidot toimia kuluttajana. (POPS 2004, 226.)

Löfströmin mukaan yhteiskuntaoppi on ollut jo lähtökohdiltaan yhteiskuntaan ja vallitseviin arvoihin sosiaalistava oppiaine. Yhteiskuntaopin taitotavoitteet ovat opetussuunnitelman perusteissa yleisemmin kuvailevia, koska yhteiskuntaopilla ei ole ollut yhtä tiedonalataustaa ja siihen perustuvaa tiedon luonnetta. (Löfström 2012a, 148.) Löfström huomauttaa, että oppilaan kyky tulkita kriittisesti esimerkiksi median välittämiä lähteitä tulisi käsittää siten, että oppilas osaa rajata kyseisiin lähteisiin liittyviä ongelmia. Yhteiskuntaopin taidoissa on hänen mukaansa pohjimmiltaan kyse siitä, että yhteiskuntaopin opetuksen tulisi ohjata oppilaita havaitsemaan mielipide-eroja ja päättämään, mitä seurauksia niillä on eri toimijoille. Taitoja mittaavien tehtävien tulisi puolestaan olla sellaisia, jotka ohjaavat oppilasta etsimään tietoja, vertailemaan ja erittelemään niitä sekä päättämään ja perustelemään esimerkiksi niistä johtuvia seurauksia. (Löfström 2012a, 151.)

Yhteiskuntaopin sisällöt on opetussuunnitelmien perusteissa määritelty eri tarkkuudella eri vuosikymmeninä siitä huolimatta, että varsinaisissa asiasisällöissä ei sinänsä ole tapahtunut suuria muutoksia. Syy siihen, miksi opetussuunnitelmissa on oppiaineiden sisältöjä määritelty eri tarkkuudella, löytyy Virran (2000, 58–59) ja Salmion (2004, 60–63) mukaan siitä, että opetussuunnitelman tehtävät on nähty eri aikoina joko amerikkalaistyyllisenä curriculumina, jossa korostuvat oppilaspainotteisuus ja opetuksen tavoitteiden määrittely tai saksalaismallisenä Lehrplanina, jossa keskeisenä ovat puolestaan opetuksen sisällöt. Salmion mukaan suomalainen opetussuunnitelma on sekoitus kumpaakin (Salmio 2004, 62). Virta (2000) toteaa, että eri vuosikymmeninä peruskoulun opetussuunnitelmien sisältöjen määrittelyn laajuus kuvaa sitä muutosta Suomessa, jolloin siirryttiin Lehrplan-tyyppisestä, opetuksen sisällöt tarkasti määrittelevästä opetussuunnitelmasta curriculum-tyyppiseen ajatteluun, jossa opetussuunnitelma nähdään jatkuvana prosessina (Virta 2000, 58).

Apajalahden (2004) mukaan eri vuosikymmeninä peruskoulun opetussuunnitelmien perustetekstien erot sisällöissä ja tavoitteissa näkyvät ennen kaikkea siinä, että vuoden 1985 opetussuunnitelman perusteissa oli määritelty vuosiluokittaiset tavoitteet ja keskeiset sisällöt valtakunnallisesti, mutta vuoden 1994 perusteissa vain ala- ja yläasteen päättyessä (Apajalahti 2004, 126). Suuri väljyys ja laaja tulkinnan mahdollisuus vuoden 1994 peruskoulun opetussuunnitelman perusteissa näyttää johtaneen siihen, että tavoitteen ja oppilaan arvioinnin yhteys hämärtyi.⁴⁸ Vuonna 2004 vahvistetuissa perusopetuksen opetussuunnitelman perusteissa määriteltiin

48 Tulkinta vuoden 1994 peruskoulun opetussuunnitelman perusteiden väljyydestä löytyy Opetushallituksen sivustolta: http://www.oph.fi/saadokset_ja_ohjeet/ohjeita_koulutuksen_jarjestamiseen/perusopetuksen_jarjestaminen/perusopetuksen_oppilaan_arviointi/oppilaan_arviointi_menneina_vuosikymmenina.

nivelkohdille oppiainekohtaiset kuvaukset hyvästä osaamisesta ja perusopetuksen päättövaiheeseen arviointikriteerit. Historian osalta tämä tarkoittaa hyvän osaamisen kuvausta 6. vuosiluokan ja päättöarvioinnin kriteerejä 8. vuosiluokan päättyessä ja yhteiskuntaopissa päättöarvioinnin kriteerejä 9. vuosiluokan päättyessä.

Perusopetuksen opetussuunnitelmien perusteissa yhteiskuntaopin oppiaineen sisällöt pysyivät pitkälti eri vuosikymmeninä samoina. Suomessa yhteiskuntaopissa opetetaan monenlaisia sisältöjä, kuten politiikkaa, taloutta, sosiaalipolitiikkaa ja oikeuskasvatusta (POPS 2004, 226–227). Löfström, Virta ja van den Berg (2010) arvelevat sisältöalueiden kirjon viestivän omalta osaltaan siitä, että oppiaineella ei ole vielä näkyvää ja vakiintunutta asemaa (Löfström, Virta & van den Berg 2010, 7).

Yhteiskuntaopin oppiaineen sisällöt määriteltiin yksityiskohtaisesti vuosien 1970 ja 1985 peruskoulun opetussuunnitelmien perusteissa. Lähtökohtana oli, että opetus alkoi oppilaille tutuista aiheista, kuten perhe, työ ja kunta, joista edettiin kohti abstraktimpia aihealueita, kuten vaalijärjestelmä, eduskunta ja turvallisuuspolitiikka (POPS 1985, 134). 1970-luvulta aina vuoteen 1994 asti yhteiskuntaopissa, kuten muissakin oppiaineissa, oppiaineiden sisällöt jäsenettiin erillisiksi saarekkeiksi, joissa yhtä aihealuetta tai saareketta käsiteltiin kahdesta kolmeen tuntia (Komiteanmietintö 1970: A5, 217–218). Opetussuunnitelmassa esitetystä saarekeajattelusta on tässä esimerkkinä vuoden 1970 yhteiskuntaopin opetussuunnitelmassa vaalit, johon kytkettiin vaalien toimittaminen, ehdokasasettelu, tulosten laskeminen, valtiolliset ja yhteiskunnalliset mielipiteet, poliittisen mielipiteen synty, aatesuunnat ja poliittiset puolueet. Kyseisessä saarekkeessa kansalaistaidon kanssa integroitaviksi aiheiksi nimettiin vaalituloksen laskeminen ja poliittisten puolueitten esittely (Komiteanmietintö 1970: A5, 227).

Sekä vuoden 1970 että vuoden 1985 yhteiskuntaopin opetussuunnitelmassa korostuvat opetettavat sisällöt ja opetettavat tiedot. Vuoden 1970 Komiteanmietinnössä pidettiin tärkeänä, että oppilaille muodostuisi tiedollinen kokonaiskuva yhteiskunnasta ja talouselämästä. Kansalaistaidon opetuksessa pyrittiin samaan aikaan kuin yhteiskuntaopissa käsittelemään samoja aihealueita, jotta tunneilla käytävien opetuskeskustelujen avulla oppilaat saisivat henkilökohtaisen kosketuksen yhteiskunnallisiin ja taloudellisiin asioihin ja he ymmärtäisivät käsiteltävinä olevien asioiden merkityksen. (Komiteanmietintö 1970: A5, 218.) Yksittäiset yhteiskuntaopin ja taloustiedon sisällöt oli luetteloitu tarpeellisten ja opetettavien tietojen listaksi. Vuoden 1985 perustetekstissä todettiin yhteiskuntaopin sisällöistä, että niihin on pyritty sisällyttämään asioita, jotka oppilaan iän ja elämäkokemuksen kannalta ovat hänelle läheisiä ja helppotajuisia (POPS 1985, 134, 143). Opettajaa puolestaan kehoitettiin havainnollistamaan, elävöittämään ja konkretisoimaan eri asiasisältöjä kiinnittämällä opetettava asia oppilaan kotiseutuun tai kotimaakuntaan (POPS 1985, 134). Vuoden 1994 perusteteksti poikkesi olennaisesti kahdesta aiemmasta opetussuunnitelmien perusteista myös sisältöjen kirjaamisen osalta, sillä se oli hyvin lyhyt ja perustetekstissä jätettiin paljon soveltamis- ja tulkitsemismahdollisuuksia

paikalliselle tasolle. Historian ja yhteiskuntaopin sisällöt määriteltiin niin, että siihen kuului yleistä historiaa, Suomen historiaa ja nykysuomalaista yhteiskuntaa (POPS 1994, 97). Peruskoulun opetussuunnitelman perusteissa kehoitettiin opetuksessa ottamaan alueelliset ja paikalliset erityispiirteet huomioon (POPS 1994, 97). Varsinainen yhteiskuntaopin sisältö määriteltiin yhdellä kappaleella siten, että ”Yhteiskuntaopissa tarkastellaan Suomen valtiollista ja yhteiskunnallista järjestelmää, kansalaisen asemaa yhteiskunnassa, Suomen kansantalouden peruspiirteitä ja talouselämän toimintaa sekä sosiaalietiikan keskeisiä kysymyksiä.” (POPS 1994, 98).

Virran mukaan vuoden 1994 perusteissa ei määritelty yksityiskohtaisesti historian tai yhteiskuntaopin opetussisältöjä, vaan siinä nimettiin lyhyesti opetettavat asiat. Virta esittää kantanaan, että nykysuomalaista yhteiskuntaa olisi tullut tarkastella toiminnallisissa yhteyksissä, eikä opetus olisi saanut rajoittua instituutioiden ja järjestelmän toteamiseen, kuten jo vuoden 1985 opetussuunnitelman perusteissa oli tehty. (Virta 2000, 58.) Löfströmin mukaan peruskoulun opetussuunnitelmien perusteissa yhteiskuntaopin oppimääräkuvauksissa on aiempina vuosikymmeninä esitetty laajasti aineen tiedollis- taidollisia tavoitteita, jossa ”tieto” oli esitetty luettelomaisesti sen mukaan, miten asioiden oletettiin olevan. Tietoon sosiaalistaminen tapahtui esittämällä tieto persoonattomana ja kuvaamalla se ikään kuin muualta tulleena. Jos vielä yhteiskuntaopin opiskelussa analysoinnin ja problematisoinnin harjoittelu ja soveltaminen jäivät vähäiseksi, se sosiaalisti Löfströmin mukaan suoraan demokratian arvoihin, eikä opiskelulla pyritty aktiiviseen ja omakohtaiseen tiedon muodostukseen. (Löfström 2001a, 13.) Kuten Löfström (2001a, 13), myös Virta tuo esiin, että yhteiskuntaopin opetukseen liittyy haaste sen mukaan, kuvataanko opetuksessa yhteiskuntaa instituutiokeskeisesti vai yksilökeskeisesti, hahmottuuko oppilaalle kuva järjestelmästä, joka on staattinen vai dynaaminen ja muuttuva sekä painottuuko opetuksessa virallinen kuva päätöksenteosta vai voidaanko opetuksessa valottaa kulissien takaista toimintaa. Tärkeää Virran mukaan yhteiskuntaopin opetuksessa on, missä määrin se voi sisältää yhteiskuntakritiikkiä. (Virta 2000, 81–82.) Suutarisen mukaan yhteiskuntaopin sisällölliset haasteet liittyvät muun muassa siihen, että ne ovat hyvin käsitteellisiä ja etäällä nuorten arkitodellisuudesta (Suutarinen 2000, 50–51). Samaa mieltä Suutarisen kanssa on myös Virta, joka toteaa, että esimerkiksi valtiolliset instituutiot ja niihin liittyvät sisällöt ovat kaukana oppilaan kokemuspöyrästä, jolloin oppiminen jää helposti pintapuoliseksi ja ulkoa opituksi. Oppilaalla voi olla myös vaikeuksia oppia yhteiskuntaopin sisältöjä, koska yhteiskuntaopissa käytettävät käsitteet ovat abstrakteja ja sisällöt monisäikeisiä. (Virta 2000, 81.)

Virran mukaan yhteiskuntaopin opetuksessa on keskeistä sen välittämä ihmiskuva, jota ei kuitenkaan hänen mielestään ole otettu riittävästi huomioon yhteiskuntaopin opetusta ja sisältöjä koskevassa keskustelussa. Yhteiskuntaopin opetussisällöistä osa on vahvasti arvosidonnaisia ja siihen saattaa Virran mielestä liittyä suuria periaatteellisia ongelmia. Demokratia, kansalaisen oikeudet, ihmisoikeudet ja yhteiskunnalliseen vaikuttamiseen liittyvät sisällöt ovat haastavia opettaa arvovapaasti, koska niin

opettajat kuin oppimateriaalit tuovat omat tulkintansa aiheiden käsittelyyn. (Virta 2000, 82.) Opetushallitus selvitti vuonna 2012 demokratiakasvatusta koskevan selvityksen osana, koetaanko kouluissa joidenkin yhteiskunnallisten aihealueiden opettaminen vaikeaksi. Vastanneista peruskouluista 12 % koki vaikeaksi opettaa joitain osa-alueita, ja vastaajat nimesivät kokemiaan hankaluuksia seuraavasti: ”Tiedotusvälineitä seurataan vähän, haetaan pinnallista tietoa.” ”Tämän päivän oppilaat kokevat hyvinvointivaltion itsestäänselvyytenä, he kokevat että jotkut hoitavat asiat puolestamme. Yksilökeskeisyys vallalla.” ”Vanhemmat ovat välillä yliherkkiä aiheelle.” ”Joskus on vaikeaa opettaa neutraalisti yhteiskunnallisia asioita.” (Opetushallitus 2012, 43.) Osa demokratiaselvityksen vastauksista tuki Suutarisen ja Virran edellä esittämiä kommentteja, kuten ”Miten ne saa lapsille tarpeeksi yksinkertaisesti selitettyä kun tuntuu, ettei itsekään aina ymmärrä!” tai ”Miten saadaan ”kuivista” faktatietoa sisältävistä perustiedoista opiskelijoita kiinnostavia, jos heillä ei ole omakohtaista kosketusta ko. asiaan.” (Opetushallitus 2012, 45.) Erityisesti alakoulut kommentoivat sitä, että politiikka on heidän oppilailleen liian vaikeaa ja raskasta. Toisaalta peruskoulujen vastauksissa korostettiin samaan aikaan ”yhteiskunnallista heräämistä osana päätöksenteon ja demokratian opiskelua”. (Opetushallitus 2012, 47.)

Vuoden 2004 opetussuunnitelman perusteissa palattiin vuoden 1985 perusteiden muotoon sikäli, että opetussuunnitelman perusteista tehtiin yksityiskohtaisemmat. Uutta oli se, että historialle ja yhteiskuntaopille kirjoitettiin, kuten erillisille oppiaineille kuuluikin, erilliset perustetekstit. Yhteiskuntaopin keskeisimmiksi sisällöiksi määriteltiin ensin yleisellä tasolla suomalainen yhteiskunta ja talouselämä sekä Euroopan unioni. Tämän jälkeen määritellyt oppiaineen sisältöalueet muistuttavat paljolti vuoden 1985 perusteissa mainittuja sisältöjä, joskin vuoden 2004 perusteteksti on selvästi yleisluontoisempi ja aihealueet määritelty siten, että perusteet antoivat riittävät raamit koulukohtaisten opetussuunnitelmien laatimiseen.

Yhteiskuntaopin osaamiseen kuuluu vuoden 2004 Perusopetuksen opetussuunnitelmien perusteiden mukaan se, miten yksilö toimii erilaisissa yhteisöissä, kuten perheessä, kotikunnassa, kansalaisena ja EU:ssa. Yhteiskuntaopin osaamiseen kuuluvat myös hyvinvoinnin eri ulottuvuuksien ja tasa-arvon sekä kestävä kehityksen merkityksen ymmärtäminen hyvinvoinnin edistämiseksi. Vaikuttamistaitoihin kuuluvat paitsi kansalaisen mahdollisuudet vaikuttaa, myös demokratia, vaalit ja äänestäminen, politiikan ja hallinnon toimijat kunnallisella, valtakunnallisella ja EU-tasolla sekä media ja yhteiskunnallinen vaikuttaminen. Yhteiskuntaopin sisältöjen moninaisuus ilmenee muun muassa kansalaisen turvallisuutta käsittelevässä sisältöalueessa, johon on sisällytetty kuuluvaksi oikeusjärjestelmän osia, kuten yksilön oikeuksia ja velvollisuuksia, oikeudellista vastuuta, liikenneturvallisuutta ja turvallisuuspolitiikkaa. Taloutta käsitellään yksityisen taloudenpidon, työnannon, yrittäjyyden, kuluttamisen, kotitalouksien, ulkomaankaupan, globaalitalouden ja talouspolitiikan näkökulmista. (POPS 2004, 226–227.) Keskeisiksi yhteiskuntaopin osaamisen alueiksi, joita tulee myös arvioida, määritellään kuuluvaksi yhteiskunnallisen

tiedon hankkiminen ja käyttäminen sekä yhteiskunnallisen tiedon ymmärtäminen (POPS 2004, 227). Voimassa olevan opetussuunnitelman perusteisiin on kirjattu ajatus siitä, että oppilaan tulisi kiinnostua osallistumisesta, kuluttamisesta ja yrittäjyydestä sekä ymmärtää tekojensa laillisuuteen liittyvät seuraukset (POPS 2004, 226–227).

Löfströmin, Virran ja van den Bergin mukaan (2010, 7) vuoden 2004 opetussuunnitelman perusteisiin on kirjattu kapea näkökulma sosiaaliseen lukutaitoon. Edellä jo useaan kertaan mainittu perinteinen instituutiokeskeisyys johtaa heidän mukaansa tietojen opettamiseen, eikä analyyttiseen ja kriittiseen valmiuteen, jota oppilailla tulisi olla. Myös oppikirjat vahvistavat toteavaa piirrettä opetuksessa. (Virta 2006, 97–98, 100, 104; Löfström, Virta & van den Berg 2010, 8.) Löfström (2012) esittää, että yhteiskuntaopin opetuksessa nuorille tulisi toisaalta antaa valmiudet toimia yhteiskunnassa ja sopeutua siihen sellaisena kuin se on tässä ja nyt, toisaalta nuorille tulisi antaa uusia työvälineitä nykyisestä poikkeavien taloudellisten ratkaisujen ja päätösten tekemiseen. Suomalaisessa yhteiskuntaopin opetuksessa on Löfströmin mukaan pidetty tärkeänä, että nuoret oppivat ensin tietyn perustietopakettin yhteiskunnan instituutioista, ja vasta sen jälkeen he ovat voineet harjoitella yhteiskunnallisen ajattelun, analyysin, synteessin ja argumentoinnin taitoja. (Löfström 2012a, 147.) Yhteiskuntaopin seuranta-arvioinnin opettajakyselyn tulokset tukevat ainakin osittain Löfströmin edellä esittämää oletusta siitä, että opetuksessa painottuu yhä tiedolliste tavoitteet, taidollisten tavoitteiden jäädessä taka-alalle. Kun yhteiskuntaopin opettajia pyydettiin nimeämään neljä sisältöaluetta, joita he olivat viime vuosina opetuksessaan painottaneet, koskivat ne taloustietoa ja sen sisältöjä, yksilön oikeuksia ja velvollisuuksia, Suomen poliittista järjestelmää ja vallankäytön keskeisimpiä instituutioita sekä osallisuutta ja vaikuttamista. (Ouakrim-Soivio & Kuusela 2012, 35.)

Löfströmin, Virran ja van den Bergin mukaan yhteiskuntaoppia on hyödynnetty, kun nuoret on haluttu sosiaalista valtion arvoihin ja valtiollisiin instituutioihin. Yhteiskuntaopista tuli väline ”kylmien faktojen” opettamiselle eikä siinä ole ollut sijaa kriittiselle analysoinnille tai vallitsevan rakenteen kyseenalaistamiselle. Syy Löfströmin, Virran ja van den Bergin mukaan löytyy siitä, että instituutioiden opettaminen ei anna mahdollisuutta tulkita poliittisia prosesseja. Opettajien voi olla vaikeaa luoda johdonmukaisia ja analyyttisiä näkökulmia poliittisista ja taloudellisista aihealueista. Tämä taas saattaa johtua siitä, että yhteiskuntaoppia opettavat opettajat ovat yleensä historian opettajia, joiden omat tiedot ja taidot opettaa yhteiskuntatietoita ovat vallinneen koulutustaustan vuoksi heikompia kuin historian opettaminen. (Löfström, Virta & van den Berg 2010, 8.)

Kansainväliset tutkimukset osoittavat, että suomalaiset nuoret osaavat hyvin yhteiskuntaopin sisältöihin liittyviä faktoja (Suutarinen ym. 2001, 4, 6–7; Suoninen, Kupari, Törmäkangas 2010, 7–8, 144). Myös yhteiskunnallisten aineiden oppimistulokset osoittavat, että oppilaat osaavat hyvin opetussuunnitelman perusteissa

yhteiskuntaopille asetettuja tietoja ja taitoja. Oppilaiden keskimääräinen ratkaisusuus yhteiskuntaopin seuranta-arvioinnissa oli 64 % (Ouakrim-Soivio & Kuusela 2012, 60). Nuorilla näyttää kuitenkin olevan vaikeuksia politiikan ja talouden toisiinsa kytkemisessä (Suutarinen ym. 2001, 41; van den Berg 2010, 9–10). Suutarisen mukaan (2002) politiikan ja talouden kytkeminen toisiinsa edellyttää yhteiskunnan käsitteellistämistä vaikka se olisi monimutkaista ja yhteiskunnan rakenteet ja voimasuhteet menisivätkin lomittain (Suutarinen 2002, 11, 19–20). Suutarisen näkemys saa tukea yhteiskuntaopin seuranta-arvioinnin tuloksista, sillä oppilaat osasivat heikoimmin niitä tehtäviä, joissa mitattiin yhteiskunnallisen päätöksenteon ja taloudellisten ratkaisujen vaihtoehtojen ja niiden seurausten pohdintaan liittyviä taitoja (Ouakrim-Soivio & Kuusela 2012, 60).

3.3.3 Oppilaan arviointi yhteiskuntaopissa

Opetushallituksen mukaan oppilaan arviointia koskevat määräykset ja menettelyt ovat vaihdelleet suuresti peruskoulujärjestelmän aikana, samoin arviointikäytännöt ja todistusarvosanojen muodostumisen periaatteet. Vuonna 2004 pantiin toimeen järjestelmä, joka yhtenäisti arviointia ja antoi entistä paremmat edellytykset valtakunnallisesti vertailukelpoisten päättöarvosanojen antamiselle.⁴⁹

Menettelyjen moninaisuus ilmenee yhteiskuntaopin oppilaan arvioinnissa kahdella tavalla. Ensinnäkin yhteiskuntaoppia arvioitiin osana historian oppiainetta vuoteen 1999 asti, jolloin Opetushallitus julkaisi Perusopetuksen päättöarvioinnin kriteerit. Kyseisessä julkaisussa historia ja yhteiskuntaoppi muodostivat omat, itsenäiset oppiaineensa, joille oli määritelty oppiainekohtaiset arviointikriteerit. Näin siitä huolimatta, että ne silloisen tuntijaon ja opetussuunnitelman perusteiden mukaan olivat kuitenkin yksi oppiaine. Toinen käytäntöjen moninaisuudesta kertova seikka on se tapa ja laajuus, jolla oppilaan arviointia opetussuunnitelmien perusteissa ohjattiin. Yhteiskuntaopin oppilasarvioinnin ohjeistus oli vuoden 1970 Komiteanmietinnön perustetekstissä varsin niukka. Perusteet oli kirjoitettu kaikille oppiaineille erikseen, samoin ohjeet oppimistulosten arvioimiseksi. (Komiteanmietintö 1970: A5, 231). Komiteanmietinnön yleisessä osassa määriteltiin koulutyön tulosten arviointi ja siinä erityisesti oppimistulosten arvioinnin tavoitteet, koulutyön tulosten arvioinnissa käytettävät mittausvälineet sekä yleisohjeet esimerkiksi kokeiden laatimiseksi, koetulosten hyödyntämiseksi ja arvosanojen antamiseksi (Komiteanmietintö 1970: A4, 158–170). Opettajia ohjeistettiin myös kokeiden arvioimisessa ja niiden hyödyntämisessä osana oppilaan arviointia. Sekä historian että yhteiskuntaopin opetuksessa pidetään 2000-luvulla tärkeänä, että oppilas osaa muodostaa omia käsityksiään ja mielipiteitään sekä perustella niitä (POPS 2004, 220, 222, 226). Nykyiseen ajattelutapaan verrattuna vuoden 1970 komiteanmietinnön historian ja

49 Arvio löytyy Opetushallituksen sivustolta: http://www.oph.fi/saadokset_ja_ohjeet/ohjeita_koulutuksen_jarjestamiseen/perusopetuksen_jarjestaminen/perusopetuksen_oppilaan_arviointi/oppilaan_arviointi_menneina_vuosikymmenina

yhteiskuntaopin arviointiohjeen näkökulma oli erilainen, sillä tehtävien asettelusta oppilaan tuli tehdä johtopäätös siitä, mitä osa-alueita opettaja piti tärkeänä (Komiteamietintö 1970: A5, 231).

Vuoden 1985 opetussuunnitelman perusteissa ei ollut oppiainekohtaisia arviointiohjeita historialle ja yhteiskuntaopille. Sen sijaan perusteissa jatkettiin varsin yksityiskohtaisten yleisohjeitten antamista oppilasarvioinnin suorittamiseksi (POPS 1985, 29–32). Yleisohjeet kirjoitettiin siten, että ne olivat sovellettavissa kaikkiin peruskoulussa opetettaviin oppiaineisiin sekä niin, että ohjeet sisälsivät sekä arviointiin että arvosteluun liittyvät ohjeet. Arvioinnilla eli evaluaatiolla ymmärrettiin kaikkea sitä toimintaa, jonka tarkoituksena oli verrata koulun toiminnan tuloksia yhteiskunnan koulujärjestelmälle asettamiin tavoitteisiin. Arvostelulla tarkoitettiin yksittäisen oppilaan tai oppilasryhmän suoritusten vertaamista johonkin ennalta asetettuun kriteeriin ja vertailutuloksen ilmoittamista arvosanana. (POPS 1985, 29.) Oppilaan arvioinnissa tuli ottaa huomioon oppilaitten erilaiset lähtökohdat. Arvioinnin tuli sekä antaa palautetta oppilaan edistymisestä että kannustaa opiskeluun (POPS 1985, 29). Huolimatta siitä, että vuoden 1985 opetussuunnitelman perusteet määrittivät valtakunnalliset ja kuntakohtaiset tavoitteet ja arviointiperusteet, Peruskoulun opetussuunnitelman perusteissa tuotiin esiin myös ns. tavoitepohjainen arviointi, jossa arvosanojen tuli ilmaista sitä, miten oppilas on onnistunut suoriutumaan suhteessa omiin tavoitteisiinsa, sekä kannustaa oppilasta käyttämään omia oppimisedellytyksiään mahdollisimman tehokkaasti.⁵⁰ Perusteissa todettiin, että ”oli hyväksyttävää, että oppilaat saavuttavat tavoitteet eri tavoin ja myös hitaasti edistyvälle oppilaalle tuli antaa mahdollisuus muidenkin kuin heikkojen arvosanojen saamiseen”. (POPS 1985, 30.)

Vuoden 1994 Peruskoulun opetussuunnitelman perusteiden väljyys ja yleisluonteisuus johtivat siihen, että jo kaksi vuotta opetussuunnitelman perusteiden julkaisemisen jälkeen perustettiin työryhmä laatimaan kriteerejä oppilaan arvioinnin tueksi. Vuoden 1999 yhteiskuntaopin päättöarvioinnin kriteerien mukaan oppilaalla tuli päättövaiheessa olla käsitys yhteiskunnallisen tiedon luonteesta, hänen tuli osata hankkia ja käyttää sekä ymmärtää yhteiskunnallista tietoa. Käsitykseen yhteiskunnallisen tiedon luonteesta kuului muun muassa, että oppilas tiesi yhteiskuntatiedon koostuvan eri yhteiskuntatieteistä ja käsittelevän ihmisen toimintaa ja vuorovaikutusta yksilönä, ryhmissä ja järjestyneissä instituutioissa. Oppilaan tuli ymmärtää, että ihmisen toiminnan yhteiskuntatieteellinen selittäminen on historiallis-kulttuurista ja tulkinnallista. Oppilaalta odotettiin tietojen lisäksi erilaisten taitojen hallintaa, sillä hänen tuli päättövaiheessa osata hankkia ja soveltaa yhteiskuntaa koskevaa tietoa, tulkita tilastoja ja kaavioita, pystyä tarkastelemaan yhteiskunnallista tietoa kriittisesti, pystyä perustelemaan käsityksiään yhteiskunnallisista asioista, osata käyttää julkisia palveluita sekä käyttää yhteiskunnan vaikutuskanavia. (Perusopetuksen päättöarvioinnin kriteerit 1999, 99.)

50 Tavoitepohjaista arviointia on käsitelty jo aiemmin tässä tutkimuksessa luvussa 2.

Peruskoulun opetussuunnitelmien perusteissa (Komiteanmietintö 1970: A4, Komiteanmietintö 1970: A5; POPS 1985; POPS 1994) tavoitteiden väljään ja yleisluonteeseen kirjaamisen liittyy olennaisesti se hankaluus, että tavoitteiden toteutumista on ollut vaikea seurata ja arvioida. Vuosien 1970, 1985 ja 1994 perusteteksteissä ei ole myöskään kiinnitetty huomiota siihen, että tavoitteiden ja oppilasarviointin välillä olisi riittävä yhteys. Oppilaan arvioinnin tulisi kytkeytyä niihin tavoitteisiin, jotka opetussuunnitelmassa on ennalta asetettu ja määritelty (Atjonen 2005, 144; 2007, 76), ja asetettuja tavoitteita tulisi myös arvioida. Historiassa ja yhteiskuntaopissa näyttää siltä, että peruskoulun opetussuunnitelmien perusteissa arviointiohjeet liittyvät ennen kaikkea sisältöjen opettamisen ja oppimisen arviointiin, mutta eivät asetettujen tavoitteiden toteutumisen arviointiin (Komiteanmietintö 1970: A4, 167–168; POPS 1985, 29–31; POPS 1994, 25–29).

Vuoden 1999 perusopetuksen päättöarvioinnin kriteereissä taitojen mukaantulo näkyi kohdassa, jossa yhteiskunnallisen tiedon hankkimisen rinnalle nousi tiedon käyttäminen ja soveltaminen arjessa. Taitojen ja arvojen tuominen osaksi opetussuunnitelman tavoitteita tarkoitti Virran mukaan arvioinnin kehittämistä kohti jatkuvaa oppimisprosessin arviointia, jolloin kyse on siitä, että arviointi toimii oppilaan tukena ennemmin kuin siitä, että arvioinnilla pyritään kontrolloimaan oppilaan omaksumia tietoja. Keskeistä hänen mielestään on myös se, ovatko arvioitavat tiedot ja taidot yksilön ja yhteiskunnan kannalta mielekkäitä. (Virta 1999, 19; 2000, 80.) Se, miten opetussuunnitelman perusteissa määritellyt taidot kirjattiin koulu- tai kuntakohtaisiin opetussuunnitelmiin sekä koulukohtaisiin arviointiohjeisiin, jäi opetuksenjärjestäjän päätettäväksi.

Päättöarvioinnin kriteerien (1999) rakenne siirtyi samanlaisena vuoden 2004 opetussuunnitelmien perustetekstiin. Perusopetuksen opetussuunnitelman perusteisiin (2004) kirjattiin oppiaineittain tavoitteiden ja sisältöjen lisäksi myös päättöarvioinnin kriteerit arvosanalle 8. Kriteereistä tuli velvoittavia, eivätkä ne olleet enää suosituksia. Vuoden 2004 Perusopetuksen opetussuunnitelman perusteissa yhteiskuntaopin oppiaineelle asetettu tehtävä on ohjata oppilasta kasvamaan yhteiskunnan aktiiviseksi ja vastuulliseksi toimijaksi, antaa perustiedot ja -taidot yhteiskunnan rakenteesta ja toiminnasta sekä kansalaisen vaikutusmahdollisuuksista. Opetuksen yleisten ja yksityiskohtaisten tavoitteiden, jotka on käsitelty luvuissa 3.3.1 ja 3.3.2, tarkoituksena on tukea oppilaan kasvua suvaitseväiseksi ja demokraattiseksi kansalaiseksi ja antaa hänelle kokemuksia yhteiskunnallisesta toimimisesta ja demokraattisesta vaikuttamisesta. (POPS 2004, 226.) Yhteiskuntaopin oppiaineen keskeiset sisällöt määriteltiin aihealueittain siten, että ne käsittelevät suomalaista yhteiskuntaa ja talouselämää sekä Euroopan unionia ja siinä yksilöä yhteisön jäsenenä, yksilön hyvinvointia, vaikuttamista ja päätöksentekoa, kansalaisen turvallisuutta, taloudenpitoa, kansantalotta ja talouspolitiikkaa (POPS 2004, 226–227).

Koska yhteiskuntaopin oppimäärä opetetaan toistaiseksi yhden lukuvuoden aikana, sille ei ole määritelty nivelvaiheen kohtaan hyvän osaamisen kriteerejä. Päätösarvioinnin kriteerit arvosanalle 8 voidaan jakaa kahteen osa-alueeseen: yhteiskunnallisen tiedon hankkimiseen ja käyttämiseen ja yhteiskunnallisen tiedon ymmärtämiseen. (POPS 2004, 227.) Yhteiskunnallisen tiedon hankkimisen ja käyttämisen kriteerit päätösarvosanalle 8 kuvataan siten, että oppilas kykenee tulkitsemaan kriittisesti median välittämiä tietoja, tilastoja ja graafisia esityksiä, hän pystyy perustelemaan käsityksiään yhteiskunnallisista asioista ja osaa vertailla yhteiskunnallisen päätöksenteon ja taloudellisten ratkaisujen eri vaihtoehtoja ja niiden seurauksia (POPS 2004, 227). Yhteiskunnallisen tiedon ymmärtäminen kuvataan puolestaan niin, että oppilas ymmärtää, että yhteiskunnallisessa päätöksenteossa ja taloudellisissa ratkaisuissa on olemassa useita vaihtoehtoja ja hän ymmärtää yhteiskunnallisen ja taloudellisen toiminnan eettisiä kysymyksiä (POPS 2004, 227).

3.3.4 Perusteiden antama tuki yhteiskuntaopin arvioinnissa ja arvosanan antamisessa

Suomalainen desentralisoitu koulutusjärjestelmä on siirtänyt sekä päätös- että toimivaltaa koulutuksenjärjestäjille ja kouluille. Oppilaan arvioinnissa se tarkoittaa, että käytännössä opettajat arvioivat itse oppilaansa Opetushallituksen antamia Perusopetuksen opetussuunnitelman perusteita (2004) sekä kunta- tai koulukohtaisia opetussuunnitelmia noudattaen. Koska perusopetuslaki ja -asetus antavat lainsäädännöllisen viitekehyksen, mutta Opetushallituksen vahvistama Perusopetuksen opetussuunnitelman perusteet (2004) muodostavat ne normit, joita koulutuksen järjestäjät, koulut ja opettajat tulkitsevat, keskitytään tässä perusopetuksen opetussuunnitelman perusteteksteihin. Kaikissa oppilaan arviointia ohjaavissa säädösteksteissä on ilmaistu vahvasti periaate oppilaan arvioinnin yhdenmukaisuudesta, joka käytännössä tarkoittaa esimerkiksi arvosanojen valtakunnallista vertailtavuutta jatko-opintoihin hakeuduttaessa.

Aiempiä opetussuunnitelmien perusteita (Komiteanmietintö 1970: A4; Komiteanmietintö 1970: A5; POPS 1985; POPS 1994) ja niiden oppilaan arviointia käsitteleviä osioita on arvosteltu kerroksellisuudesta sekä siitä, että ne olivat epäselviä ja että niissä vaikutti taustalla arvioinnin kirjava traditio. Siitä huolimatta, että Opetushallitus verkkosivuillaan⁵¹ viestittää voimassa olevan opetussuunnitelman perusteiden selkeydestä aiempiin perusteisiin verrattuna, on perustetekstissä puutteita, jotka voivat johtaa joko virhetulkintoihin, vanhojen käytänteitten soveltamiseen tai siihen, että voimassa olevia perusteita ei osata, tiedetä tai huomata noudattaa oppilaan arvioinnissa. Opetussuunnitelman perustetekstin tulisi olla selkeää ja yksiselitteistä, sillä opetussuunnitelman perusteet ohjaavat suomalaisessa koulutusjärjestelmässä

51 Kyseinen kannanotto löytyy Opetushallituksen sivustolta: http://www.oph.fi/saadokset_ja_ohjeet/ohjeita_koulutuksen_jarjestamiseen/perusopetuksen_jarjestaminen/perusopetuksen_oppilaan_arviointi/oppilaan_arviointi_menneina_vuosikymmenina

opetuksen lisäksi oppilaan arviointia, päättöarvosanaa ja kansallisia seuranta-arviointeja. Jo aiemmin tässä luvussa 3 on käsitelty niitä ristiriitaisuuksia, jotka liittyvät Opetushallituksen asettaman perusopetuksen päättöarvioinnin kriteereitä valmistelleen työryhmän muistioihin (1996; 1997) ja Päättöarvioinnin kriteerit (1999)-julkaisun ohjeisiin, vaikka kaikkia ohjeita ja periaatteita ei ole enää luettavissa Päättöarvioinnin kriteereissä (1999).

Oppilaan arviointi perustuu aina oppimiskäsitykseen (Jakku-Sihvonen & Heinonen 2001, 78–80; Atjonen 2005, 144), mutta Perusopetuksen opetussuunnitelman perusteissa (2004) oppimiskäsitystä käsittelevässä luvussa (POPS 2004, 16) ei nimetä kognitiivis-konstruktivistista tai sosio-konstruktivistista oppimiskäsitystä, eikä siihen viitata oppilaan arvioinnissa, vaikka sille tyypilliset oppilaan omien tietorakenteiden kehittyminen, yksilöllisyyden huomioonottaminen, kokonaisuuksien hahmottaminen tai oppilaskeskeisyys on sisäankirjoitettu oppilaan arviointiosuuteen (vrt. POPS 2004, 260). Käsitteen avaaminen ja oppilaan arviointiosuuden kiinnittäminen oppimiskäsitykseen ja sen teoreettiseen viitekehykseen auttaisi arviointityötä tekevää opettajaa hahmottamaan laajemmin, mistä perusteissa olevat normit ovat syntyneet ja miksi oppilaan arviointia suoritetaan suhteessa hänen omien tietojensa ja taitojensa kehittymiseen, eikä yksittäisen oppilaan arviointia enää tulisi suhteuttaa muun ryhmän tai luokan tasoon.

Nykyisiä oppilaan arvioinnin normeja vaivaa kerrostuneisuus ja päällekkäisyys, jota kuvaa hyvin Opetushallituksen verkkosivuille kerätty artikkelikokoelma oppilaan arvioinnista, johon tässä tutkimuksessa on jo viitattu aiemmin. Opetushallituksen sivustolla olevien artikkeleiden lisäksi oppilaan arviointia on käsitelty ainakin Opetushallituksen tiedotteessa arviointikriteerien käyttöön ottamisesta (2007)⁵² sekä Opetushallituksen ylläpitämässä Edu.fi-portaalissa,⁵³ jossa on oppiainekohtaista materiaalia perusopetuksen oppilaan arvioinnin tueksi. Täydentävät verkkotekstit tai tiedotteet eivät viesti Perusopetuksen opetussuunnitelman perusteissa (2004) määritellyn oppilaan arviointiosuuden tai päättöarvioinnin selkeydestä tai siitä, että normitekstit olisivat luettavissa ja tulkittavissa yksiselitteisesti. Jotta oppilaan arviointi ja arvosanat voisivat perustua yhdenmukaisiin käytänteisiin, tulisi perustetekstin olla tulkittavissa mahdollisimman yksiselitteisesti.

Virta kuvasi jo vuonna 1999 virallista arviointikieltä idealistiseksi (Virta 1999, 20). Idealismin lisäksi arviointikieli saattaa sisältää ristiriitaisuuksia. Esimerkki arviointikielen ristiriitaisuudesta löytyy Perusopetuksen opetussuunnitelman perusteiden (2004) ensimmäisestä oppilaan arviointia käsittelevästä kappaleesta, jossa arvioinnin tehtäväksi määritellään toisaalta ohjata ja kannustaa opiskelua, toisaalta auttaa oppilasta muodostamaan realistinen kuva oppimisestaan (POPS 2004, 260).

52 Opetushallituksen tiedote on sivustolla: http://www.oph.fi/download/30179_Perusopetuksen_paattoarviointi_uudistuu.pdf

53 Perusopetuksen oppilaan arviointiin löytyy oppiaineittain ohjeita Opetushallituksen ylläpitäältä sivustolta: <http://edu.fi/perusopetus>

Todellisuudessa kyseiset lauseet sopivat hyvin opinnoissaan menestyvien oppilaiden oppimisen ja kasvun kuvaamiseen, mutta heikosti menestyvät oppilaat, jotka tarvitsivat ohjausta ja kannustamista, eivät sitä saa ainakaan silloin, kun heille kuvataan realistisesti heidän oppimistaan.

Perustetekstissä todetaan, että opintojen aikaisen arvioinnin tulee perustua monipuoliseen näyttöön (POPS 2004, 260). Monipuolinen näyttö on jo käsitteenä tulokinnallinen. Lankinen ja Lahtinen tulkitsevat sitä, mahdollisesti vanhan säädöspohjan kannalta niin, että arvosanan tulee perustua muuhunkin kuin kokeissa annettuun näyttöön (Lankinen & Lahtinen 2009, 237). Atjonen (2005, 145; 2007, 65) ja Virta (1999, 58–77) lähestyvät monipuolisen arvioinnin käsitettä kognitiivisen oppimiskäsityksen kautta ja erilaisia autenttisen arvioinnin menetelmiä hyödyntäen, joita ovat esimerkiksi jatkuva havainnointi, portfolioit, oppimispäiväkirjat, aineisto- tai näyttökokeet tai arviointikeskustelut.

Eniten arviointia käsittelevässä osuudessa huomio kiinnittyy päättöarviointia ohjaavien kriteereiden yleisluonteisuuteen. Arvosanaa 8 kuvataan siten, että ”oppilas saa arvosanan 8, mikäli hän osoittaa keskimäärin oppiaineen kriteereiden edellyttämää tasoa” (POPS 2004, 264), jota seuraavassa lauseessa täydennetään toteamalla, että ”joidenkin kriteerien saavuttamatta jättämisen voi kompensoida muiden kriteerien ylittäminen”. Vastaavasti arvosana 5 kuvataan siten, että ”oppilas on saavuttanut perusopetuksessa vaadittavat tiedot ja taidot välttävästi, mikäli hän pystyy osoittamaan jossain määrin kriteerien edellyttämää osaamista” (POPS 2004, 264). Kun päättöarvosana on toistaiseksi tärkein määrittäjä oppilaille heidän hakeutuessaan jatko-opintoihin, antavat nykyiset kriteerit varsin vähän tukea opettajan arviointityöhön ja yhdenvertaisen arvioinnin perustaksi.

Jos Suomessa käytettäviä päättöarvioinnin kriteerien kuvauksia verrataan esimerkiksi Ruotsissa vuonna 2011 käyttöön otettuihin opetus suunnitelman perusteiden mukaisiin yksityiskohtaisiin tasokuvauksiin, ero on huomattava. Esimerkkinä oppiaineen taitotasokuvausten suhteen voi toimia yhteiskuntaoppi. Ruotsissa käytössä olevat arvosanat A, C ja E on kiinnitetty yksityiskohtaisesti oppiaineelle asetettuihin tavoitteisiin, tietoihin ja taitoihin. Yhteiskuntaopille, kuten muillekin perusopetuksessa opetettaville aineille, on laadittu käsikirja, jossa on perusteellisten kuvausten lisäksi esimerkkejä opettajan arviointityön tueksi. (Skolverket 2012, 28–50.) Opettajan arviointityön tueksi on laadittu oppiainekohtaiset sivustot, joista esimerkiksi yhteiskuntaopin arviointisivustolla on paljon oppilaan arvioinnin tueksi laadittua materiaalia käytännön sovelluksineen ja esimerkkeineen, kuten jo aiemmin tämän tutkimuksen luvun 2 yhteydessä on esitetty.⁵⁴

54 Ruotsin kouluviraston sivustolla on erikseen oppiaineittain oppilaan arviointiin tukimateriaalia. Esimerkiksi yhteiskuntaopin oppilaan arvioinnin sivusto löytyy osoitteesta: <http://www.skolverket.se/prov-och-bedomning/ovrigt-bedomningsstod/2.6011/7-9/2.2975/2.6121/samhallskunskap-1.106465>

Suomessa yhteiskunnallisten aineiden seuranta-arviointiin osallistuneiden koulujen opettajilta kysyttiin muun muassa sitä, miten hyvin he saavat yhteiskuntaopin ja historian oppilaan arviointiin tukea Perusopetuksen opetussuunnitelman perusteista (2004). Kyselyyn vastanneista opettajista noin puolet oli sitä mieltä, että he saavat riittävästi ohjausta opetussuunnitelman perusteista yhteiskuntaopin oppilasarviointiin, mutta noin neljäsosa ilmoitti, etteivät he saa riittävästi tarvitsemaansa tukea (Ouakrim-Soivio & Kuusela 2012, 35). Kun opettajia pyydettiin nimeämään neljä omasta mielestään tärkeintä yhteiskuntaopin tavoitetta tai kriteeriä, 32 % vastanneista opettajista ei joko vastannut kysymykseen lainkaan tai he nimesivät sisältöjä tai vastauksissa oli selvästi ilmaistu, etteivät he tiedä perusteissa määriteltyjä yhteiskuntaopin tavoitteita tai kriteerejä (Ouakrim-Soivio & Kuusela 2012, 35).

Löfström totesi vuoden 1999 yhteiskuntaopin päättöarvioinnin kriteereistä, että ne olivat liian yleiset ja ylimalkaisesti määritelty (Löfström 2001a, 15) ja saman huomion hän on tehnyt myös vuoden 2004 yhteiskuntaopin kriteerien osalta (Löfström 2012a, 148–149). Virran mukaan vuoden 1999 päättöarvioinnin kriteerit olivat ylätasoa kuvauksia, jotka eivät kovinkaan selkeästi tarkentaneet, mitä oppilaan olisi todella osattava suorittaa. Kriteerit jättivät hänen mielestään liian paljon tulkinnanvaraa siitä, mitä kriteerit merkitsivät ja miten niiden avulla ajateltiin toteuttaa oppilaiden vertailua niin, että samaan aikaan oppilaiden oikeusturvaa edistetään. (Virta 2000, 79.)

Sama ylimalkaisuus ja tulkinnanvaraisuus liittyvät vuoden 2004 yhteiskuntaopin päättöarviointiin. Taulukkoon 3 on koottu yhteiskuntaopin tavoitteet ja oikealle puolelle se, mitä päättöarvioinnin kriteerit arvosanalle 8 sisältävät.

TAULUKKO 3. Yhteiskuntaopin oppiaineen tavoitteet ja päättöarvioinnin kriteerit arvosanalle 8.

Yhteiskuntaopin tavoitteet	Päättöarvioinnin kriteerit arvosanalle 8
<p>Oppilas</p> <ul style="list-style-type: none"> • saa käsityksen yhteiskunnallisen tiedon luonteesta • oppii hankkimaan ja soveltamaan yhteiskuntaa ja talouselämää käsittelevää informaatiota kriittisesti ja toimimaan aktiivisena vaikuttajana • oppii tuntemaan julkiset palvelut • saa valmiuksia työnteon kunnioittamiseen • oppii yrittäjyyden perusteet ja ymmärtää yrittäjyyden merkityksen yhteiskunnan hyvinvoinnin tekijänä • oppii ymmärtämään yhteiskunnallisten päätösten vaikutuksia kansalaisten elämään • kiinnostuu yhteiskunnallisesta osallistumisesta ja vaikuttamisesta • oppii tarkastelemaan ja kehittämään osaamistaan vastuullisena kuluttajana ja yhteiskunnallisena toimijana ja tuntee tekojensa oikeudelliset seuraamukset. 	<p>Yhteiskunnallisen tiedon hankkiminen ja käyttäminen</p> <p>Oppilas</p> <ul style="list-style-type: none"> • kykenee tulkitsemaan kriittisesti median välittämiä tietoja, tilastoja ja graafisia esityksiä • pystyy perustelevaan käsityksiään yhteiskunnallisista asioista ja • osaa vertailla yhteiskunnallisen päätöksenteon ja taloudellisten ratkaisujen eri vaihtoehtoja ja niiden seurauksia. <p>Yhteiskunnallisen tiedon ymmärtäminen</p> <p>Oppilas</p> <ul style="list-style-type: none"> • ymmärtää, että yhteiskunnallisessa päätöksenteossa ja taloudellisissa ratkaisuissa on olemassa useita vaihtoehtoja ja • ymmärtää yhteiskunnallisen ja taloudellisen toiminnan eettisiä kysymyksiä.

Tarkasteltaessa yhteiskuntaopin tavoitteita ja kriteereitä rinnakkain, tulee varsin näkyväksi, että osa tavoitteista ei liity kriteereihin. Osa tavoitteista on tiedollisia, kuten tavoite, jonka mukaan ”oppilas saa käsityksen yhteiskunnallisen tiedon luonteesta”, osa on taidollisia, kuten tavoite, ”oppilas oppii tuntemaan julkiset palvelut”. Osa tavoitteista liittyy asenteisiin, kuten esimerkiksi ”oppilas saa valmiuksia työntöön kunnioittamiseen”. Atjosen mukaan arviointi ei ”asetu oikeisiin uomiinsa”, eikä se kohdennu oikein, jos kriteerit eivät mittaa asetettuja tavoitteita (Atjonen 2005, 144). Vertailemalla taulukkoon 3 koottuja yhteiskuntaopin tavoitteita ja oppiaineelle määriteltyjä kriteereitä, on havaittavissa selvästi, että yhteiskuntaopin oppiaineen tavoitteet ja kriteerit keskustelevat kyllä samasta aiheesta, yhteiskuntaopista, mutta ne eivät keskustelee keskenään (vrt. myös Ouakrim-Soivio 2013, 149–150).

Jotta yhteiskuntaopin oppiaineen tavoitteiden ja arviointikriteereiden keskinäisen yhteyden väljyys konkretisoituu, antaa historian tavoitteet ja kriteerit hyvän, vertailuvan esimerkin siitä, miten oppiaineet poikkeavat tavoitteiden asettelussa ja arviointikriteereiden määrittelyssä toisistaan. Historian oppiaineen tavoitteeksi on asetettu, että oppilas oppii ymmärtämään historian tietojen olevan historioitsijoiden tulkintoja, jotka saattavat muuttua lähteiden tai tarkastelutapojen myötä (POPS 2004, 220). Kyseistä tavoitetta vastaa päättöarvioinnin kriteeri, jonka mukaan oppilas pystyy lukemaan erilaisia lähteitä ja tulkitsemaan niitä (POPS 2004, 223). Tavoitetta, jonka mukaan oppilas oppii ymmärtämään erilaisia tapoja jakaa historia aikakausiin (POPS 2004, 220), vastaa päättöarvioinnin kriteeri, jossa oppilas kykenee sijoittamaan tapahtumat ajallisiin yhteyksiin ja niiden avulla aikajärjestykseen (POPS 2004, 224). Tällä tavalla historian oppiaineessa määritellyt tavoitteet kiinnitetään niitä vastaaviin päättöarvioinnin kriteereihin.

Peruskoulun opetussuunnitelman perusteita (1994) ja Perusopetuksen päättöarvioinnin kriteereitä (1999) on arvosteltu niiden yleisluonteisuuden vuoksi. Sama ylimalkaisuus näyttää siirtyneen vuoden 2004 perusteisiin ainakin yhteiskuntaopin oppiaineen tavoitteiden osalta. Kun tarkastellaan esimerkiksi sitä, mitä oppilaan tulisi oppia yhteiskuntaopissa – ”oppilas saa käsityksen yhteiskunnallisen tiedon luonteesta, hän oppii tuntemaan julkiset palvelut, oppilas ymmärtää yrittäjyyden merkityksen yhteiskunnan hyvinvoinnin tekijänä, hän kiinnostuu yhteiskunnallisesta osallistumisesta ja vaikuttamisesta tai hän tuntee tekojensa oikeudelliset seuraamukset” (POPS 2004, 226) – tavoitteet ovat varsin väljiä ilmauksia. Se, että oppilas tuntee julkiset palvelut, ei välttämättä tarkoita sitä, että tarvitessaan hän osaisi käyttää niitä. Se, että oppilas kiinnostuu yhteiskunnallisesta osallistumisesta, ei myöskään vielä kerro siitä, että hän toimisi aktiivisena kansalaisena. Lisäksi tavoitteissa asetetun mukaista toimintaa, yrittäjyyden merkityksen ymmärtämistä, tekojen oikeudellisten seurausten ymmärtämistä tai julkisten palvelujen tuntemista ei ole yksinkertaista mitata millään mittarilla tai antaa arvosanaa suhteessa näin väljästi asetettuihin tavoitteisiin.

Mitä tulee yhteiskuntaopin päättöarvosanan 8 kriteereiden (POPS 2004, 227) ja Perusopetuksen opetussuunnitelman perusteissa (2004) arvosanan 8 tieto- ja taitotason kriteereiden (POPS 2004, 264) yhteensovittamiseen, se ei juuri anna kiinnekohtia opettajan arviointityöhön. Arvioinnin kohteina ovat yhteiskunnallisen tiedon hankkiminen, jossa oppilaan tulee kyetä tulkitsemaan kriittisesti median välittämiä tietoja, tilastoja ja graafisia esityksiä ja pystyä perustelemaan käsityksiään yhteiskunnallisista asioista, sekä yhteiskunnallisen tiedon ymmärtäminen, jossa oppilaan tulee ymmärtää, että yhteiskunnallisessa päätöksenteossa ja taloudellisissa ratkaisuisissa on olemassa useita vaihtoehtoja ja ymmärtää yhteiskunnallisen ja taloudellisen toiminnan eettisiä kysymyksiä (POPS 2004, 227). Kun yhteiskuntaopin päättöarvioinnin kriteereitä sovelletaan arvosanan 8 ohjeisiin, ”oppilas saa arvosanan 8, mikäli hän osoittaa keskimäärin oppiaineen kriteerin edellyttämää tasoa” tai arvosanan 5 kriteereihin ”oppilas saa, mikäli hän pystyy osoittamaan jossain määrin kriteerien edellyttämää tasoa” (POPS 2004, 264), opettajalle jää varsin suuri väljyys ja tulkinnan mahdollisuus itse määrittää, mitä hän arvosanaa antaessaan painottaa.

Perusopetuksen opetussuunnitelman perusteiden (2004) mukaan tietojen lisäksi tulisi arvioida myös taitoja (POPS 2004, 260). Kun katsotaan yhteiskuntaopin oppiaineen tavoitteita, sisältöjä ja päättöarvioinnin kriteerejä, keskeiset sisällöt on määriteltä selkeästi ja yksiselitteisesti. Löfström on huomauttanut siitä, että suomalainen yhteiskuntaoppi on perinteisesti painottunut antamaan ensisijaisesti tietoja yhteiskunnan instituutioista, ja vasta sen jälkeen oppilaan ajatellaan voivan harjoitella yhteiskunnallisen ajattelun, analyysin, synteesin ja argumentoinnin taitoja (Löfström 2012a, 147–148). Kun katsotaan, miten opetussuunnitelman perusteissa tiedot ja taidot määritellään, on varsin ymmärrettävää, että opetuksen painopiste on yhä sisällöissä.

Yhteiskuntaopin oppiaineen perustetekstissä taidot ovat jääneet sisältöjen varjoon ja taitotavoitteiden rinnalla tiedot korostuvat yhä. Perusteissa ei varsinaisesti ole erotettu tietojen rinnalle taitoja (POPS 2004, 226–227). Opetussuunnitelmatekstin taitotavoitteita voi Löfströmin mukaan tulkita siten, että osa tavoitteista on tiedollisia, osa taidollisia. Näitä tiedollisia ja taidollisia tavoitteita ovat ne, että oppilas oppii hankkimaan ja soveltamaan yhteiskuntaa ja talouselämää käsittelevää informaatiota kriittisesti ja toimimaan aktiivisena vaikuttajana, tai oppilas oppii hankkimaan ja soveltamaan yhteiskuntaa ja talouselämää käsittelevää informaatiota kriittisesti ja toimimaan aktiivisena vaikuttajana. Asennetavoitteita ovat esimerkiksi ne, että oppilas kiinnostuu yhteiskunnallisesta osallistumisesta ja vaikuttamisesta tai oppilas saa valmiuksia työnteon kunnioittamiseen. (Löfström 2012a, 148–149.) Jos kuitenkin halutaan ohjata opetusta taitojen opettamiseen, tulisi taitojen olla opetussuunnitelman perusteissa kirjoitettu selkeästi ja siten, että ne olisi otettu myös huomioon yhteiskuntaopin päättöarvioinnin kriteereissä.

Perusopetuksen opetussuunnitelman perusteet (2004) antavat oppilaan arvioinnin ja erityisesti päättöarvioinnin suhteen arvioijalle suuria vapauksia tulkita arvosanan antamista määrittäviä kohtia varsin väljästi. Yhteiskuntaopin päättöarvioinnin suurimmat ongelmat liittyvät niiden yleisluonteisuuden lisäksi siihen, että oppiaineen tavoitteita ei ole kiinnitetty päättöarviointikriteereihin. Hyvän osaamisen kuvaus arvosanalle 8 ohjeistaa opettajaa arvioimaan eri asioita kuin oppiaineen tavoitteet ovat opetukselle asettaneet. Taitojen puuttuminen tai niiden tulkitseminen oppiaineen tavoitteiden pohjalta johtaa väistämättä siihen, että opettajat painottavat opetuksessaan yhä tietoja, jotka ovat varsin perinteisesti, mutta selkeästi keskeisinä sisältöinä perustetekstiin kirjattu. Tätä taustaa vasten ei ole odottamatonta, että oppilaat osaavat hyvin yhteiskuntaopin faktoja, kuten jo aiemmin luvussa 3.3.2 todettiin, mutta tietojen soveltaminen on hankalampaa, kuten yhteiskunnallisten aineiden seuranta-arviointi osoitti (Ouakrim-Soivio & Kuusela 2012, 60–61).

Kun tarkastellaan yhteiskuntaopin opetuksen painopisteitä, opetustuntimääriä sekä oppiaineen tiedollisia ja taidollisia tavoitteita eri vuosikymmeninä, ovat yhteiskuntaopin opetuksen peruslinjat säilyneet suhteellisen samanlaisina siitä huolimatta, onko yhteiskuntaoppia opetettu historian oppiaineen yhteydessä vai itsenäisenä oppiaineena. Opetussuunnitelmien perusteissa erot ovat muodostuneet pikemminkin siitä, kuinka laajasti tai suppeasti esimerkiksi yhteiskuntaopin opetukselle asetettuja tavoitteita, tietoja ja taitoja on perusteteksteissä määritelty ja miten yksityiskohtaisesti opettajien arviointityötä on ohjeistettu.

Tässä kolmannessa luvussa on esitetty perusteluita sille, miksi yhteiskuntaopin päättöarvioinnin kriteerit eivät välttämättä tue opettajan arviointityötä eivätkä päättöarvioinnin kriteerit näytä toimivan yhdenmukaisesti oppilaiden osaamista mitaavana ”mittatikkuna”. Näitä syitä ovat muun muassa arviointia koskevien normien kerrostuneisuus ja se, että edellisiä arviointinormeja ja ohjeita ei ole purettu perusteellisesti uusien astuessa voimaan, päättöarviointikriteerien ylimalkaisuus ja se, etteivät päättöarvioinnin kriteerit kiinnity yhteiskuntaopin oppiaineelle asetettuihin tavoitteisiin niin kuin niiden kuuluisi. Saattaa myös olla, että osittain edellä mainituista syistä, myös yhteiskuntaopissa suuri osa kouluista ankkuroi arvosanan 8 enemmän oman koulunsa keskiarvoitukseen kuin Opetushallituksen kriteereihin, kuten Kuusela (2006, 81) on tutkimuksessaan osoittanut.

4 VALIDIUS-KÄSITE JA SEN TEORIATAUSTAA

Koska tämän tutkimuksen tehtävänä on tarkastella arvosanojen validiutta, määrittää validiuden käsitettä ja sen teoriataustaa tässä luvussa tarkemmin.⁵⁵ Validiudella eli tutkimuksen pätevyydellä ja luotettavuudella tarkoitetaan joko tutkimusmenetelmän tai mittarin kykyä selvittää sitä, mitä sillä on ajateltu selvittää. Mittarit ja menetelmät eivät nimittäin aina vastaa sitä, mitä tutkija kuvittelee tutkivansa (Hirsjärvi, Remes & Sajavaara 2009, 231). Validiudella voidaan tarkoittaa myös mittaustuloksista tehtävien päätelmien sopivuutta, mielekkyyttä ja käyttökelpoisuutta (Nummenmaa ym. 1997, 203).

Mittarilla tai mittaamisella tarkoitetaan joko koko testipatteristoa, kuten mittavälineitä, joilla tuotetaan tietoa tutkittavasta kohteesta, tai yksittäistä testiä. Vaikka yksinkertaisimmillaan mittari voi olla yksi ainoa kysymys, koostuu se yleensä useammasta osiosta. Mittarin käyttö perustuu pyrkimykseen havainnoida ilmiötä mahdollisimman objektiivisesti. Testien haasteena on, että kehitetyllä mittarilla ei aina tavoiteta tutkittavan ilmiön todellista luonnetta. Testillä saadaan selville vain se, mitä kyseisellä mittarilla on mahdollista selvittää. (Zeller 1998, 322; Metsämuuronen 2006, 57.) Mikäli käytettävissä on valmis mittari, jonka validius ja reliabelius on tutkittu, sitä on syytä käyttää, koska mittari on yleensä testattu suurella määrällä ihmisiä ja mittarin luotettavuus on tutkittu ja kuvattu (Kane 1990; Metsämuuronen 2006, 54, 115).

Tutkimusasetelman tai tutkimuksen validius voidaan jakaa ulkoiseen (*external*) ja sisäiseen (*internal*). Ulkoisen validiuden tarkastelulla pyritään karsimaan tutkimusasetelman luotettavuutta uhkaavat tekijät (Jakku-Sihvonen 2001, 113; Metsämuuronen 2006, 55; vrt. Virta 1999, 78) ja siinä tarkastellaan, onko tutkimus yleistettävissä ja jos on, niin mihin ryhmiin. Sisäinen validius tarkoittaa tutkimuksen omaa luotettavuutta, jossa pohditaan muun muassa sitä, ovatko käsitteet oikeita, onko teoria valittu oikein, onko mittari oikein muodostettu, mitkä tekijät vaikuttavat mittaustilanteessa luotettavuuteen alentavasti sekä mitataanko mittarilla sitä, mitä on sillä on tarkoitus mitata (Metsämuuronen 2006, 55). Sisäisen validiuden osoittamisessa on tärkeää, että arvioinnin kohteena oleva todellisuus on konstruoitu mahdollisimman oikein (Zeller 1988, 332; Jakku-Sihvonen 2001, 113). Esimerkiksi Greer (1969, 160) piti havaintoja ja niistä tehtäviä formulointeja vaikeimpina ja keskeisimpänä kaikista tieteellisistä yrityksistä. Myös se, millaisella mittarilla asiaa mitataan, liittyy sisäiseen validiuteen. Esimerkkinä voisi olla yhteiskuntaopin esseevastauksen pisteittäminen mahdollisimman objektiivisesti ja yhteismitallisesti. Tuottamistehtävät ovat käytännössä mahdottomia tulkita täysin yksiselitteisesti, ja siksi niiden pisteittäminen täysin samalla tavalla on mahdotonta. Oleellista on saattaa pisteitysohjeet mahdollisimman yksiselitteisiksi esimerkeiksi jakamalla

55 Tämän tutkimuksen oma validiustarkastelu käydään luvussa 5, jossa tarkastellaan niitä systemaattisia tekijöitä, jotka saattavat alentaa tämän tutkimuksen tai tutkimuksessa tehtyjen mittausten luotettavuutta.

arvioitava teksti osatekijöihin (Metsämuuronen 2006, 118) ja merkitsemällä pis-teitysohjeisiin osatekijöiden tuottamat pisteet.

Modernin validiusteorian mukaan ei ole olemassa eri validiuden lajeja, vaan vali-diutta tulee tarkastella esimerkiksi testin, kokeen tai tutkimusasetelman asettaman lähtökohdan mukaan sopivimmasta näkökulmasta (Messick 1989b, 5).⁵⁶ Viimeis-ten vuosikymmenten aikana validiuden määrittelyssä on tapahtunut olennainen muutos myös siinä, että validiuden tarkastelun kohteena ovat saadut tulokset, ei mittari, sekä tulosten tulkinta ja tulosten hyödyntäminen siihen tarkoitukseen, mihin mittari on alun perin laadittu (Cronbach 1971, 447; Messick 1994, 774; ks. myös Rantanen 2003, 19).

4.1 Erilaisia näkökulmia validiuden tarkasteluun

Tutkimuksen, testin tai kokeen sisäistä validiutta on perinteisesti tarkasteltu kol-mesta näkökulmasta sen sisällön, kriteereiden tai konstruktion kautta (Messick 1994, 741). Koska tässä tutkimuksessa validiutta tarkastellaan lisäksi konvergenssin näkökulmasta, määritellään tässä yhteydessä myös konvergenssivalidiuden käsite.

Sisällön validiustarkastelussa (*content validity*) on olennaista, että sillä katetaan kohdealueena oleva keskeinen sisältö (Virta 1999, 80). Sisällön validiutta tutkit-taessa varmistetaan, että mittarissa tai ylipäänsä tutkimuksessa käytetyt käsitteet ovat teorian mukaisia ja oikein operationalisoituja ja että ne kattavat riittävän laa-jasti kyseisen ilmiön (Metsämuuronen 2006, 116). Sisällön validiudella voidaan ilmaista esimerkiksi sitä, onko seuranta-arviointiin valitut tehtävät edustava otos Perusopetuksen opetussuunnitelman perusteissa (2004) määrittelyistä oppiaineen tavoitteista, tiedoista ja taidoista (vrt. Rantanen 2003, 19). Validiuden tarkastelua sisällön näkökulmasta on arvosteltu erityisesti siksi, että sitä pidetään subjektiivivi-sena ja puolueellisena. Sisällön validiustarkastelu perustuu testin tai koetehtävien edustavuuden ja niiden relevanssin arviointiin. Milloin sisällön validiustarkastelu tehdään tehtäväryhmän toimesta, ryhmällä on luonnollisesti taipumus vahvistaa itse esittämäänsä tulkintaa. (Kane 2006, 19.) Messickin mukaan sisällön validiustar-kasteluun perustuvat havainnot eivät pidä sisällään testitulosten tulkintaa ja niiden validiustarkastelua, ja siksi sitä ei voida käyttää täysipainoisesti validiustarkastelussa. Messick korostaakin sisällön validiutta hyvin rajallisena näkökulmana testin, kokeen tai tutkimustehtävän validiustarkasteluun. (Messick 1989a, 17.)

56 Messickin teoretisoima moderni validiuskäsitys on alkanut saada sijaa vasta 2000-luvulla. Esimerkiksi Zeller (1988, 323–329) nimeää eri validiustyyppisiä, joista hän nostaa keskeisimmiksi sisältövalidiuden, kriteeriva-lidiuden ja käsitevalidiuden. Myös Rantanen (2003, 19) käsittelee tutkimuksessaan eri validiuden lajeja, jos-kin hän toteaa, että Messickin moderni validiusteoria lähtee siitä, että käsitevalidius ”sisältää kaikki validi-u-den lajit” (Rantanen 2003, 20).

Kriteerivalidius (*criterion validity*) on sisäisen validiuden laji, jossa verrataan mittarilla saatua arvoa johonkin sellaiseen arvoon, joka toimii validiuden kriteerinä. Kriteerivalidius on perinteisesti jaettu ennustevalidiuteen (*predictive validity*), josta esimerkkinä on opintomenestystä ennustavat valintakokeet, ja samanaikaisvalidiuteen (*current validity*), josta esimerkkinä voisi olla uuden tehtäväsarjan vertaaminen jo koeteltuun standardoituun testiin (Rantanen 2003, 19).

Kriteerivalidius toimii tarkastelunäkökulmana hyvin silloin, kun on jo olemassa olevat, uskottavat kriteerit, joihin tulosta verrataan (Kane 2006, 18). Tällaisena kriteerinä voisi toimia esimerkiksi toisella mittarilla samanaikaisesti saatu arvo, jolloin puhutaan yhtäaikais- tai samanaikaisvalidiudesta. Kriteerivalidisuuden mittana käytetään yleensä korrelaatiokerrointa (Metsämuuronen 2006, 65). Mittarin validiuskertoimia voi olla useita riippuen siitä, mitä pidetään kriteerinä (Jakku-Sihvonen 2001, 125; Metsämuuronen 2006, 121). Ongelmallisena on juuri keskenään samankaltaisten kriteereiden luominen. Ebelin mukaan jos kyseenalaistetaan yksi kriteeri, voidaan samalla kyseenalaistaa kaikki muutkin kriteerit (Ebel 1961, 642).

Mikäli ilmiö todella noudattaa jotain teoriaa tai mallia, sen tulisi saada tukea aineistosta. Tätä sisäisen validiuden tarkastelutapaa kutsutaan konstruktio- tai käsitevalidiudeksi (*construct validity*). Käsitevalidiudella kuvataan sitä, miten hyvin tutkimuksen kohde on onnistuttu käsitteellistämään ja missä määrin mittaus kytkeytyy teoreettiseen viitekehukseen. (Jakku-Sihvonen 2001, 125; Rantanen 2003, 19; Kane 2006, 18–19; Metsämuuronen 2006, 118.) Käsitevalidiudessa yhdistyvät sekä testi- tai tutkimustulosten tulkinta että tulosten hyödyntäminen, sillä käsitevalidiudessa pohditaan, vastaako mittarin tai tutkimusmenetelmän käsite sen teoreettista viitekehystä (Metsämuuronen 2009, 29). Käsitevalidius kuvataan usein myös prosessina, jossa aineistoon perustuva havainto tai näyttö tukee päättelyä siitä, miten johdonmukaisesti empiiriset havainnot vastaavat testitulosta. Havaintoja arvioidaan pääasiassa suhteessa siihen, missä määrin ne ovat johdonmukaisessa yhteydessä muihin mittauksiin. (Messick 1980, 1015.) Käsitevalidiuden teoreettinen viitekehys yhdistää empiirisen testauksen ja teorian pohjalta laaditut rationaaliset hypoteesit testituloksiin ja niiden olennaisiin yhteyksiin (Messick 1995, 741). Nykykäsitys testin, kokeen tai tutkimusasetelman validiudesta perustuu pitkälti käsitevalidiuden periaatteisiin ja siksi käsitevalidiutta käsitellään perusteellisemmin luvussa 4.2.

Konvergenssivalidiudella (*convergent validity*) tarkoitetaan sitä, konvergoivatko mittaustulokset keskenään eli osoittavatko mittaustulokset samaan suuntaan (Kane 2006, 36–37; ks. Ketokivi 2009, 67–68). Konvergenssin mittana käytetään korrelaatiota. Jos korrelaatio on vahva, mittaustulokset konvergoituvat samaan suuntaan. (Kane 2006, 36–37.)

4.2 Modernin validiusteorian muotoutuminen

Validiuden käsite ja teoria ovat muotoutuneet viimeisten kuuden vuosikymmenen kuluessa. Alun perin validiudella tarkoitettiin mittaustarkkuutta. Validiutta tarkasteltiin suhteessa ”oikeisiin” arvoihin tai kriteereihin, joihin saatuja mittaustuloksia verrattiin. Validiudella tarkoitettiin siis sitä, miten saadut tulokset ennustivat ennalta asetettujen kriteereiden mukaisia tuloksia. (Kane 2001, 319.) Alussa validiuden osoittamisessa korostui pikemminkin argumentointi kuin laskeminen tai todistaminen, jolloin keskeistä validiuden tarkastelussa oli se, millä perusteella mittausta voitiin pitää asiallisena ja luotettavana (Metsämuuronen 2006, 122). 1950- ja 1970-lukujen välisenä aikana validiutta pyrittiin käyttämään eräänlaisena työkalupakkina, jossa eri validiusnäkökulmat palvelivat joko kasvatustieteellisiä tai psykologisia testejä. Sisältövalidiudella pyrittiin perustelevaan erilaisten koulusaa-vutustestien validiutta, kun taas käsitevalidius istui paremmin teorialähtöisten ja tulkinnallisten tulosten validiustarkasteluun.

Standards for Educational and Psychological Testing määritteli 1950-luvun alussa validiutta siten, että siitä saatava tieto indikoi kokeen tai testin käyttäjälle, missä määrin koe tai testi kykeni saavuttamaan sille asetetut tavoitteet (Brennan 2006, 1). Nummenmaan ja muiden mukaan periaatteessa oli mahdollista, että tutkija ei pystynyt perustelevaan sitä, mitä hänen mittarinsa mittasi ja miksi niillä oli yhteys muihin muuttujiin (Nummenmaa ym. 1997, 206). Kokeen tai testin validiutta tarkasteltiin ensi alkuun neljästä eri näkökulmasta: sen sisällön, sen sisältämän väitteen, samanaikaisuuden ja sen rakenteen mukaan. Varsin pian validiutta alettiin kuitenkin tarkastella kolmesta näkökulmasta sisällön, kriteereiden tai konstruktin eli käsitteiden mukaan, mitä Guion (1980) alkoi kutsua validiuden kolminaisuus-opiksi. (Brennan 2006, 2.) 1950-luvun alussa validiustarkastelun näkökulma liittyi myös testien tai tutkimuksen toimivuuteen. Cureton (1951) määritteli testin tai kokeen validiuden siten, että validiudessa oli olennaista, että mittari tai testi mittaa sitä, mitä sen oli ajateltu mittaavan (Cureton 1951, 621). Tarkemmassa määrittelyssä Cureton tarkasteli validiutta sen mukaan, oliko kyse testin tai tutkimuksen merkityksestä (*relevance*) vai reliabiliudesta (*reliability*) eli satunnaisten virheiden puuttumisesta. Curetonin määritteli validiutta myös tosiasiallisten tulosten (*actual test scores*) korrelaationa suhteessa ”oikeiden kriteereiden” määrittämiin tuloksiin (*”true” criterion scores*). (Cureton 1951, 622–623.)

1960-luvulle asti validiudella kuvattiin sitä, mitataanko sitä, mitä on tarkoituskin mitata. Mittaajalla oletettiin olevan tiedossa jokin kriteeri mittaukselle, johon validiutta suhteutettiin ja verrattiin. Validius liitettiin usein kriteerivalidiuden käsitteeseen, jossa saatua koetulosta verrattiin johonkin toiseen validina kriteereinä pidettyyn arvoon. Kriteereihin perustuvan validiustarkastelun haasteena oli ja on yhä hyvin määriteltujen, havaintoihin perustuvien ja validien kriteereiden määritteleminen. Silloin, kun kriteereiden asettaminen on ollut vaikeaa, eräänä ratkaisuna on käytetty kriteereiden vertaamista ennalta haluttuun suoritukseen, josta

on muodostunut eräänlainen ”tavoitellun suorituksen mitta”. Esitetyn kaltainen asiantuntija-arvioon perustuva validiustarkastelu on kuitenkin herättänyt arvostelua siksi, että se on menettelytavoiltaan subjektiivista ja validius on perustunut pitkälti arviointia suorittavien henkilöiden ennakoasenteisiin. (Kane 2001, 320; Kane 2006, 18.)

American Psychological Association (1966) kuvasi validiuden sisällöllisiä, kriteereihin perustuvia tai rakenteeseen liittyviä näkökulmaeroja niin, että sisällön validius kuvaa sitä, miten hyvin testien tai kokeiden perusteella on tehty sisältöön perustuvat johtopäätökset. Kriteerivalidius vertaa testituloksia ulkoiseen muuttujaan tai muuttujiin ja käsitevalidius luo kiinnepohdan sellaisille päätelmille, joita ei voi tehdä suoraan. Kane kuvaa näitä näkökulmaeroja siten, että kriteerivalidiudessa saatua tulosta tarkastellaan asetettujen kriteerien mukaan, sisältövalidiudessa suorituksen mittaustulosten vaikuttavat asiantuntija-arviot siitä, miten tuotosta tulisi mitata, ja käsitevalidiudessa validiustarkastelu tapahtuu aina laajan analyysin tuloksena. (Kane 2001, 323.)

1970-luvulla American Psychological Association sai valmiiksi standardit, joissa korostettiin validiuden olevan mittaustulosten perusteella tehtävien päätelmien asianmukaisuutta. Oleellisena alettiin pitää sitä, että mittaustulosten tulkintaa pysytettiin perustelemaan. (Kane 2001, 324.) Messickin mukaan käsitevalidius loi rationaalisen perustan testien käyttämiselle sen ennustettavuuden ja merkityksen vuoksi (Messick 1980, 1012). 1970-luvulla oli validius-käsitteen määrittelyssä päästy yhteisymmärrykseen siitä, ettei ole olemassa erilaisia validiuden lajeja, vaan validiutta tulee tarkastella erilaisista näkökulmista, sillä kyse on ennen kaikkea havaintojen tai näytön perusteella tehtävien johtopäätösten tulkinnoista ja näiden validiudesta (Messick 1989b, 5). Vuosikymmen myöhemmin 1980-luvulla validius-käsite oli laajentunut tarkoittamaan myös testitulosten perusteella tehtyjen päätelmien tarkoituksenmukaisuutta, mielekkyyttä ja hyödyllisyyttä (Messick 1994, 774; Brennan 2006, 2; vrt. myös Rantanen 2003, 19).

Validiuden käsite on muotoutunut kuuden viimeisen vuosikymmen ajan validiuden lajeista aina moderniin validiusteoriaan, joka perustuu Messickin teoretisoimaan ajatukseen käsitevalidiuden ykseydestä. Samalla validiuden tarkastelussa on päädytty monimutkaisiin ja pitkälle kehittyneisiin ajatusmallien tarkastelua edellyttäviin formulointeihin, kuten esimerkiksi seuraavassa luvussa esiteltävään Kanen validointiprosessiin.

4.2.1 Käsitevaledius validiustarkastelun näkökulmana

Empiirisessä tutkimuksessa validiutta ovat perinteisesti pohtineet käyttäytymistieteilijät. Validius-käsitteen ja validiuden teoreettisen viitekehyksen muotoutumiseen vaikuttaneita tutkimuksia, kirjoituksia ja keskusteluja ovat ylläpitäneet muun muassa Cronbach, Messick ja Kane. Heidän näkemyksensä validiudesta perustuvat pitkälti käsitevalidiuden määrittelylle ja tämän validiustarkastelun näkökulman teoretisoimiselle.

Kanen mukaan validius-käsitteen muotoutuminen tapahtui vaiheittain. Ensimmäinen vaihe keskittyi testien tai arviointien yleisen uskottavuuden, esitettyjen tulkintojen ja niiden hyödyntämisen arviointiin. Validiustarkastelussa nousi keskeiseksi testituloksen tulkinta, ei testi sinänsä (vrt. Cronbach 1971, 447; Messick 1989a, 13). Curetonin (1951), Cronbachin ja Meehlin (1955), Cronbachin (1971) sekä Messickin (1975) teoretisoinnit vaikuttivat olennaisesti siihen, että testitulosten tulkinta alettiin ottaa huomioon validiustarkastelussa. Toisessa vaiheessa luotiin yleiset käsitevalidiutta koskevat periaatteet ja edelleen voimassa olevat validiuden määritelmät, joita kehitti ennen kaikkea Messick. Kolmannessa vaiheessa testi- tai tutkimustulosten tulkinnasta aiheutuva toiminta ja sen seuraukset liitettiin kuuluvaksi validointiin, jota teoretisoi erityisesti Kane. Vielä nykyisin kiistellään siitä, mikä merkitys testitulosten seurauksilla ja tulosten hyödyntämisellä on validiustarkastelussa. Esimerkiksi Pophamin (1997) mukaan seurausten tarkastelu ei saisi liittyä validiustarkasteluun millään tavalla. Nykyisin yleisenä validiustarkastelun lähtökohtana pidetään sitä, että ne tahot, jotka esittävät testiä hyödynnettäväksi tiettyä tarkoitusta varten, myös perustelevan sitä, miksi testi soveltuu juuri kyseiseen tarkoitukseen. Viimeinen vaihe modernin validiusteorian muotoutumisessa käsittää validiuden sisällyttämisen tai yhdistämisen erilaisten mittausten, testien tai tutkimusten tulkintojen arviointiin. (Kane 2001, 328.) Nykykäsitys validiudesta sisältää päättelyn ja oletusten tueksi laajennetun analyysin, johon kuuluu esitettyjen tulkintojen perusteleva ja muiden mahdollisten, kilpailevien tulkintojen esilletuominen (Messick 1989a; 1989b). Moss on lisännyt validius-käsitettä muovanneiden tutkijoiden ketjuun myös Kanen, sillä tämän esittämä yhtenäinen teoria validiudesta (*unified theory of validity*) ja validiuskäsitteen yhtenäisyydestä (*validity as unified concept*), syntyivät Curetonin, Cronbachin, Messickin validiutta teoretisoineiden tutkimusten ja artikkeleiden pohjalta (Moss 2007, 471).

Alun perin Cronbach ja Meehl (1955) omaksuivat validiutta koskevan ajattelunsa teoreettiseksi perustaksi hypoteettis-deduktiivisen lähestymistavan, jossa paikkansapitävät perusoletukset, aksioomat, liitettiin määritelyihin termeihin, joita he kutsuivat teoreettisiksi ajatusmalleiksi (*theoretical constructs*) tai käsitteiksi. Aksioomat eivät sinänsä perustuneet minkään havainnointitulosten pohjalle, vaan ne olivat johdettavissa tai pääteltävissä kulloisenkin teorian pohjalta. Saatua testitulosta tulkittiin suhteessa siihen, kuinka hyvin se toteutti teoriasta johdettua ennustetta. Mikäli havainnot tai tulokset eivät vastanneet teoriasta johdettuja ennusteita, teoriaa pidettiin vääränä tai mittaustapaa ei pidetty sopivana indikoimaan teorian

sisältämää ajatusmallia tai käsitettä. Jos havainnot tai tulokset olivat epäjohdonmukaisia suhteessa teoriaan, tulos ei ollut validi ja sitä ei voitu hyväksyä. (Kane 2001, 321; Kane 2006, 20.) Hypoteettis-deduktiivisessa mallissa oletettiin, että on olemassa hyvin määritelty teoria, josta empiiriset ennusteet voidaan johtaa (Kane 2006, 20). Konstrukti- eli käsitevalidiudella ei ollut tarkoitus korvata aiempia ”validiuden lajeja” kuten sisältö- tai kriteerivalidiutta. Alun perin Cronbach ja Meehl esittivät käsitevalidiutta yhdeksi validiustarkastelun näkökulmaksi kriteeri- ja sisältövalidiuden rinnalle esimerkiksi silloin, kun koe mittasi laadullisia ominaisuuksia (Cronbach & Meehl 1955, 282) tai sellaisia ominaisuuksia, joiden mittaamiseksi ei ollut mahdollista luoda riittäviä kriteerejä (Cronbach & Meehl 1955, 299). Vaikka Cronbach ja Meehl pitivät käsitevalidiutta kaikenkattavana validiuden näkökulmana, he eivät luoneet sille yhtenäistä teoriaa tai validiustarkastelun teoreettista viitekehystä eivätkä he pitäneet käsitevalidiutta edes ensisijaisena validiuden tarkastelunäkökulmana (Kane 2001, 322; Kane 2006, 20).

Cronbach pyrki erottamaan erilaisia validoinnin lähestymistapoja, joissa hän liitti validiuden ajattelumallin teoreettisiin muuttujiin. Hän totesi ”ettei ole olemassa mitään olennaisia ja ennustettavia kriteereitä eikä myöskään tiettyjä sisältöjä, joiden perusteella esimerkiksi sisältöalueita voitaisiin luokitella” (Cronbach 1971, 462). Cronbach piti tärkeänä, että validiustarkastelu perustui kattavaan arviointiin ja erillaiseen näyttöön (Cronbach 1971, 445). Käsitevalidius ei voinut hänen mukaansa perustua vain yhden validiuskertoimen tai yhden asiantuntijan mielipiteeseen, vaan tuloksen selittämiseksi tuli hyödyntää teoriaa, joka selitti tuloksen syitä ja vaikutuksia (Cronbach 1971, 443). Se, minkä tuli olla validi, ei ollut testi tai koe sinänsä, vaan dataan⁵⁷ tai aineistoon perustuva tulkinta (Cronbach 1971, 447).

Campbell ja Fiske (1959) olivat huolissaan siitä, onko testien tai kokeiden rakenneiden mittaaminen riittävää validiustarkastelun kannalta (Campbell & Fiske 1959, 100) ja siksi he kehittivät validiuden tarkasteluun monimenetelmällisen matriisin, jolla saatettiin määritellä muuttujia. Kehittämässään menetelmässä he olivat edelläkävijöitä, sillä matriisissa yhdistettiin eri ominaisuuksia toisiinsa teoreettisiin ominaisuuksiin validiuden varmistamiseksi. Heidän mielestään konvergenssi eli samaan suuntaan ja divergenssi eli eri suuntiin osoittavat havainnot olivat tärkeä osa käsitevalidiuden tarkastelussa. (Kane 2006, 36.)

57 Data ei Messickin mukaan ole sama asia kuin tieto tai informaatio, sillä tieto tai informaatio on jo tulkittua dataa: ”Data are not information; information is that which results from the interpretation of data” (Messick 1980, 1014). Vaikka suomen kielessä data-sanana käyttö on kömpelöä, käytetään sitä tässä tutkimuksessa Messickin ajatuksen mukaisesti, kuvaamaan vielä tulkitsematonta aineistoa.

1970-luvun lopulle tultaessa oli muodostunut kaksi suuntausta validiusteorian kehittämiseksi. Ensimmäinen suuntaus pyrki laatimaan selkeitä määritelmiä, joiden avulla voitaisiin validoida erilaista näyttöä testitulosten tulkitsemiseksi tai tulosten hyödyntämiseksi. Toinen, kilpaileva suuntaus pyrki kehittämään yhdenmukaista validius-käsitettä (*unified validity*). (Kane 2006, 20.) Vaikka Cronbach ja Meehl (1955) eivät koskaan luoneet varsinaista teoriaa validiudesta, rakensi Messick validiudelle yhtenäistä, teoreettista viitekehystä nimenomaan Cronbachin ja Meehlin käsitevalidiutta koskevan ajattelumallin pohjalta (Kane 2001, 325). Myöhemmissä teoretisoinneissaan Cronbach (1980) painotti validiustarkastelun sosiaalista ulottuvuutta ja asiayhteyden merkitystä testitulosten tulkinnassa. Cronbachin lisäksi myös Messick on painottanut omissa tutkimuksissaan validiuden sosiaalista ja yhteiskunnallista ulottuvuutta (Messick 1980, 1012), sillä hänen mielestään validiudessa yhdistyy arvio siitä, ”missä määrin empiirinen näyttö ja teoreettinen ajattelutapa tukevat riittävästi ja tarkoituksenmukaisesti testituloksiin ja muihin arviointeihin perustuvia johtopäätöksiä” (Messick 1989b, 13). Testin validiustarkastelussa keskeistä ovat sen tulkittavuus, merkitys ja tulosten hyödynnettävyys. Validius on integroiva ja arvioiva käsitys testitulosten perusteella tehtävistä johtopäätöksistä ja testin hyödyntämisen sosiaalisista seurauksista (Brennan 2006, 2). Kuten Kaplan, myös Messick yhdisti eettiset, arvioihin perustuvat kysymykset mittaamisen kaikkia tasoja koskeviksi testipatteriston luomisesta ja analysoimisesta aina testituloksen tulkintaan. Olennaista Messickin mukaan oli nostaa validiustarkastelussa pohdittavaksi, keiden arvojen tai standardien mukaan validiutta tarkastellaan ja mitä seurauksia esimerkiksi kielteisillä tai myönteisillä testi- tai arviointituloksilla voi olla. (Kaplan 1964, 198; Messick 1980, 1013.)

Messick määritteli validiutta kahden keskeisen kysymyksen kautta: onko testi tai koe riittävän hyvä mittaamaan niitä ominaisuuksia, joita sen tulisi mitata ja soveltuuko testi tai koe käytettäväksi esitettyyn tarkoitukseen esitetyllä tavalla. Ensimmäiseen kysymykseen hänen mielestään löytyy vastaus testien psykometristen ominaisuuksien avulla ja tarkastelemalla asetelmaa käsitevalidiuden näkökulmasta. Toiseen kysymykseen vastaaminen edellytti hänen mielestään eettistä näkökulmaa sekä sosiaalisten arvojen kytkemistä vastaukseen. (Messick 1980, 1012.) Vaikka validiutta voitiin tarkastella monesta muustakin näkökulmasta kuin sisällön, kriteereiden tai käsitteen mukaan, Messickin mielestään validiutta koskevat kysymykset oli mahdollista luokitella kahteen ryhmään: kysymyksiin, jotka vastaavat, onko testi tai koe mitannut sitä, mitä sen on oletettu mittaavan, ja kysymyksiin, joiden avulla voidaan tarkastella mittauksen hyödyllisyyttä ja ennustettavuutta. Messick korosti, että erilaiset, testitulosten perusteella tehtävät johtopäätökset edellyttivät erilaista näyttöä, kuten erilaista dataa, faktoja tai argumentteja, ei ”erityyppistä” validiutta. Käsitys siitä, että validius on kaikenkattava arvio testin tulkinnasta ja sen käyttämisestä, perustuu siihen, että tulkinnassa keskitytään näytön luonteeseen sekä näytön ja testituloksen perusteella tehtävien johtopäätösten väliseen suhteeseen. (Messick 1980, 1013–1014.)

Koska Messickin määritelmä validius-käsitteestä on usein ymmärretty väärin niin, että hän olisi teoreettisessa viitekehyksessään kuvannut validiuden eri lajeja, joiden lähtökohdat validiuden tarkastelulle olisivat perustuneet erilaisiin argumentteihin,⁵⁸ päätyi hän muotoilemaan validiutta koskevan määritelmänsä Guionia lainaten: ”validius on yleinen tarkastelutapa, jossa perustellaan testin tulkintaa ja sen hyödyntämistä” (Guion 1978, 500; Messick 1980, 1014). Validius-käsitteen ykseyden ymmärtäminen oli Messickin mukaan hankalaa, varsinkin kun eri tarkastelun lähtökohtia tai näkökulmia yhdisti validius- sana. Väärinymmärryksen vaaraa lisäsi hänen mielestään myös se, että useat validiuden eri tarkastelutavat saattoivat olla samankaltaisia tai samanaikaisesti käytettävissä, vaikka sisällöllisesti tai merkitykseltään ne eivät olleet samoja. (Messick 1980, 1014.)

Jotta validiustarkastelun erilaiset näkökulmat olisivat selkeämmin ymmärrettävissä, Messick laati luokittelevan ja validiuden eri tarkastelunäkökulmia kuvaavan listan. Siinä hän nimesi erilaisia näkökulmia, joiden perusteella validiutta saatettiin tarkastella esimerkiksi testin tai tutkimusasetelman sisällön tai tutkimuksen yleistettävyyden lähtökohdista. (Messick 1980, 1015.) Taulukkoon 4 on koottu Messickin esittämiä näkökulmia, joiden avulla validiutta voidaan tarkastella. Oikeassa sarakkeessa on kuvaus siitä, mikä validiustarkastelun näkökulma kulloinkin on keskeistä. Koska käännöksessä on pyritty seuraamaan Messickin ajatusta validiuden tarkastelunäkökulmista, käännökset eivät ole sanatarkkoja.

58 Näin validiutta tulkitisivat esimerkiksi Zeller (1988, 322), Dunnette ja Borman (1979) sekä Lissitzin ja Samuelson (2007a; 2007b).

TAULUKKO 4. Messickin luokittelemia näkökulmia validiuden tarkastelemiseksi.

Lähde: Messick 1980, 1015.

Validiuden näkökulma	Kuvaus siitä, mikä validiudessa keskeistä
Sisällön validius (<i>Content validity</i>)	Sisällön merkitys keskeinen, yksityiskohtainen määrittelyala, sisällön kattavuus ja edustavuus tärkeitä (<i>Content relevance—domain specifications, Content coverage—domain representativeness</i>)
Kriteerivalidius (<i>Criterion validity</i>)	Kytkeytyy kriteereihin (<i>Criterion relatedness</i>)
Ennustevalidius (<i>Predictive validity</i>)	Merkitys ennustettavuuden kannalta (<i>Predictive utility</i>)
Samanaikaisvalidius (<i>Concurrent validity</i>)	Merkitys diagnostinen (<i>Diagnostic utility, Substitutability</i>)
Käsitevalidius (<i>Construct validity</i>)	Tulkinnan mielekkyys (<i>Interpretive meaningfulness</i>)
Konvergenssivalidius (<i>Convergent validity</i>)	Konvergenssin johdonmukaisuus (<i>Convergent coherence</i>)
Erotteleva validius (<i>Discriminant validity</i>)	Erottelukyky (<i>Discriminant distinctiveness</i>)
Ominaisuuteen liittyvä validius (<i>Trait validity</i>)	Ominaisuuksien vastaavuus (<i>Trait correspondence</i>)
Nomologinen validius (<i>Nomological validity</i>)	Kausaalinen selittäminen, yleistä lakia selittävä ilmiö (<i>Nomological relatedness</i>)
Faktoreihin liittyvä validius (<i>Factorial validity</i>)	Faktoreiden kokoonpano (<i>Factorial composition</i>)
Sisällöllinen validius (<i>Substantive validity</i>)	Sisällöllinen yhdenmukaisuus (<i>Substantive consistency</i>)
Rakenteellinen validius (<i>Structural validity</i>)	Rakenteita noudattava (<i>Structural fidelity</i>)
Ulkoinen validius (<i>External validity</i>)	Ulkoinen yhdenmukaisuus (<i>External relatedness</i>)
Populaatio (<i>Population validity</i>)	Populaatioon yleistettävä (<i>Population generalizability</i>)
Ekologinen (<i>Ecological validity</i>)	Ekologinen yleistettävyyys (<i>Ecological generalizability</i>)
Ajallinen validius (<i>Temporal validity</i>)	Ajallinen jatkuvuus ja yleistettävyyys aikakaudesta toiseen (<i>Temporal continuity—across developmental levels</i> <i>Temporal generalizability—across historical periods</i>)
Tehtävään liittyvä validius (<i>Task validity</i>)	Tehtävän yleistettävyyys (<i>Task generalizability</i>)

Jotta erilaisia validiudennäkökulmia saattoi erotella toisistaan, niitä tuli tulkita tutkimus- tai testituloksen tarkoituksenmukaisuuden, mielekkyyden ja hyödyllisyyden näkökulmista (Messick 1994, 744).

Monista eri validiuden tarkastelunäkökulmista huolimatta Messick korosti, että kaikki mittaaminen tuli liittää käsitevalidiuden tarkasteluun (Messick 1980, 1019). Messickin validiuskäsitys perustui siihen, että rakenteellisesti validin testin, kokeen tai tutkimuksen on sisällöllisesti oltava olennaista, edustavaa ja teknisesti laadukasta. Käsitevalidiudessa yhdistyvät Messickin mukaan sekä kriteerit että sisältö, sillä käsitevalidiuden tarkastelussa testataan järkeviä hypoteeseja niiden teoreettisesti olennaisissa yhteyksissä (Messick 1980, 1015). Perusteluna näkemykselleen käsitevalidiuden yhtenäisyydestä Messick esitteli kuusi tarkastelutapaa, jotka liittyvät keskeisesti käsitevalidiuden sisältöön, luotettavuuteen, rakenteellisuuteen, yleistettävyyteen, merkittävyteen ja ulkoiseen näkökulmaan. Nämä eri tarkastelutavat sopivat Messickin mukaan yleensä validiustarkastelun kriteeriksi tai standardeiksi niin käyttäytymistieteissä kuin psykologiassa. (Messick 1995, 744.)

Messick piti tärkeänä, että arvioitavat tai mitattavat tiedot, taidot tai muut ominaisuudet sekä näiden rajapinnat määrittämään mahdollisimman täsmällisesti. Hän piti sisällön tarkastelua hyödyllisenä esimerkiksi työtehtäviä analysoidessa tai opetus-suunnitelma-analyysessä laadittaessa. Luotettavuuden näkökulmalla Messick viittaa teoreettisen ajattelutavan johdonmukaisuuteen, joka ilmenee havainnoiduissa testivastauksissa sekä empiiriseen näyttöön, joka kytkee vastaajan arvioinnin tehtävään. Rakenteellisessa näkökulmassa arvioidaan testituloksen rakenteen luotettavuutta ja tarkkuutta suhteessa määriteltyyn ajatusmalliin tai käsitteeseen. Yleistettävyydessä tarkastellaan sitä, missä määrin testitulokset ja niiden tulkinnat ovat yleistettävissä esimerkiksi eri väestöryhmissä. Käsitevalidiuden ulkoisella näkökulmalla Messick tarkoitti eri mittausmenetelmien käyttöä ja niiden avulla saatujen tulosten keskinäistä vertailua. (Messick 1995, 745.) Mittauksen tai arvioinnin tulee joko konvergoida eli osoittaa samaan suuntaan tai diskriminoida eli erotella (*discriminate*). Erottelevan mittauksen tulee osoittaa erottelevuus suhteessa toiseen mittaukseen tai arviointiin. (Messick 1995, 746.) Merkittävyydellä Messick tarkoitti testituloksen tulkinnan vaikutuksia, samoin kuin sitä, miten testituloksen käyttöä arvioidaan. Tulosten seurauksia ja vaikutuksia voidaan arvioida osana käsitevalidiutta, jolloin tarkastellaan testien tosiasiallisia ja mahdollisia seurauksia. Näin erityisesti, kun tarkastellaan epävalidiutta aiheuttavia tekijöitä, kuten puolueellisuutta, ennakkosenteita tai epäoikeudenmukaisuutta. Messickin mukaan tutkimus- ja testitulosten tulkinta sekä niiden sosiaalinen ulottuvuus voivat olla vaikutuksiltaan joko myönteisiä tai kielteisiä. Esimerkkinä myönteisestä tuloksen vaikutuksesta voisi olla kansainvälisten oppimistulosvertailujen seurauksena kansalliseen koulutuspolitiikkaan tehtävät uudistukset. Kielteisenä esimerkkinä voisi puolestaan olla oppilaiden suoritusarviointi, silloin kun tulokset ja niistä tehtävät tulkinnat ovat puolueellisia tai epäoikeudenmukaisia. (Messick 1995, 745–746.)

Cronbachin, Messickin ja Kanen näkemyksiä validiudesta yhdistää ennen kaikkea se, että he pitävät käsitevalidiutta kattavimpana validiuden näkökulmana. 1990- ja 2000-luvuilla validiutta teoretisoinut Kane on validiuskäsitteensä kuitenkin enemmän Cronbachin kuin Messickin linjoilla. Cronbachin (1988) mukaan tulosten tulkinnan ja käytön validointi perustuivat arvioinnin argumentoinnissa käytettyyn logiikkaan. (Kane 2006, 22.) Kane laajensi Cronbachin esittämää ajatusta käsitteellistämällä validiutta niin, että se tarjosi kokonaisarvioinnin testituloksesta ja sen tulkinnasta.⁵⁹ Kane tarkasteli validiutta ennen kaikkea validointiprosessin kautta. Vaikka validius ja validointi⁶⁰ ovat kaksi erillistä, mittaamiseen liittyvää käsitettä, ne myös kytkeytyvät toisiinsa. Kane korostaa yleistä validoinnin metodologiaa, joka perustuu validiuden käsitteellistämiseen argumentoinnissa⁶¹ (Brennan 2006, 2). Kane määritteli validoinnin kahdella tavalla. Validoinnissa on ensinnäkin kyse siitä, että esitettyä tulkintaa tuetaan ja sen käyttöä perustellaan tehtyjen havaintojen tai saadun näytön perusteella. Validoinnin tehtävä on tällöin kuin asianajajan, joka rakentaa puolustusta esitettyjen tulkintojen pohjalle. Toiseksi validoinnissa on kyse siitä, missä laajuudessa esitetyt tulkinnat ja niiden käyttö ovat asianmukaisia ja uskottavia. (Kane 2006, 17.) Validoinnissa ei ole kyse ”teknisten temppeujen” tekemisestä, vaan siinä edellytetään, että tavoitteet, validoinnin kriteerit ja validointia koskeva yleinen lähestymistapa ovat yhdenmukaisia (Kane 2001, 329). Validoinnissa testin tulkinta ja sen käyttö tulee olla perusteltua (Kane 2006, 22). Kanen formuloinnissa, validiusargumentointi⁶² (*validity argument*), tarjoaa kokonaisarvioinnin mahdollisesta tulkinnasta aina testituloksen hyödyntämiseen. Validoinnin tavoitteena on johdonmukainen analyysi, jossa näyttö, jota koetellaan mahdollisen tulkinnan ja sen hyödyntämisen puolesta tai sitä vastaan, tähtää uskottavien, keskenään kilpailevien tulkintojen saamiseksi (Brennan 2006, 2).

Kane kuvasi validointia kuusivaiheisena prosessina, joka perustuu argumentointiin, ja jossa arvioidaan tulkinta-argumentteja (*interpretative argument*) (Kane 1990, ii, 33). Tulkinta-argumentit muodostavat viitekehiksen, jossa validiusargumenttia kehitetään. Validointiprosessi on Kanen mukaan sikäli päättymätön, että validointia voi kehittää aina vain paremmaksi, ilman, että se valmistuisi koskaan lopullisesti. Validoinnin kolme ensimmäistä vaihetta muodostivat ns. formatiivisen ja kolme seuraavaa ns. summatiivisen vaiheen. (Kane 1990, 21.) Myöhemmin Kane kutsuu ns. formatiivista vaihetta kehittämisvaiheeksi (*development stage*), sillä siinä keskitytään mittausmenetelmien kehittämiseen ja niitä vastaaviin tulkinta-argumentteihin. Summatiivista vaihetta Kane kutsuu myöhemmin arviointivaiheeksi (*appraisal*

59 Kanen käsitystä validiusargumentoinnista käsitellään vielä luvussa 4.3.1, jossa tarkastellaan erityisesti testien validiutta. Validiusargumentoinnin käyttöä tässä tutkimuksessa tarkastellaan puolestaan luvussa 4.4.

60 Sivistyssanakirjan mukaan validoinnilla voidaan tarkoittaa suomen kielessä sitä, että validoitava kohde täyttää tietyt kriteerit tai sillä voidaan tarkoittaa sitä, että lopputulos on oikein muodostettu. Joissain tapauksessa validoinnissa voi olla kyse siitä, että esimerkiksi tulosten tai johtopäätösten oikeellisuus on tarkastettu. Määritelmä löytyy osoitteesta: <http://suomisanakirja.fi/validointi>

61 Tässä tutkimuksessa englannin kielen *argument*-sana on käännetty asiayhteydestä riippuen joko suomen kielen sanoilla argumentti, perustelu tai argumentointi.

62 Hildén käyttää termistä *validity argument* käännöstä validiusargumentti ja termistä *interpretive argument* käännöstä tulkinta-argumentointi (Hildén 2008, 757).

stage), sillä siinä painopiste on puolestaan tulkinta-argumenttien kriittisessä arvioinnissa. (Kane 1990, 21; vrt. Kane 2006, 25–26.) Käytännössä nämä kaksi edellä esitettyä vaihetta limittyvät keskenään. Tavoitteena on kehittää arviointiprosessi, joka tukee esitetyn testituloksen tulkintaa ja sen käyttöä. Kane pitää välttämättömänä, että testin laatijalla on ennakkokäsitys siitä, miten testistä saadaan kehitettyä paras mahdollinen. Alustava ehdotus testituloksen tulkinnasta ja testituloksen hyödyntämisestä laaditaan ensin hyvin yleiseksi. Tämän jälkeen testin laatijat määrittelevät yleisellä tasolla sen, miten testille asetettava tavoite määritellään, jonka jälkeen kehitetään mittausprosessi. Johdonmukaisen mittausprosessin aikaansaamiseksi edellytetään, että tulkinta-argumentit ovat uskottavia. Formatiivisen eli kehittämisvaiheen tarkoituksena on kehittää testiä ja uskottavia tulkinta-argumentteja, jotka kuvaavat esitettyjä testitulosten tulkintoja ja niiden hyödyntämistä. (Kane 1990, 21.) Validointiprosessi alkaa rakentamalla mahdollisimman selkeä tulkinta-argumentti, jossa mahdollisesti myös tunnistetaan ne erityiset oletukset, jotka tukevat johtopäätöstä. Tämän jälkeen arvioidaan tulkinta-argumentin uskottavuutta ja johdonmukaisuutta oletusten ja johtopäätösten rationaalisuuden perusteella. Formatiivinen vaihe on pikemminkin analyttinen kuin empiirinen. Kane huomauttaa, että joissain tapauksissa oletuksia ja johtopäätöksiä on mahdollista tarkastella aiempien tutkimusten tai empiirisen aineiston valossa, joskus taas oletusten arviointi saattaa perustua ainoastaan yleiseen tietoon tai kokemukseen. Toisinaan validointiprosessissa havainnot tai näyttö tukevat suoraan tulkinta-argumenttia, toisinaan taas näyttö voi olla niin poikkeuksellista, että tulkintaa on mahdotonta jatkaa ja validointiprosessi on keskeytettävä. Mikäli tulkinta-argumenttiin tai testiprosessiin joudutaan tekemään suuria muutoksia, Kane kehottaa palaamaan validointiprosessissa alkuun ja muotoilemaan tulkinta-argumenttia uudelleen. Kaikkein yleisintä on kuitenkin se, että näytön tai havaintojen perusteella tehdään muutoksia joko tulkinta-argumenttiin tai testiprosessiin, jonka jälkeen validointiprosessi etenee summatiiviseen vaiheeseen. (Kane 1990, 35.)

Kun testin kehittäminen voidaan katsoa päättyneeksi, siirrytään summatiiviseen eli arviointivaiheeseen, jossa tulkinta-argumentteja haastetaan. Kanen mukaan voidaan olettaa, että testin kehittämisvaiheessa on saatu aikaan täsmällinen ja johdonmukainen tulkinta-argumentti, joka yhdistää testisuorituksen (*test performance*) esitettyyn tulkintaan ja sen käyttöön. (Kane 2006, 26.) Summatiivisessa vaiheessa tutkitaan ja arvioidaan esitettyjä oletuksia empiiristen havaintojen tai näytön perusteella. Validointiprosessissa pyritään tunnistamaan mahdolliset tulkinta-argumentin heikkoudet ja siksi kaikkein vaikeimmat, haavoittuvimmat, heikoimmat ja ongelmallisimmat tulkinta-argumentit pyritään koettelemaan eli tutkimaan empiirisen näytön avulla. (Kane 1999, 36.) Tulkinta-argumentit voivat olla ongelmallisia esimerkiksi siksi, että empiirisen näytön perusteella voidaan osoittaa, että oletukset eivät ole todenmukaisia tai havainnot eivät tue esitettyä oletusta (Kane 1990, 25). Mikäli empiirinen aineisto tai havainnot osoittavat oletukset vääriksi, tulee Kanen mukaan prosessissa palata takaisin alkuun. Mikäli taas aineisto tukee oletuksia, voidaan validointiprosessissa jatkaa loppuun. Jos kaikki tulkinta-argumentista tehtävät johtopäätökset ja oletukset ovat ongelmattomia, voidaan tulkinnan validius

hyväksyä. Jos taas argumentti ei ole riittävän hyvä, on palattava prosessissa takaisin joko kehittämisvaiheen tai arviointivaiheen alkuun. (Kane 1990, 36; ks. myös Kane 2001, 330; Kane 2006, 24–25.) Kanen mukaan arviointivaiheessa kaikkein kyseenalaisimmat oletukset, joita tulkinta-argumentista voidaan tehdä, ovat yleensä informatiivisimpia. Jos tulkinta-argumentit kestävät koettelemisen, niiden uskottavuus kasvaa. Jos taas tulkinta-argumentit eivät kestä haastamista, ne tulee joko tarkistaa tai hylätä. (Kane 2006, 26.)

Kane on luokitellut tulkinta-argumenttien viisi ominaisuutta, joita hän pitää validiuden kannalta olennaisina. Tulkinta-argumentit ovat 1) ihmisten luomia ja määrittelmiä, joten niitä voidaan kehittää, tarkistaa ja hylätä, 2) jäsentyneet siten, että tehtäessä johtopäätöksiä, niiden rooli voi joidenkin oletusten yhteydessä olla suuri, joidenkin toisten oletusten yhteydessä taas pieni, 3) dynaamisia, sillä niiden tarkastelun kohde voi laajentua, pienentyä tai muuttua kokonaan, 4) muokattavia tiettyihin tilanteisiin sopiviksi ja 5) avoimia, sillä ne ovat epätäydellisiä. (Kane 1990, 28.) Tulkinta-argumenttien tulee myös aina olla selkeitä, johdonmukaisia ja uskottavia, sillä niiden tärkein tehtävä on tehdä oletuksista ja johtopäätöksistä mahdollisimman selkeitä (Kane 2006, 29).

Validoinnissa on Kanen mukaan kyse siitä, että testituloksen tulkinta on oikeaa eli validia. Validointi edellyttää, että saatujen empiiristen havaintojen, näytön tai tulosten perusteella tehtyjen tulkintojen täytyy olla uskottavia ja objektiivisia ja tulkintoja tulee voida koetella argumentoimalla joko niiden puolesta tai niitä vastaan. (Kane 2006, 17.) Testin tai tutkimuksen sisältö ja sen tulkinta ovat validoinnissa keskeisiä, sillä niiden avulla määritellään tehtyjä havaintoja tai kerättyä aineistoa (Kane 2001, 330). Argumentointiin perustuva lähestymistapa tasapainottaa toisaalta testituloksen tulkintaan perustuvaa validointia, toisaalta yksityiskohtien tulkintaa (Kane 2001, 329). Kun argumentin uskottavuutta arvioidaan, validoidaan tulkinta-argumentti. Tulkinta-argumenttia tukevat havainnot tai näyttö argumenttoivat validiutta ja siitä tehtäviä tulkintoja. Validiusargumentoinnin yksityiskohdat riippuvat puolestaan yksittäisistä johtopäätöksistä tai oletuksista, joita tulkinta-argumenttien avulla tehdään. (Kane 1990, ii; ks. myös Kane 2006, 22.)

Kanen mukaan validiusargumentointi heijastaa yleisiä, käsitevalidiuden luonteen kuulumia periaatteita. Tulkinta-argumentti tarjoaa selkeän väitteen päätelmien ja oletusten perusteella laadituista tulkinnoista ja testitulosten käytöstä. Kyseisiä päätelmiä ja oletuksia tulee arvioida analyysien ja empiiristen tutkimusten avulla. Kanen mukaan argumentointiin perustuva lähestymistapa on validoinnissa hyödyllistä, sillä tieteellisen tutkimuksen tai mittausprosessien avulla saatu näyttö validiudesta tukee tulkinnoista ja oletuksista laadittuja tulkinta-argumentteja. (Kane 2006, 23; ks. myös Kane 1990, 18–19). Kane pitää argumentaatioon perustuvan validoinnin etuna myös sitä, että validointia voidaan käyttää minkä tahansa testituloksen tulkinnassa, eikä se suosi tiettyä testityyppiä muiden kustannuksella. Validointi tarjoaa mahdollisuuden luoda suuntaviivat, joiden avulla voidaan systemaattisesti arvioida esitetyn testituloksen tulkinnan ja testin hyödyntämisen validiutta.

Vaikka validointi ei anna mitään absoluuttista totuutta validiudesta, se antaa kuitenkin mahdollisuuden mitata sitä, miten validiustarkastelussa on edistytty. Lisäksi validointiprosessi yhdenmukaistaa tulkinta-argumenttien ja oletusten testaamista. (Kane 1990, 38–39.)

Vaikka yhtenäinen, käsitevalidiuteen perustuva, validiuden tarkastelutapa sai laajalti hyväksyntää, se ei kyennyt antamaan selkeitä ohjeita testituloksen tulkinnan ja käytön validoimiseksi. Kaikki eivät liioin olleet yksimielisiä Messickin teoretisoimasta yhtenäisestä käsitevalidiudesta. Moderni, yhtenäinen validiusteoria on herättänyt myös arvostelua ja kritiikkiä, jota käsitellään seuraavassa luvussa.

4.2.2 Modernin validiuskäsitteen haasteita ja kritiikkiä

Vuosikymmenten saatossa Messickin teoria käsitevalidiuden ykseydestä on saanut myös kritiikkiä. Käsitevalidiuden tarkastelunäkökulman soveltamista on arvosteltu erityisesti silloin, kun tutkimusasetelmaa ei ole johdettu vahvasta teoriasta. Jos vahvaa teoriaa ei ole ollut, käsitevalidiudella on taipumusta olla hyvin avointa. Käsitevalidiuden tarkastelussa ei ole myöskään ollut itsestään selvää, mistä validiustarkastelu tulisi aloittaa tai miten prosessia tulisi mitata. (Kane 2006, 21.) Esimerkiksi Cronbach piti validoinnille olennaisena piirteenä, että se on pitkälinen, jopa loputon prosessi (Cronbach 1989, 151), jossa kaikkea mahdollista dataa voidaan pitää olennaisena validiustarkastelun näkökulmasta (Kane 2006, 22). Sekä Cronbach (1988) että Kane (2001) ovat esittäneet omat ratkaisunsa yhtenäisen käsitevalidiuden mukanaan tuomien ongelmien ratkaisemiseksi.

Cronbachin (1988) mukaan käsitevalidiuden soveltaminen on vaikeaa erityisesti silloin, kun tutkimusasetelma ei nojaudu vankkaan tai yhtenäiseen teoriaan, kuten on usein esimerkiksi yhteiskuntatieteissä. Yhtenäisen teorian puuttuminen johtaa hänen mielestään helposti epäselviin, jopa virheellisiin tulkintoihin käsitevalidiudesta. Kuvatakseen tätä käsitevalidiuteen liittyvää ongelmaa Cronbach erottaa toisistaan käsitevalidiuteen kuuluvan heikon ohjelman (*weak program*) ja vahvan ohjelman (*strong program*).⁶³ Heikko ohjelma antaa ikään kuin alustavaa tietoa, ja siinä mikä tahansa testituloksen ja muuttujan välinen korrelaatio on tervetullutta ja toivottavaa. Heikko ohjelma on luonteeltaan niin avoin, että mitä tahansa havaintoa, vaikka vain etäisesti testitulokseen liittyvää, voidaan pitää olennaisena validiuden näkökulmasta. Käsitevalidiuden heikon ohjelman avoimuuden vuoksi ei ole kuitenkaan aina selvää, miten edes oleellisimpia havaintoja tulisi tulkita. Käsitevalidiuden heikon ohjelman haasteena on sen rajoittamattomuus (Kane 2001, 327). Jos selkeä määritelmä puuttuu, saatetaan validiustarkastelussa helposti päätyä siihen

63 Kane rinnastaa edellisessä luvussa 4.2 kuvattua argumentaatioon perustuvaa validointia Cronbachin teoretisoimaan käsitevalidiuden vahvaan ohjelmaan. Kanen mukaan hänen käyttämänsä termi ”argumentointiin perustuva validiuden lähestymistapa” tarkoittaa samaa kuin käsitevalidiuden vahva ohjelma. Vaikka Cronbach ja Meehl (1955) esittivät, että käsitevalidius liittyy vahvaan teoriaan, Kanen mukaan tulkinta-argumenttien ei välttämättä tarvitse perustua teorioihin; tulkinta-argumentteihin perustuvien väitteiden tulee ainoastaan olla riittävän selkeitä, jotta niitä voidaan koetella ja arvioida. (Kane 1990, 39.)

johtopäätökseen, että mikä tahansa havainto validiudesta liittyy käsitevalidiuteen. Jos näin olisi, se johtaisi Kanen mukaan siihen, että kaikki validiustarkastelu olisi käsitevalidiuden tarkastelua ja tulkintaa. Toisaalta ei ole myöskään itsestään selvää, että vahva ohjelma pitää sisällään kaikki validiustarkasteluun kuuluvat pyrkimykset. (Kane 2001, 326–327.)

Vahva ohjelma ei voi toteutua ilman vankkaa teoriataustaa, sillä se edellyttää käsitteiden täsmällistä ilmaisemista ja ennalta asetettuihin kysymyksiin tai ongelmiin vastaamista (Cronbach 1988, 12–13). Kane pitää kahden keskenään kilpailevan ohjelman, heikon ja vahvan, kehittämistä välttämättömänä, sillä käsitevalidiuden alkuperäinen muotoilu keskittyi teoreettisten ajatusmallien tai käsitteiden määrittelyyn formaalin teorian pohjalta. Vaikka teoreettiset muotoilut sinänsä olivat hyvin konstruoituja, ne antoivat niukasti eväitä kasvatus- tai yhteiskuntatieteille, joissa formaalit teorat olivat jo pitkälle kehittyneitä. Käsitevalidiuden vahvaa ohjelmaa ei yleensä sovellettu muussa kuin sen puhtaimmassa ja teoreettisimmassa muodossa. Siitä huolimatta, että vahvan ohjelman täytäntöönpanoon kehitettiin menetelmiä, käsitevalidius pysyi verrattain abstraktina. (Kane 2006, 326.) Vaikka käsitevalidiuden näkökulmaa väljennettiin, jotta se olisi helpommin sovellettavissa, pidettiin kuitenkin samaan aikaan kiinni tiukasta, muodollisesta teoriasta. Seurauksena oli, että käsitevalidiuden heikosta ohjelmasta tuli abstraktimpi, mutta sitä ei tuettu riittävästi teoreettisesti. (Cronbach 1988, 12.) Cronbachin ja Meehlin (1955) hahmottelema vahva ohjelma perustui kapeaan tarkastelun kohteeseen, ja sen avulla voitiin alistaa teoreettisia oletuksia ja johtopäätöksiä empiirisille haasteille. Oleelliseksi käsitevalidiuden tarkastelussa muodostui teorian tieteellinen testaaminen. Ongelmana vahvaan ohjelmaan perustuvassa lähestymistavassa oli kuitenkin sen rajalliset hyödyntämismahdollisuudet, jos kehittynyttä testiteoriaa ei ollut. (Kane 2001, 236–237.)

Kane pitää yhä yhtenäisen käsitevalidiuden teorian haasteena vahvaa ohjelmaa, joka saattaa estää validoinnin tyydyttävän tasoisen toteuttamisen. Kane pitää käsitevalidiuden vahvan ohjelman toteuttamista vaikeana jopa silloin, kun validiustarkastelun tukena on vankka teoria. Validiustarkastelu muuttuu hänen mukaansa lähes mahdottomaksi toteuttaa, mikäli teoriaa ei ole lainkaan. Teorioiden arviointi empiirisesti muuttui hankalaksi, kun käsitevalidiudessa luovuttiin perinteisestä jaottelusta teoreettisten ajattelumallien ja havaittavien ominaisuuksien välillä. Kanen mukaan teorian testaamisessa ja koettelemisessa on välttämätöntä, ettei kaikkia muuttujia johdeta samasta teoriasta. Joidenkin muuttujien on hyvä olla jopa sellaisia, että niitä voidaan tulkita ilman erityistä teoriataustaa. Vahva ja heikko ohjelma ovat ongelmallisia myös siksi, että jos vahvaa ohjelmaa ei saavuteta, saatetaan luiskahtaa helposti heikkoon ohjelmaan tai validiuskysymys saatetaan jättää kokonaan ottamatta huomioon. (Kane 2001, 331–332.) Käsitevalidiuden tarkastelua ei ole myöskään kyetty yhdenmukaistamaan operationaalisella tasolla, vaikka siitä on käyty keskustelua yli kolmen vuosikymmenen ajan. Validiudesta tehtävien havaintojen arvioimiseksi ei siis ole kriteerejä tai ainakin ne ovat vielä hyvin epävarmalla pohjalla. (Kane 2001, 327.)

Ratkaisuna edellä esitettyihin käsitevalidiuden yhtenäisyydestä aiheutuviin ongelmiin, Kane esittää argumentointiin perustuvaa validoinnin lähestymistapaa, jossa validiusargumentteja tuetaan erilaisilla tulkinta-argumenteilla ja johtopäätöksillä. Jokainen tulkinta-argumentti on ainutkertainen sikäli, että se sisältää erityisiä päätelmiä ja oletuksia, joita sovelletaan vain tietyissä asiayhteyksissä. Vaikka jokainen tulkinta on Kanen mukaan ainutkertainen, voidaan tulkintoja erotella toisistaan. Kanen mukaan on olemassa kahdenlaisia tulkintoja, joista toinen liittyy havaittaviin ominaisuuksiin (*observable attributes*) ja toinen teoreettisiin ajattelumalleihin (*theoretical constructs*) (Kane 2001, 332). Kane määrittelee havaittavat ominaisuudet keskimääräiseksi suoritustasoksi ja niiden määrittely voidaan tehdä ilman nimenomaista teoriaa. Sen sijaan teoreettiset ajattelumallit perustuvat tiettyyn teoriaan tai käsitteeseen, eikä niitä voi suoraan määrittellä havaintojen perusteella. (Kane 2001, 333.) Havaittavien ominaisuuksien ja teoreettisten ajattelumallien ero voi olla kontekstisidonnaista. Muuttuja voi olla havaittava ominaisuus erityisesti sellaisissa tapauksissa, joissa se ei tukeudu tutkittavina oleviin teoreettisiin oletuksiin. Kun havaittava ominaisuus on määritelty, sillä on suhteellisen yksinkertainen tulkinta-argumentti, jolla on selkeä validointistrategia. (Kane 2001, 334.) Kane pitää havaittavia ominaisuuksia tärkeinä, koska niillä voidaan määrittellä teorialle tavoitteet, niillä voidaan tulkita kahteen eri teoriaan perustuvaa tiettyä suoritusta, jomankumman teorian eduksi ja niillä voi olla jossain tietyssä yhteydessä käytännön merkitystä (Kane 2001, 335).

Viimeaikainen käsitevalidiuteen kohdistunut kritiikki ja siitä kumpuava keskustelu kuvaa toisaalta käsitevalidiuden kompleksisuutta, toisaalta käsitevalidiuteen yhä liittyviä tulkintaerimielisyyksiä. Lissitz ja Samuelson (2007a; 2007b) ovat kritisoineet erityisesti kolmea käsitevalidiuteen keskeisesti liittyvää seikkaa: sen teoriataustaa, siihen liittyvää terminologiaa sekä käsitevalidiuden ja sisällön validiuden keskinäistä suhdetta (Lissitz & Samuelson 2007a, 456; 2007b, 482). Lissitzin ja Samuelsonin validiusteoriaa ja siinä erityisesti Messickin käsitevalidiutta kohtaan esittämä kritiikki on synnyttänyt keskustelun, johon ovat osallistuneet validiutta jo pitkään tutkineet Gorin (2007), Embertson (2007) ja Moss (2007). Vaikka Lissitzin ja Samuelsonin esittämä kritiikki ei ole saanut myönteistä vastaanottoa, pikemminkin päinvastoin, ovat Moss, Gorin ja Embertson paitsi tuoneet esiin omia näkemyksiään käsitevalidiudesta myös ottaneet kantaa Cronbachin ja Meehlin, Messickin sekä Kanen näkemyksiin ja teoretisointeihin käsitevalidiudesta.

4.3 Arviointi ja validius

Nykyisin validius kytkeytyy jo käsitteenä arviointiin; kyse on siitä, että arvioidaan sitä, mitä on aiottu arvioida (Koppinen, Korpinen ja Pollari 1994,32). Jakku-Sihvosen (2001) mukaan arvioinnin validius voidaan määritellä sen mukaan, kuinka pitkälle empiiriset havainnot ja teoreettiset perustelut tukevat arvioinnin tuloksia ja niiden perusteella tehtyjä johtopäätöksiä sekä niiden paikkansapitävyyttä (Jakku-Sihvonen 2001, 111). Messickin mukaan arvioinnin peruslähtökohtiin kuuluu, että se on vertailtavaa, oikeudenmukaista, reliaabelia ja validia (Messick 1994, 7; Messick 1995, 742). Oikeudenmukaisuus johtaa myös vaatimukseen yhdenvertaisuudesta, jolla tarkoitetaan testituloksen ja siitä tehtävien tulkintojen arvovapautta ja puolueettomuutta. Messick pitää arvioinnin perusvaatimuksia keskeisinä, koska ne sisältävät arvoja, joilla on merkitystä, kun arvioinnin perusteella tehdään päätöksiä. Validiudella on esimerkiksi politiikan ja tieteen näkökulmista yhteiskunnallista arvoa ja merkitystä, eikä arvioinnin validiutta voi tarkastella ainoastaan testitulosten ja asetettujen kriteereiden välisten korrelaatiokerrointen avulla (Messick 1995, 742). Arviointitiedon luotettavuuden tarkastelu on laaja ja monisyinen prosessi, ja siksi myös validius-käsite on ymmärrettävä laajasti (Jakku-Sihvonen 2001, 113).

Kuten jo aiemmin tässä luvussa on esitetty, validius-käsite on tieteellisessä keskustelussa viimeisten vuosikymmenten aikana syventynyt (Virta 1999,78), mikä on myös helpottanut arviointitiedon validiuden tarkastelua (Jakku-Sihvonen 2001, 113). Esimerkiksi Scheurich suhtautuu perinteiseen validiustarkasteluun kriittisesti, sillä hänen mielestään se tukahduttaa aineistoihin sisältyvän moniäänisyyden. Hänen mukaansa validiustarkastelussa ja totuuspohdinnoissa tulisi sen sijaan kunnioittaa tutkimuskohteen moniäänisyyttä ja erilaisuutta sekä tarkastella validiutta monipuolisena, moniulotteisena ja muuttuvana. (Scheurich 1997, 85–88.)

Uusien arviointitapojen tavoitteena on ollut lisätä arvioinnin validiutta (Broadfoot 1995, 9; Brookhart 2007, 54–55) ja arvioinnin peruslähtökohdat tulisi ottaa huomioon kaikessa arvioinnissa (Messick 1994, 13; Virta 1999, 79). Jotta arviointi olisi validia, sen tulisi kattaa koko arvioitava sisältöalue mahdollisimman laajasti esimerkiksi laajuudeltaan suppeampien tehtävien avulla, mutta sen lisäksi arvioitavat tiedot ja taidot tulisi olla mahdollista esittää myös syvällisesti (Virta 1999, 85). Arvioinnin validiudessa tulee kiinnittää huomiota myös arvioijan suorittaman arvioinnin pätevyyteen, arviointituloksen tulkintaan ja analyysiin (Virta 1999, 86; Metsämuuronen 2006, 118) sekä arvioinnin asianmukaisuuteen ja mielekkyyteen. Tällöin validiutta tarkastellaan esimerkiksi sen kannalta, kuinka merkityksellisiä oppilaan tiedot ja taidot ovat, tai sen kannalta, miten hyvin arviointi osoittaa oppilaan osaamista tai miten hyvin arviointi vastaa opetussuunnitelmaa ja siinä ilmaistuja tavoitteita (Virta 1999, 91–92). Kun arviointi perustuu valtakunnallisiin kriteereihin, on Virran mukaan olennaista, että kouluissa noudatettava opetussuunnitelma ja sen toteutus vastaavat yleisiä valtakunnallisia kriteerejä. Mikäli näin ei olisi, eri koulujen oppilaat joutuisivat jo opetustilanteessa eriarvoiseen asemaan,

eikä oppilailla olisi mahdollisuutta oppia niitä tietoja ja taitoja, joita opetussuunnitelmassa painotetaan. Näin ollen arvioinnilla ei myöskään olisi opetussuunnitelmasta validiutta. (Virta 1999, 89.) Arviointikäytäntöjen epäyhtenäisyys ja satunnaisuus ovat olleet kritiikin kohteina ja arvioinnin kehittämisen lähtökohtina jo vuosikymmenten ajan. Arviointitiedon tulee olla objektiivista, jolloin oppilaan arviointi ei saa olla riippuvaista siitä, kuka arvioi (Virta 1999, 90; Atjonen 2007, 208; Vänttinen 2011, 120).

Koska oppilaan arvosanan antamisperusteina ovat usein sekä Suomessa että muualla kokeissa tai testeissä osoitetun osaamisen lisäksi oppilaan luokassa osoittama osaaminen, tarkastellaan seuraavassa kokeiden ja testien sekä opettajan luokkatyöskentelyn arvioinnin validiuteen liittyviä kysymyksiä.

4.3.1 Testin validius

Validiuden tarkastelu liittyy olennaisesti sekä arvioinnin että testien toimivuuteen, testitulosten tulkintaan ja niiden hyödynnettävyyteen.⁶⁴ Messick lähestyy testituloksen käsitettä (*test score*) laajasti, jolloin sillä tarkoitetaan erilaisissa testeissä, kyselyissä, havainnoinneissa tai muissa arvioinneissa ilmenevää yhdenmukaisuutta tai lainalaisuutta. Validiuden periaatteet ovat Messickin mukaan sovellettavissa kaikissa arvioinneissa, myös suoritusarvioinneissa. Messick pitää validiuden periaatteiden sovellettavuutta tärkeänä, sillä esimerkiksi suoritusarviointia on pidetty keskeisenä välineenä, kun koulutusta on uudistettu standardeihin perustuvaksi. Hänen mielestään kyse on sellaisesta poliittisesti merkittävästä seurauksesta, johon suoritusarvioinnin validiustarkastelussa tulisi puuttua. (Messick 1995, 741–742.)

Messick esitti testin validiuteen liittyvää kritiikkiä jo varhaisessa vaiheessa. Hänen mukaansa joidenkin testien tai kokeiden heikko laatu saattoi johtaa siihen, että niitä käytettiin väärin tai virheellisesti (Messick 1980, 1012). Messickin kärkevin kritiikki kohdistui kuitenkin perinteiseen validiuskäsitykseen, jossa validius jaettiin eri ”tyyppiseen validiuteen” tai ”validiuden eri lajeihin” sisältövalidiuden, kriteerivalidiuden ja käsitevalidiuden kesken. Messickin mukaan perinteinen validiuden tarkastelutapa ei ollut riittävää minkään asian testaamiseksi ainoastaan sisältövalidiudesta tai kriteerivalidiudesta käsin. (Messick 1989 6–7.) Lisäksi Messick piti kyseistä jaottelua hajanaisena sillä, siinä ei otettu huomioon näyttöä, jonka mukaan arvot vaikuttavat testeistä saatujen tulosten merkitykseen ja sitä kautta ihmisten toimintaan. Vanha jaottelutapa ei myöskään ottanut huomioon sitä, minkälaisia sosiaalisia seurauksia tulosten hyödyntämisellä mahdollisesti oli. Sen sijaan yhtenäinen validius-käsite, jossa validiutta tarkastellaan eri näkökulmista, nivoo toisiinsa käsitevalidiuden perusnäkökohdat eli testituloksen merkitykseen ja tulkintaan liittyvät

64 Testien validiustarkastelua ei ole tarkoituksenmukaista irrottaa erilliseksi osa-alueekseen ja siksi jo aiemmin tämän tutkimuksen luvussa 4.2 on osana modernin validiusteorian muotoutumista ja validointia käsitelty testitulosten ja niistä tehtävien tulkintojen validiutta.

sosiaaliset arvot sekä testin hyödyntämisen. (Messick 1989b, 5–7; Messick 1995, 741.) Messickin mukaan kasvatustieteissä ja psykologiassa arvioinnin validius, reli-aabelius, vertailtavuus ja oikeudenmukaisuus eivät ole ainoastaan mittaamiselle asetettuja periaatteita, vaan ne ovat myös sosiaalisia arvoja, jotka ovat tärkeitä aina, kun arvioinnin perusteella tehdään päätöksiä (Messick 1995, 742).

Kuten jo aiemmin luvussa 4.2 on esitetty, Messick määritteli validiutta niin, että siinä yhdistyy arvio, ”missä määrin empiirinen näyttö ja teoreettiset ajattelutavat tukevat riittävästi ja tarkoituksenmukaisesti testitulosten tai muiden arviointitulosten perusteella laadittuja johtopäätöksiä ja toimintaa” (Messick 1980, 1023; 1989a, 13; 1989b, 5; 1995, 741). Messick määritteli validiutta myös siten, että ” käsitevalidius yhdistää kriteerit ja sisällön yhteiseksi viitekehyykseksi, jossa testataan järkeviä hypoteeseja suhteessa niiden olennaisiin, teoreettisiin yhteyksiin.” Messickin mukaan validiutta voidaan lähestyä ns. peruskysymyksellä siitä, ” missä määrin, empiiriseen näyttöön tai teoreettiseen ajatusmalliin perustuvia testituloksia tulisi voida tulkita ja hyödyntää ”. (Messick 1989b, 5.) Validius ei näin ollen liity ainoastaan testien tai arviointien ominaisuuteen, vaan siinä on kyse myös testitulosten merkityksestä (Messick 1995, 741). Yhtenäisen validiuskäsityksen ydin on tarkoituksenmukaisuuteen, mielekkyyteen ja hyödyllisyyteen perustuva päättely, joka pohjautuu luotettavaan ja empiirisesti perusteltuun tulkintaan testin tuloksista (Messick 1989b, 5).

Messick esitti, että tulosten validointi on empiiristä arviointia mittauksen merkityksestä ja seurauksista. Validoinnissa yhdistyvät tieteellinen tutkimus järkeviin perusteluihin, joilla puolletaan tai mitätöidään testitulosten tulkintaa tai käyttöä. (Messick 1995, 742.) Se mitä validoidaan ei ole testi tai havaintotyökalu sinänsä, vaan testituloksiin perustuva tulkinta, johtopäätös tai toiminta. Kun tulkitsevaa päättelyä validoidaan, selvitetään missä määrin näyttö on sopusoinnussa päättelyn kanssa ja vahvistetaan, että näyttö tukee huomommin vaihtoehtoisia tulkintoja. Testin validoinnissa tarvitaan sekä samaan suuntaan että eri suuntaan osoittavaa näyttöä. Kun validoidaan toimintaa, validoidaan päätelmiä siitä, mikä merkitys tuloksella on ja mikä sen painoarvon on suhteessa tulosten perusteella tehtäviin toimenpiteisiin. Messick korostaa, että erityisesti silloin, kun testitulosta käytetään päätöksenteon perusteena, validoinnissa tulee kiinnittää erityistä huomiota testituloksen merkitykseen, hyödyllisyyteen ja sosiaalisiin seurauksiin. (Messick 1989b, 5.)

Koska Messick piti validiutta yleisenä arviona testitulosten perusteella tehtävien päätelmien riittävydestä ja tarkoituksenmukaisuudesta, perustuu testitulosten arviointi hänen mukaansa neljään kohtaan, jotka ovat 1) induktiivinen yhteenveto samaan suuntaan tai eri suuntaan osoittavasta tutkimusnäytöstä sen suhteen onko testin tulos tulkittavissa erityisesti jonkin ajatusmallin tai käsitteen näkökulmasta, 2) testituloksen painoarvon ja testituloksen tulkinnan arviointi, 3) ajatusmallin tai käsitteen merkityksen perustelu ja näyttö sekä tulosten käyttö tietyissä tilanteissa ja 4) testitulosten mahdollisten sosiaalisten seurausten arviointi suhteessa tosiasialisiin seurauksiin. (Messick 1980, 1023.)

Messick laati oman käsitevalidiuteen perustuvan mallinsa, eräänlaisen nelikentän, jonka avulla testin tai kokeen validiutta voidaan tarkastella (Messick 1980, 1023; Messick 1995, 748).

TAULUKKO 5. Messickin validiutta kuvaava nelikenttä. Lähde: Messick 1980, 1023; 1995, 748.

Perustelut	Testin tulkinta (Test Interpretation)	Testin käyttö (Test Use)
Näyttön perustuva (<i>Evidential Basis</i>)	Käsitevalidius (<i>Construct validity</i>)	Käsitevalidius + merkitys/ hyödyllisyys (<i>Construct validity</i>) + (<i>relevance/utility</i>)
Seurauksiin perustuva (<i>Consequential Basis</i>)	Arvojen vaikutus + Käsitevalidius ⁶⁵ (<i>Value implications</i>) + (<i>Construct validity</i>)*	Sosiaaliset vaikutukset + Arvojen vaikutus* (<i>Social consequences</i>) + (<i>Value implications</i>)*

Messickin mukaan testin validiuteen kytkeytyi kaksi puolta, jotka ovat perustelut ja testaamisen tarkoitus tai testitulokset. Perustelut ovat joko näyttöön perustuvia (*evidential basis*) tai todennäköisiin seurauksiin (*consequential basis*) tukeutuvia. Testauksen tarkoitukseen tai testitulokseen kuuluu puolestaan testituloksen tulkinta (*test interpretation*) ja testin käyttö (*test use*). (Messick 1980, 1023; 1995, 748.) Kuten taulukosta 5 ilmenee, nelikentässä on kuvattu sekä testaamisen merkitystä ja arvoja että testitulosten tulkintaa ja testin käyttöä. Nelikentän avulla Messick on kuvannut sitä, miten näyttöön tai todennäköisiin seurauksiin perustuva puoli liittyy luokittelun toisen puolen eli testin tulkinnan tai testin käytön kanssa. Taulukossa 5. kuvattujen neljän solun avulla voidaan Messickin mukaan vastata validiuden peruskysymykseen eli siihen, missä määrin empiiriseen näyttöön tai todennäköisiin seurauksiin perustuvia testituloksia voidaan tulkita ja käyttää. Nelikentän keskeinen ominaisuus on se, että käsitevalidius esiintyy jokaisessa solussa, koska ”testituloksen tarkastelu käsitevalidiuden näkökulmasta on se yhdistävä voima, joka kokoaa kaikki validiuteen liittyvät seikat yhtenäiseksi käsitteeksi”. (Messick 1995, 748.) Nelikentän avulla Messick siis pyrki kuvaamaan myös käsitevalidiuden ykseyttä. Käsitevalidiuden lisäksi soluissa on kuvattu myös muita ominaisuuksia, jotka nousevat esiin testaamiseen liittyvien perustelujen ja testitoiminnan pohjalta. (Messick 1995, 748–749).

65 Messickin nelikenttään on merkitty tähdellä ne lisäykset, jotka hän teki myöhemmin (vrt. Messick 1980, 1023; Messick 1995, 748).

Messickin mukaan nelikentän yksittäisiä soluja voidaan tarkastella erikseen. Ensimmäisessä solussa on esitetty, että käsitevalidius on näyttöön perustuvaa testin tulkintaa. Se edellyttää sekä samaan suuntaan osoittavaa (*convergent*) että erottelevaa (*discriminant*) näyttöä, jossa on kuvattu esimerkiksi teoreettisesti olennainen empiirinen yhteys tiettyä käsitettä mittaavan testin ja jonkun toisen tätä samaa käsitettä mittaavan menetelmän välillä. (Messick 1980, 1019; 1995, 748.) Toisessa solussa kuvataan näyttöön perustuvaa testin käyttöä, joka on käsitevalidiuden tarkastelua, mutta sillä ehdolla, että yleinen näyttö tukee testituloksen merkitystä. Testitulokset saavat erityisen merkityksen, kun niitä sovelletaan johonkin tiettyyn tarkoitukseen tai tietyssä yhteydessä. Käytännön esimerkkinä voisi olla tilanne, jossa pyritään arvioimaan testauksen hyödyllisyyttä suhteessa testistä aiheutuviin kustannuksiin. (Messick 1980, 1019.)

Kolmannessa solussa on kuvattu sitä, miten seurauksiin perustuva testin tulkinta on testituloksista tehtävien päätelmien arviointia. Messickin mukaan on keskeistä, että teoreettiset ja arvottavat päätelmät ovat keskenään yhteismitallisia, koska ne liittyvät erottamattomasti toisiinsa. Testituloksen tulkintaa tarvitaan arvottavien johtopäätösten arvioimiseksi ja päinvastoin. Siksi sellainen testituloksen tulkinta, joka perustuu todennäköisiin seurauksiin, tulee käsittää sekä käsitevalidiutena että seurauksena, joka syntyy, kun testituloksen merkitystä arvioidaan. Vaikka monet tutkijat kiistelevät arvojen ja seurausten mukaan ottamisesta validiustarkasteluun, Messickin katsoo, että arvot liittyvät luontaisesti testitulosten arviointiin, sillä ne antavat testituloksille merkityksen. Validiuden arviointi on arvojen arviointia. (Messick 1980, 1021; 1995, 748.) Testin validius perustuu siis myös siihen, kuinka testituloksia käytetään (Messick 1980, 1025). Arvot ovat validiustarkastelun ulottuvuus, joka voidaan liittää kaikkeen psykologian tai kasvatustieteen mittaamiseen. Messick perustelee näkemystään sillä, että psykologiassa ja kasvatustieteissä muuttujat tuottavat joko suoraan tai epäsuorasti tietoa ihmisen ominaisuuksista. Aina kun testiä suunnitellaan, analysoidaan, tulkitaan tai hyödynnetään, on kyse myös siitä, kenen lähtökohdista käsin arvotetaan, mitataan tai määritellään. Toinen keskeinen arvioihin liittyvä kysymys on se, mitä seurauksia negatiivisilla testituloksilla on. Messick pitää arvoja kaikkeen mittaamiseen, niin suhteelliseen, kriteeriperusteiseen kuin luokassa suoritettavaan havainnointiin, olennaisesti kuuluvana ulottuvuutena. (Messick 1980, 1013; 1995 748.)

Neljännessä solussa keskitytään tarkastelemaan sitä, mikä on tulkinnasta tehtävien arvottavien johtopäätösten sosiaalinen tai yhteiskunnallinen merkitys, sillä tulosten ja niistä tehtävien päätelmien perusteella käynnistetään usein jokin toimintatapa. Testitulokseen perustuva toiminta yhdistää käsitteen mittaamisen käytännössä sovellettaviin kysymyksiin ja tätä kautta sosiaalipolitiikkaan. Neljännessä solussa esitetty testin käyttö ja sen seurausten lähtökohdat ovat arvioinnin mahdollisten ja todellisten sosiaalisten seurausten testaamista ja testitulosten soveltamista. Messickin mukaan viime kädessä testin todellinen arvo mitataan siinä, kuinka hyvin testi palvelee tuloksia hyödyntäviä tahoja, kuten päätöksentekijöitä tai poliitikkoja. (Messick 1980, 1020–1021; 1995, 748.)

Kane lähestyy validiutta argumentoinnin työkaluna (Brennan 2006, 2; Kane 2006, 23) ja hänen muotoilunsa validiusargumentista tarjoaa yleisarvion testitulosten tulkinnasta aina testin hyödyntämiseen. Yksittäisen tutkimuksen validiustarkastelu voi Kanen mukaan perustua testitulosten käyttöön, sen tulkintaan, tilastollisiin analyyseihin tai sisällön tai kriteereiden analysointiin, mutta validiusargumentointi vaatii kokonaisuudessaan erilaisten ja eri lähteistä peräisin olevien, havaintoihin perustuvien tietojen yhdistämistä (Kane 2006, 23). Validiusargumentointi perustuu siihen, ovatko sen tulkinta-argumentit valideja. Tulkinta-argumentti luo perustan, kun valitaan näyttöä validiusargumentointiin (Kane 1990, 10).

Kane korostaa, että validiusargumentoinnissa on välttämätöntä olla olemassa selkeät ja valmiit väittämät, joissa on otettu huomioon, miten ja missä yhteydessä testiä ja sen tuloksia aiotaan käyttää. Testin käyttöä ja niiden tulkintaa koskevat ehdotukset voivat olla yksityiskohtaisia ja ne voivat muodostaa oletusten ja tulkintojen ”ryppään” aina testin suorittamisesta siitä saatavien tulosten tulkintoihin tai testitulosten perusteella tehtäviin johtopäätöksiin asti. Johtopäätösten tulee olla yksityiskohtaisia ja johtopäätöksiä tukevia oletuksia tulee arvioida tarkoituksenmukaisella näytöllä ja uskottavilla, vaihtoehtoisilla tulkinnoilla. Yleensä testien tulokset ja niiden käyttö ovat kontekstisidonnaisia ja siksi myös yksityiskohtaiset johtopäätökset ja oletukset saattavat muuttua asiayhteyden tai testin käytön myötä. Validoinnissa tulkinta-argumenttien avulla voidaan tehdä tarkennuksia oletuksiin ja validiusargumenttien avulla voidaan puolestaan arvioida esitettyjä tulkintoja sekä testituloksen käyttöä. (Kane 2001, 322.)

Kanen mukaan validiuden käsitteeseen liittyy heikkouksia. Hän pitää validiuteen kuuluvana puutteena esimerkiksi sitä, ettei ole olemassa selviä ohjeita tai suuntaviivoja, joita voitaisiin käyttää valittaessa näyttöä, jota käytetään testin tuloksen validointiin. Suuntaviivojen puute saattaa aiheuttaa sen, että validointia tuetaan vain sellaisella näytöllä, joka on helposti tai sopivasti saatavilla. (Kane 1990, 6.) Kyseisen puutteen korjaamiseksi Kane esittää, että testituloksen tulkintaa analysoidaan suhteessa tulkinta-argumentteihin (Kane 2006, 23).

Testituloksia määrittelevät tulkinnat ovat yleensä seurausta erilaisten johtopäätösten muodostamasta ”ryppästä”, johon kuuluu muun muassa yleistyksiä, ekstrapolointia, oletuksia, selityksiä, teorioista johdettuja johtopäätöksiä ja testituloksiin perustuvia päätöksiä. Lähes kaikki testitulosten tulkinnat sisältävät elementtejä yleistyksistä tarkkoihin havaintoihin. (Kane 1990, 15.) Kaikki tulkinta edellyttää ainakin implisiittisesti teoriaan perustuvia johtopäätöksiä, joihin voi sisältyä selityksiä tai yhteyksiä toisista ajatusmalleista. Mikäli testitulokset eivät ole olennaisia päätöksenteoksen kannalta, ei silloin myöskään ole selvää, miksi testi tehdään. Testin käytön oikeutus nojaa oletukseen mahdollisista tuloksista ja niiden perusteella tehtävistä päätöksistä (Kane 1990, 16; ks. myös Messick 1988).

Jokaisesta tulkinta-argumentista tehtävä johtopäätös perustuu oletukseen, että johtopäätökset perustellaan. Validoinnin näkökulmasta oletukset ovat tulkinta-argumenteissa avainasemassa. Yleensä tulkinta-argumenttien puutteet liittyvät virheelisiin tai puutteellisiin oletuksiin, pikemminkin kuin puutteelliseen logiikkaan; loogiset virheet ovat yleensä helpompia määritellä ja korjata kuin heikkoudet oletuksissa, etenkin jos tulkinta-argumentti ei ole selkeä. Tulkinta-argumentteja ei voi eritellä matemaattisella tarkkuudella, vaan ne määritellään yleisin ehdoin. Koska tulkinta-argumentit ovat yleensä monimutkaisia ja ne saattavat sisältää monenlaisia oletuksia, tulkinta-argumentteja on vaikea määritellä selkeästi ja niiden tehokas arviointi on myös hankalaa. Erityisen ongelmallisina Kane pitää piiloon jääviä oletuksia, jotka eivät ole selvästi tunnistettavissa osana argumenttia. (Kane 1990, 17; 2006, 26.) Jokainen tulkinta-argumentti on Kanen mukaan uniikki, samoin näyttö, joka tukee tulkinta-argumenttia ja validiusargumenttia, on ainutkertaista (Kane 1990, 34).

Kun tulkinta-argumentteja analysoidaan, niistä tulee käydä selvästi ilmi, että testit ja niiden painotukset vastaavat todellisuutta ja siinä esiintyviä arvoja (Cronbach 1988, 5). Kane esitti, että sopiva peruste keskustella tulkinta-argumenttien rakenteesta on keskittyä sellaisiin päätelmiin ja perusteluihin, jotka esiintyvät yleisesti testitulosten tulkinnoissa. Brennanin mukaan Kanen käsitys validiudesta muistuttaa pitkälti ohjelma-arviointia (Brennan 2006, 2), jossa arvioidaan ja tehdään johtopäätöksiä arvioitavan ohjelman, kuten esimerkiksi opetussuunnitelman, saavutuksista, merkityksestä ja arvosta (Heinonen 2001a, 56).

Validiusargumentti tarjoaa perustelun hyväksyä tulkinta-argumentti ja sitä kautta hyväksyä myös esitetty tulkinta. Validiusargumenttina voidaan käyttää empiiristä aineistoa, edellisten tutkimusten tuloksia ja erilaisia perusteluja, joilla tuetaan validiusargumentoinnin eri osia. Jos validiusargumentti tukee tulkinta-argumenttia, sen tulee heijastaa tulkinta-argumentin rakennetta. Paras tapa tarkistaa tulkinta-argumentti on, että validiusargumentti keskittyy siihen osaan tulkinta-argumenttia, joka on kaikkein kyseenalaisin ja ongelmallisin. (Kane 1990, 18; Kane 2006, 26.) Tulkinta-argumentteja haastetaan ja koetellaan siten, että niiden rinnalle pyritään esittämään kilpailevia ja vielä uskottavampia argumentteja (Kane 2006, 29).

Jo aiemmin Cronbach (1988) ja Messick (1980; 1989a; 1989b; 1995) esittivät, että testin käytön seurauksia tulee voida arvioida. Heidän mielestään validiusargumentointi tarjoaa kokonaisarviointiin testi- tai tutkimustulosten tulkinnasta ja niiden hyödyntämisestä tuottamalla yhdenmukaisen analyysin koko kerätystä aineistosta. Myös Kanen mukaan seurausten arviointi on olennainen osa testin käytön validoinnissa, joskin seuraukset voivat olla kauaskantoisia, eikä niitä ole aina helppo määrittää (Kane 2001, 336). Jotta testitulosten perusteella muodostettua tulkintaa voidaan arvioida, on Kanen mukaan välttämätöntä, että testin tulkinta ja käyttötarkoitus ilmaistaan selkeästi (Kane 2001, 329; Kane 2006, 22). Seurausten merkityksen tarkastelemiseksi validoinnissa, Kane jakaa tulkinta-argumentit kahteen osaan, joista

ensimmäinen on kuvaava (*descriptive part*) ja toinen on ohjaileva (*prescriptive part*) osa. Kuvaavassa osassa argumentteihin kuuluu päätelmien ”rypäs”, joka muodostuu testituloksista aina yksilöitä kuvaileviin lausuntoihin. Ohjaileva osa muodostuu päätöksistä, joita tehdään kuvaavan osan perusteella. (Kane 2006, 337.) Kanen mukaan testien ja niiden tulosten seurausten arvioimiseksi ei ole olemassa helppoa tapaa, vaan jokaisessa mittaustenmenettelyssä tulee arvioida niiden omia ansioita. Vaikka Kanen mukaan erilainen näyttö on olennaista eri arviointimenetelmien seurausten arvioimiseksi, pitää hän hyödyllisenä erottaa toisistaan ne tulkinta-argumentit, jotka vain kuvailevat, sellaisista tulkinta-argumenteista, jotka puoltavat testitulosten perusteella tehtäviä päätöksiä tai toimenpiteitä. Seurausten arvioinnissa tulisi hänen mielestään erottaa näyttö, jolla validoidaan erilaisia tulkinta-argumentteja, sillä päätösprosessien validointi on aina ollut riippuvaista päätösten seurausten arvioinnista. (Kane 2006, 338–339.)

4.3.2 Luokkatyöskentelyn arvioinnin validius

Luokkatyöskentelyn arvioinnissa on perinteisesti kiinnitetty vähän huomiota varsinaisten tulosten arviointiin, sillä suuri osa luokkatyöskentelyn arvioinnista perustuu laadulliseen arviointiin, jossa keskeistä on suorituksen tulkinta. Luokkatyöskentelyn arvioinnissa tavoitteena ei ole ollut luokitella havaintoja suhteessa empiirisiin lainalaisuuksiin, vaan havainnot on pyritty suhteuttamaan oikeaan kontekstiin ja ymmärrettävään taustaan. (Kane 2006, 47.) Luokassa tapahtuvan työskentelyn arviointia ja sen validiutta on alettu pohtia laajemmin vasta 2000-luvun alussa. Tois-taiseksi sitä ovat käsitelleet Brookhartin (2003; 2007) lisäksi muun muassa Kane (2006), Moss (2003), Smith (2003) ja Shepard (2006). He ovat esittäneet validiutta koskevan teorian uudelleenmuotoilemista niin, että sitä voitaisiin käyttää myös luokkatyöskentelyn arvioinnissa (Shepard 2006, 642).

Mittaan koskevan teorian pääperiaatteena on ollut, että validius ja reliiabiilius ovat riippuvaisia testin käyttötarkoituksesta. Luokkatyöskentelyn arviointi ja siinä validiuden ja reliiabiiliuden tarkastelu ovat myös kontekstisidonnaisia, sillä peruskoulun oppilaiden työskentelyn arviointi on erilaista kuin esimerkiksi korkeakoulun sisäänpääsykokeiden arviointi. (Shepard 2006, 641.) Shepard yhdistää luokkatyöskentelyn arvioinnin validius- ja reliiabiiliustarkastelussa laadullisia tutkimusmenetelmiä, formatiivisen arvioinnin käytänteitä sekä havainnointia alustavan olettamuksen muodostamiseksi. Luokkatyöskentelyn arvioinnissa validiuden ja reliiabiiliuden tarkastelun rajapinta on hämärä. Se, mitä Smith (2003) tutkimuksessaan kutsuu riittäväksi tiedoksi reliiabiiliuden näkökulmasta (Smith 2003, 30), vastaa Mossin (2003) käsitystä luokkatyöskentelyn arvioinnin validiuden tarkastelusta. (Shepard 2006, 642.)

Luokkatyöskentelyn arvioinnin validiuteen liittyvässä meta-analyysissä Brookhart on nostanut esiin tutkimuksissa esille tulleita haasteita, joista ensimmäinen koskee opettajien tietoja ja taitoja arvioinnin validiudesta ja reliabiliudesta. Tutkimukset (Impara, Plake & Fager 1993; Campbell & Evans 2000; Mertler 2000) osoittivat, että opettajilla ei ole käsitystä siitä, mitkä periaatteet ovat olennaisia arvioinnin validiuden ja reliabiliuden kannalta. Citzeikin, Fitzgeraldin ja Rachorin (1995) tutkimus osoitti, että luokkatyöskentelyn arviointiin liittyy validiusongelmia myös siksi, että opettajat voivat valita varsin vapaasti, millä arviointimenetelmillä he arvioivat kaikille yhteisten tavoitteiden, kuten esimerkiksi opetussuunnitelmassa asetettujen tavoitteiden, saavuttamista. Lisäksi tietyissä menetelmissä, kuten havainnoinnissa, jota käytetään luokkatyöskentelyn arvioimiseksi, on erityisiä arvioinnin validiuteen liittyviä ongelmia. Esimerkiksi havainnointiin perustuvaa oppilaan arviointia pidetään usein validina sen perusteella, että oppilaan arvioinnissa saama tulos on hyvä. Bulterman-Bosin ja muiden (2002) tutkimuksen mukaan havainnointiin liittyvä validiusongelma johtuu taas siitä, että opettajat eivät pysty erottamaan havainnointiprosessissa itseään havainnoijan roolista, vaan opettajalla on helposti kaksoisrooli sekä arvioijana että havainnoijana. Koska havainnointiprosessi on kaksisuuntaista, opettajat voivat olla erilaisia havainnointitilanteissa, sen mukaan minkälainen heidän vuorovaikutuksensa kulloinkin havainnoitavaan oppilaaseen on. Havainnointiin perustuvaa luokkatyöskentelyn arvioinnin validiutta voi heikentää se, miten havainnoinnilla saatuja tuloksia arvioidaan sekä se, miten opettajan mielikuvat aiempien oppilaiden toiminnasta vastaavassa tilanteessa vaikuttavat sen hetkisten oppilaiden arviointiin. (Brookhart 2007, 44–54.)

Shephardin mukaan formatiivinen arviointi on validia vain, jos se edistää oppilaan oppimista (Shepard 2006, 642). Moss esittää, että luokkatyöskentelyn arvioinnin validiudessa on kyse arvioinnin seurauksista sekä siitä, kuinka hyvin arvioinnilla tulkitaan ja annetaan tietoa ja ohjataan ja autetaan oppilasta lisäämään osaamistaan (Moss 2003, 19–20). Jotta validiusteoriaa voitaisiin kehittää palvelemaan luokkatyöskentelyn arviointia, sen tulisi kohdistua käsitteellisesti sekä summatiiviseen että formatiiviseen arviointiin. Validiustarkastelun tulisi kohdentua siihen, miten hyvin erilaiset arviointimenetelmät kuvaisivat oppilaiden tietoja, taitoja, asenteita ja identiteetin kehittymistä summatiivisessa arvioinnissa. Samalla validiustarkastelun tulisi kyetä tukemaan formatiivista arviointia ja siinä erityisesti syvää ja laajaa asioiden osaamista ja hallintaa. Validiustarkastelun tulisi nimenomaan vastata kysymykseen, toimiiko arviointiprosessi niin kuin sen on ajateltu toimivan. Jonkin verran enemmän tietoa luokkatyöskentelyn arvioinnin validiudesta on saatu viimeaikaisista tutkimuksista, joissa on tutkittu formatiivista arviointia yhdessä oppimisteorioiden kanssa. (Shephard 2006, 642.)

Kanen (2006) mukaan luokkatyöskentelyä voidaan arvioida tulkinta-argumentin avulla. Siinä opettajalle muodostuu hänen omien kokemustensa ja ammatillisen viitekehyksen perusteella alustava käsitys oppilaasta. Ammatillinen viitekehys muodostuu Kanen mukaan opettajan pedagogisista näkemyksistä, opetettavan sisällön ja opetusmenetelmien hallinnasta, sekä opettajan kyvystä käyttää luokkatyöskentelyn

arvioinnissa omaa ammattitaitoaan, jonka avulla hän arvioi oppilaidensa suorituksia. Opettajan ja oppilaan jatkuvan vuorovaikutuksen seurauksena opettaja saattaa muokata näkemystään oppilaasta. Opettajalle voi muodostua vuorovaikutuksen myötä odotuksia oppilaan suorituksista, joita Kane kutsuu työhypoteeseiksi. Opettaja antaa oppilaalle ohjeistuksia ja laatii oletuksia oppilaan arvioimiseksi laatimiensa työhypoteesien perusteella. Opettaja saa työhypoteeseihin perustuen palautetta siitä, kuinka oikeaan hänen näkemyksensä oppilaan suorituksesta oli osunut. Uudessa tilanteessa opettaja voi myös arvioida oppilastaan kokonaisvaltaisesti, kun hän joutuu arvioimaan oppilaan suoritusta. (Kane 2006, 47–48.)

Kuten validiusargumentoinnissa yleensä, tulkinta-argumentin johdonmukaisuus ja paikkansapitävyys tulee tutkia. Sen jälkeen selvitetään, voidaanko tulkinta-argumentin avulla laatia järkevä selitys, ja jos voidaan, niin selityksen perusteella laaditut johtopäätökset ja oletukset arvioidaan. Mikäli tulkinta-argumenttia voidaan kehittää edellä kuvatun toimintamallin mukaisesti, validiusargumentointi keskittyy tarkastelemaan opettajan viitekehyksen uskottavuutta ja opettajan näkemyksiä oppilaasta.

Toulminin argumentaatiomallin mukaan, jota käsitellään perusteellisemmin luvussa 4.4, opettajan viitekehys muodostaan perusteen (*warrant*) tulkitseville arvioinneille. Peruste saa puolestaan tukea (*backing*) paitsi opettajan viitekehyksestä, myös opettajan ammatillisista kokemuksista ja siitä, missä määrin opettajalla on mahdollisuus käyttää muuta tietoa, jolla voisi olla merkitystä luokkatyöskentelyn arvioinnissa. Kanen mukaan opettamiseen luonteeseen kuuluu jatkuva ”vuorovaikutuksen ja päätöstenteon virta” ja siksi suuri osa opettajan johtopäätöstensä tueksi käyttämästä tiedosta ei ole näkyvää. Opettaja ei myöskään tallenna kaikkia havaintojensa pohjalta tekemiä perusteluja, eikä erittele kaikkia niitä johtopäätöksiä, jotka johtavat lopullisten arviointipäätösten tekemiseen. Kuitenkin, jos opettajaa pyydetäisiin tekemään niin, hän luultavasti pystyisi antamaan perustelut tekemilleen valinnoille. (Kane 2006, 48–49.) Kuten kaikkien oletettujen päätelmien tekeminen, myös opettajan arviointi voidaan haastaa ja asettaa koeteltavaksi. Oppilaasta saatu tieto tai ympäröivät olosuhteet tai näiden kahden välinen vuorovaikutus voivat muuttaa opettajan tekemää tulkintaa. (Kane 2006, 50.)

4.3.3 Arviointiin liittyvä epävalidius

Cookin ja Campbellin (1979) mukaan validiustarkastelussa on olennaista pohtia systemaattisesti niitä tekijöitä, jotka saattavat alentaa mittauksen luotettavuutta. Kanen mukaan Cook ja Campbell (1979) esittivät arvioinnin uskottavuuden uhkina aliedustettavuuden (*under-representation*) tai asiaan kuulumattoman vaihtelun (*irrelevant variance*). Jonkin ominaisuuden mittaaminen on aliedustettua, jos se ei onnistu kuvaamaan riittävän edustavasti ominaisuuksiin liittyviä erilaisia havaintoja. (Kane 2006, 38.)

Ekstrapoloinnilla eli tilastollisen yhteyden pohjalta tehtävällä päättelyllä pyritään kuroma umpeen yleistettävien havaintojen ja kohderyhmästä tehtävien havaintojen välistä eroa osoittamalla, että aliedustavuudella ei ole vaikutusta tulokseen (Kane 2006, 38). Sen lisäksi, että otos on mahdollisimman laaja, voidaan aliedustuksen mahdollisuutta vähentää määrittelemällä käsitteet ja mittausprosessi, kuten mittauksen tehtävä ja mittaustilanne, mahdollisimman laajasti. Cook ja Campbell (1979) sekä Messick pitivät monien eri mittausmenetelmien tai useiden tehtävien käyttämistä keinona vähentää arvioinnin epävalidiutta. (Messick 1989a, 35; Kane 2006, 38.) Tietyn ominaisuuden mittaamiseen kuulumaton vaihtelu johtuu yleensä tekijöistä, jotka eivät sinänsä liity mitattavaan ominaisuuteen, kuten mittaamiseen liittyviin menettelyihin tai tulokseen vaikuttaviin havainnointiolosuhteisiin (Kane 2006, 39).

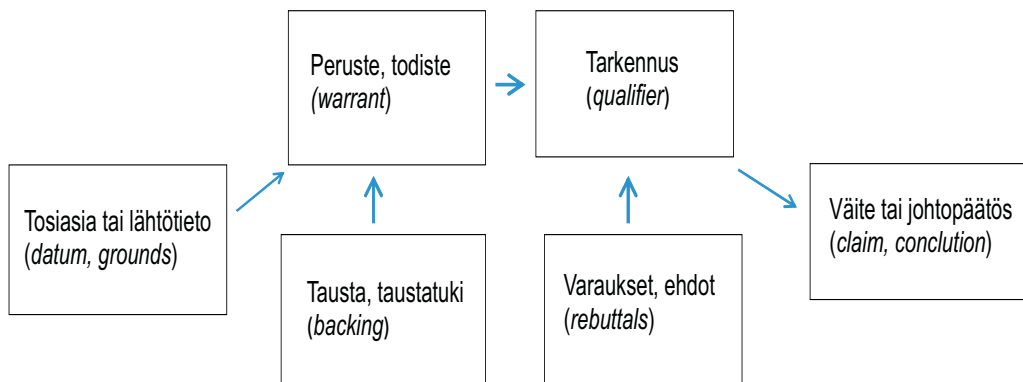
Vaikka Messick piti edellä mainittua aliedustusta ja asiaan kuulumatonta vaihtelua kaikkien arviointien validiustarkasteluihin liittyvinä uhkina, hän kutsui niitä rakenteen aliedustavuudeksi (*construct underrepresentation*) ja rakenteeseen kuulumattomaksi vaihteluksi (*construct-irrelevant variance*) (Messick 1989b, 4). Aliedustavuudella tarkoitetaan, että arviointi on liian kapea-alaista ja siksi jokin käsitevalidiuden ulottuvuus jää ottamatta huomioon (Messick 1989b, 4; 1995, 742–743). Rakenteeseen kuulumaton vaihtelu, tarkoittaa puolestaan sitä, että arviointi on liian laajaa ja siksi se sisältää ylimääräistä, joskin reliaabelia vaihtelua (Messick 1989a, 34; 1989b, 4; 1995, 742–743). Tästä syystä satunnaisesta virheestä johtuvaa vaihtelua ei lueta rakenteeseen kuulumattomaksi vaihteluksi. Satunnaista virhettä kontrolloidaan yleensä standardoimalla mittausprosessi ja keräämällä laajalla otoksella arvioitava aineisto. Sen sijaan systemaattisesta virheestä johtuva asiaan kuulumaton vaihtelu tulee käsittää niin, että se toimii erityisenä virhelähteenä. (Kane 2006, 39.)

Rakenteeseen kuulumatonta vaihtelua voi esiintyä, koska annettu tehtävä on aivan liian vaikea suoritettavaksi. Messick nimittää tätä rakenteeseen kuulumattomaksi vaikeudeksi (*construct-irrelevant difficulty*). Kyseinen epävalidius johtaa tuloksiin, jotka ovat aivan liian matalia. Eri ryhmien tai yksittäisten henkilöiden testitulosten tulkinnoissa ja tulosten hyödyntämisessä rakenteeseen kuulumaton vaikeus on huomattava puolueellisuuden ja epäoikeudenmukaisuuden lähde. (Messick 1995, 742–743.) Tehtävät voivat olla myös liian helppoja ja tästä epävalidiuden lähteestä Messick käyttää nimitystä (*construct-irrelevant easiness*) eli rakenteeseen kuulumaton helppous (Messick 1995, 742). Rakenteeseen kuulumatonta helppoutta esiintyy silloin, kun tehtävissä on ylimääräisiä vihjeitä, jonka perusteella vastaaja saattaa vastata oikein tai aineisto tai tehtävä on vastaajalle jo ennalta tuttua. Rakenteeseen kuulumaton helppous johtaa puolestaan tuloksiin, jotka ovat huomattavan hyviä. (Messick 1995, 743.)

4.4 Toulminin argumentaatiomalli ja validiusargumentointi

Toulmin julkaisi vuonna 1958 argumenttien analysoimiseksi teoreettisen mallin siihen kuuluvine terminologioineen. Toulminin päättelymallia on käytetty laajasti aina ohjelma-arvioinneista mittaamisteorioihin (Kane 2006, 27). Argumentin anatomiana tunnetun päättelymallin lähtökohta on nimenomaan argumentin sisällyksessä, ei sen muodossa. Toulminille argumentointi tarkoitti päättelyä ja sen pätevyyttä. (Toulmin 2003, 87.) Hänen mukaansa väitteen (*claim*) tukeminen perusteilla (*warrant*), esimerkiksi silloin kun johtopäätös tehdään tulkinta-argumenttiin perustuen, on velvoite. Lisäksi väitteen tulee olla mahdollisimman hyvin perusteltu. (Toulmin 2003, 90.) Toulminin mukaan hänen argumentointimallissaan ei ole kyse siitä, että väitteen tueksi esitetään lisää tosiasiallista, faktuaalista, tietoa, vaan kyse on siitä, että osoitetaan, mihin johtopäätös tai väite perustuu.

Toulminin mukaan esitettäessä tosiasioita tai lähtötietoja väitteen tai johtopäätöksen tueksi, johtaa se argumentoinnin vaiheittaiseen etenemiseen, joka ei perustu tiedon lisäämiseen, vaan siihen, että jokainen esitetty väite tai johtopäätös saa vahvistusta tai tukea olemassa olevista säännöistä, laeista tai yleispätevästä päättelystä. Peruste voi olla myös yleinen periaate, joka ilmenee argumentin tulkinnan tai vastaargumentin kautta. Perusteet tekevät lähtötiedoista merkityksellisiä suhteessa johtopäätöksiin, ja ne toimivat ikään kuin siirtymävaiheena, kun edetään lähtötiedosta väitteeseen tai johtopäätökseen. (Toulmin 2003, 91.) Kuviossa 3 on havainnollistettu edellä esitetty Toulminin päättelymalli ja sen elementit.



KUVIO 3. Toulminin päättelymalli ja sen elementit.

Toulminin mallissa päättelyketju alkaa tosiasiaista tai lähtötiedosta (*datum*), jonka perusteella muodostetaan väite (*claim*), joka voi olla myös loppupäätelmä (*conclusion*) (Toulmin 2003, 90–91). Vaikka Toulminin mukaan argumentaatio on alakoh- taista, eri alojen argumentaatiota yhdistää se, että kaikkia johtopäätöksiä voidaan analysoida edellä esitetyn kaavan mukaisesti (Toulmin 2003, 92; Kane 2006, 28). Toulminin mukaan hänen esittämänsä argumentaatiomalli pätee kaikessa argu- mentoinnista, riippumatta siitä, mihin aihepiiriin tai alaan argumentointi liittyy (Toulmin 2003, 94).

Päätelyyn voi kuulua yleistyksiä tai ennusteiden ulottamista testitulosten perus- teella tehtyjen havaintojen ulkopuolelle sekä tulosten selittämistä (Toulmin 2003, 93). Kvantitatiivisia argumentteja tulkittaessa tosiasiaa tai lähtötietoa vastaa esi- merkiksi jonkin tietyn tehtävän suoritustaso. Tällöin väite tai johtopäätökset voi- vat muodostua esimerkiksi pisteityksestä. Jokainen tulkinta-argumentti voi liittyä eri määrään päätelmiä johtopäätöksineen ja väitteineen. (Kane 2006, 27.) Vaikka ketju, tosiasia (*datum*) – esitetty perustelu tai todiste (*warrant*) – väite tai johtopäätös (*claim, conclusion*), on hyvin yleinen, perustelut tai todisteet voivat olla hyvinkin erityisiä tai täsmällisiä. Tämän vuoksi testituloksia voidaan tulkita erilaisista todis- teista tai perusteluista käsin. Todisteet väitteille eivät ole itsestäänselvyksiä ja siksi ne tulee aina perustella. Pisteityssäännöksi (*scoring rule*) kutsutaan todisteita, jotka muodostetaan pisteityksestä tehtävän päättelyn perusteella. Esitettyjen perusteiden yleistettävyyden periaatteen mukaan laaditut todisteet liittyvät havaittujen tulosten tilastollisiin yleistettävyyksiin, kuten esimerkiksi otoksen keskiarvoon. Tilastollisen yhteyden pohjalta voidaan ennustaa esimerkiksi oppilaan oletettua suoritustasoa kurssilla, jolloin todiste muodostettavalle väitteelle voi olla esimerkiksi regressioyh- tälö (*regression equation*). (Kane 2006, 28.)

Toulminin päättelymallissa empiiristen havaintojen tulee tukea todisteita. Yleensä pisteitystä perustellaan asiantuntijaryhmän tekemällä arviolla pisteityksen soveltu- vuudesta. Lähtötiedon yleistettävyyteen kuuluu, että otos on riittävän laaja ja edus- tava, sekä se, että tehdyt havainnot ovat otoksen perusteella yleistettävissä. Testitu- losten suhteen tehty empiirinen tutkimus tarjoaa regressioyhtälön kautta mahdol- lisuuden todisteisiin, jonka perusteella muodostettava väitettä voidaan perustella. Päätöstä koskevan säännön (*decision rule*) arvioimiseksi laaditaan yleensä analyysi siitä, minkälaisia seurauksia, myönteisiä tai kielteisiä, tehtävällä päätöksellä voidaan olettaa olevan. (Kane 2006, 28.)

Toulmin lisäsi päättelymalliin myöhemmin vielä muutaman lisäelementin, jolloin hänen lopullinen päättelymallinsa noudattaa kuviossa 3 esitettyä muotoa, jonka elementit ovat 1) lähtötiedot (*grounds, data*), 2) väite tai johtopäätös (*claim*), 3) peruste (*warrant*), 4) varaukset (*rebuttals*), 5) tarkennus (*qualifier*) ja 6) taustatuki (*backing*). Toulminin argumentaatiomalliin tekemä tarkennus liittyi ennen kaik- kea väitteen tai johtopäätöksen vahvuuteen, kun taas olosuhteiden muodostamat varaukset (*conditions of rebuttal*) ilmaisivat olosuhteita, joissa perusteita ei voinut

soveltaa. (Toulmin 2003, 94; Kane 2006, 28.) Validiuden näkökulmasta tarkasteltuna poikkeukset voivat muodostaa joissain tilanteissa tai joidenkin asioiden suhteen sellaisia tulkinta-argumentteja, jotka eivät ole perusteltuja (Kane 2006, 28).

Kuten jo aiemmin tämän tutkimuksen luvuissa 4.2 ja 4.3 on esitetty Cronbachin (1988) ja Kanen (1990; 2001; 2006) mukaan validiutta voidaan hyödyntää myös argumentaatiossa. Validiusargumentoinnin tarkoituksena on tarjota esitettyjen tulkintojen avulla kokonaisarviointi havaintojen puolesta ja niitä vastaan. Mikäli tulkinta-argumentit perustuvat testitulosten perusteella tehtäviin päätelmiin tai päätöksiin, validiusargumentointi voi tarjota kriittisen ulottuvuuden perusteluihin ja antaa taustatukea niiden perusteella tehtäviin päätelmiin. (Kane 2006, 28.)

Väite, jonka mukaan esitetty tulkinta tai tulkinnan hyödyntäminen on validia, sisältää automaattisesti väitteen myös siitä, että sen perusteella tehty tulkinta-argumentti on validia. Tulkinta-argumentin tulee olla ymmärrettävä, rationaalinen ja sen perusteella tehtävien oletusten tulee olla uskottavia. Analyysin, johon tulkinta-argumentti perustuu, tulee olla vakuuttavaa, ja siitä tulee ilmetä selvästi, että todellisuuden rakentaminen ja eri arvojen painotukset perustuvat testiin ja sen tulkintaan. (Cronbach 1988, 5.) Kanen lisäksi Toulmin (1958) on tarkastellut validiutta argumentoinnin näkökulmasta. Luvussa 6.4 koetellaan Toulminin argumentaatiomallin mukaan muodostettuja väitteitä: arvosanat ovat validi mittari ja arvosanat eivät ole validi mittari. Validiusargumentoinnin periaatteen mukaan näitä kahta väitettä koetellaan ja perustellaan tutkimuksen empiirisellä aineistolla sekä tuodaan mahdollisesti kilpailevia oletuksia väitteiden testaamiseksi.

Kanen mukaan käsitevalidiuden tarkastelun periaatteet sopivat myös validiusargumentointiin. Argumentteihin perustuva validiuden tarkastelutapa edellyttää, että esitetyt tulkinnat ovat tarkkaan määriteltyjä ja nämä tulkinnat ja oletukset arvioidaan olemassa oleviin havaintoihin perustuen. (Kane 2001, 339–340.) Tulkinta-argumentit ovat sellaisia, jotka perustuvat havainnoista ja päätelmistä tehtyihin johtopäätöksiin ja päätöksiin. Argumenttien tulkinnan tulee perustua ”jos-niin”-sääntöön. Sillä tarkoitetaan, että mikäli jollain havainnolla on tiettyjä ominaisuuksia, niin siihen perustuvilla tuloksilla tai tulkinnoilla on tietty arvo. Käytännön esimerkki ”jos-niin”-säännöstä voisi olla seuraava: ”jos hakijan tulos ylittää asetetun alarajan, hän tulee hyväksytyksi”. Kuten jo aiemmin luvussa 4.2 on esitetty validiusargumentointi edellyttää tulkinta-argumenttien arviointia. Jotta tulkitut testi- tai tutkimustulokset tai niiden hyödyntäminen olisivat valideja, se edellyttää, että tulkinta-argumentti on johdonmukainen, tulkinnat ovat järkeviä ja niistä tehdyt oletukset ovat uskottavia. Tulkinta-argumentin avulla on mahdollista päätyä tulkintoihin, joita voidaan arvioida ja analysoida tarkemmin. Mikäli tulkinta-argumentit sisältävät tilastollisia yleistyksiä, otoksen perusteella tehdyistä havainnoista ja niiden syistä tulisi voida tehdä koko perusjoukkoa koskevia yleistyksiä. Validiusargumentointia tulisi tällöin arvioida yleistettävyyden näkökulmasta. Jos taas tulkinnasta ei voi tehdä yleistystä, empiiristen havaintojen johdonmukaisuus

kyseisessä tapauksessa olisi epäolennaista validiuden näkökulmasta tai tulkinnan validiutta tulisi epäillä. Havaintojen yleistettävyys sisältyy validiusargumentteihin, jos tulkinta-argumentteihin sisältyy yleistettävyden ulottuvuus. (Kane 2006, 23.)

Kanen mukaan tulkinta-argumentteja tulee voida arvioida ainakin sellaisten kriteereiden, kuten selkeyden, johdonmukaisuuden sekä uskottavien päätelmien ja oletusten mukaan (Kane 2006, 29). Päätelmien, samoin kuin niistä tehtyjen johdtopäätösten, tulee perustua tehtyihin havaintoihin. Perustelujen tulee olla johdonmukaisia, ja niiden tulee olla johdonmukaisessa yhteydessä tehtyihin havaintoihin. Perustelujen tulee olla uskottavia, eikä yhtään olennaista päätelmää saa jättää tarkastelun ulkopuolelle. Jotkut päätelmistä edellyttävät empiirisiiä havaintoja, jotta ne olisivat yleistettävissä. Osa päätelmistä voidaan johtaa havainnoista ja tulkinnoista helposti, mutta toisia päätelmiä täytyy tarvittaessa dokumentoida ja analysoida yksityiskohtaisemmin. Kanen mukaan on olemassa tilanteita, joissa oletukset ovat erittäin kyseenalaisia, ja siksi niitä tulee voida perustella useisiin eri rinnakkaisiin näyttöihin perustuen. Kuten tieteellisiä teorioita yleensäkin, voidaan tulkinta-argumentteja haastaa tai kyseenalaistaa useilla eri tavoilla. Kane pitää hyvänä argumentointitapaa sellaista, jossa tulkinta-argumentin rinnalla tarkastellaan toista, vaihtoehtoista teoriaa, hypoteesia tai argumenttia. Validiusargumentointi onnistuu, mikäli esitetty tulkinta antaa riittävän taustan ja argumentaatiolla onnistutaan sulkemaan pois muut tulkinnat. (Kane 2006, 29.)

Tämän tutkimuksen luvussa 6.4 tarkastellaan empiirisellä aineistolla saatuja tuloksia eri sisäisen validiuden näkökulmista. Kriteerivalidiuden näkökulmaa käytetään, kun oppilaiden yhteiskuntaopin osaamista tarkastellaan kahdella eri mittarilla, oppilaiden saamista arvosanoilla ja seuranta-arvioinnista saaduilla tuloksilla. Lisäksi kriteerivalidiuden näkökulmaa hyödynnetään, kun saatuja yhteiskuntaopin tuloksia ristiinvalidoidaan historian tulosten kanssa, jolloin kriteerinä toimivat samalla mittarilla eli arvosanoilla tai seuranta-arviointituloksella mitattu toisen oppiaineen tulos. Kriteerivalidiuden mittana käytetään yleensä korrelaatiokerrointa. Tässä tutkimuksessa korrelaatiokertoimien avulla tarkastellaan sitä, miten oppilaan osaamista mittaavat arvosanat ja seuranta-arviointitulokset korreloivat oppilastasolla, koulujen sisällä ja koulujen välillä ja miten kahdessa eri oppiaineessa, yhteiskuntaopissa ja historiassa, tulokset korreloivat keskenään.

Messickin mukaan testin tai kokeen tulkinta ja sen havainnoitavuus perustuvat käsittevalidiuteen. Tämä tarkoittaa, että havainto joko konvergoi tai erottelee tuloksia. Teoreettisesti olennaiset empiiriset yhteydet voivat esimerkiksi konvergoitua testin tai kokeen ja siinä hyödynnettyjen, samasta teoriataustasta johdettujen mittausmenetelmien kesken tai samaa käsitettä mittaavien eri mittareiden kesken. (Messick 1980, 1019.) Konvergenssivalidiudessa on kyse siis siitä, että mittaukset ovat johdonmukaisesti yhteydessä niihin käsitteisiin tai ajatusmalleihin ja muuttujiin, joihin niiden tulee olla, kun ne on johdettu samasta teoriataustasta (Messick 1980, 1016). Oppilaiden arvosanojen ja seuranta-arviointituloksen validiutta tarkastellaan sisäiseen validiuteen liittyvään konvergenssivalidiuden näkökulmasta,

kun tarkastellaan, konvergoituvatko yhteiskuntaopin arvosanat ja seuranta-arviointi oppilastasolla ja koulutasolla. Konvergenssivalidiuden määritelmän mukaan arvosanojen ja ratkaisuosuuksien tulisi konvergoitua eli niiden tulisi osoittaa samaan suuntaan (Kane 2006, 36–37; Ketokivi 2009, 67–68).

Messickin mukaan arviointiin saattaa liittyä epävalidiutta esimerkiksi siksi, että se on liian kapea-alaista ja jokin käsitevalidiuden ulottuvuus saattaa jäädä ottamatta huomioon (Messick 1989b, 4; 1995, 742–743). Rakenteeseen kuulumaton vaihtelu tarkoittaa systemaattisesta virheestä johtuvaa asiaan kuulumatonta vaihtelua, joka toimii erityisenä virhelähteenä (Kane 2006, 39). Luvussa 6.4 tarkastellaan tutkimusaineiston pohjalta oppilaiden saamien arvosanojen validiutta tai niiden mahdollista epävalidiutta.

5 TUTKIMUKSEN LUOTETTAVUUSTARKASTELU

Tässä luvussa tarkastellaan tämän tutkimuksen validiutta ja reliabiliutta eli niitä systemaattisia tekijöitä, jotka ovat saattaneet alentaa tutkimuksen tai siinä hyödynnettyjen mittausten luotettavuutta. Validiudella tarkoitetaan, että tutkimuksessa käytetyt mittarit tai tutkimusmenetelmät ovat mitanneet juuri sitä, mitä on ollut tarkoituskin mitata (Hirsjärvi, Remes & Sajavaara 2009, 231–232). Reliabiliteetilla tarkoitetaan tutkimuksen toistettavuutta eli siitä, että tutkimuksessa hyödynnetyt mittaustulokset ovat toistettavissa ja että ne antavat ei-sattumanvaraisia tuloksia (Hirsjärvi, Remes, Sajavaara 2009, 231). Mikäli mittari on reliabiliteetti, vastaukset ovat eri mittauskerroilla samanlaisia (Metsämuuronen 2006, 64).

5.1 Tutkimusaineiston validius

Tässä yhteydessä validiutta tarkastellaan ulkoisen validiuden näkökulmasta siten, että tutkitaan, onko tulos yleistettävissä eli onko otanta onnistunut. Otannan tulisi edustaa mahdollisimman hyvin kohdejoukkoa ja olla satunnaisesti valittu. (Metsämuuronen 2009, 28.) Ulkoisessa validiudessa tarkastellaan myös, onko mittaustilanteessa ollut jotain sellaista, kuten vastauskatoa, joka olisi voinut heikentää tuloksen yleistettävyyttä. Tutkimuksen sisäisessä validiustarkastelussa arvioidaan puolestaan sitä, onko tämän tutkimuksen mittarit muodostettu oikein ja mitataanko niillä sitä, mitä on ollut tarkoituskin mitata. Osana sisäistä validiuden tarkastelua käsitellään myös asiantuntijatriangulaatiota, tehtävien esitestaamista sekä tehtävien sensorointia. Sisäistä validiutta käsittelevässä luvussa 5.1.2 tarkastellaan tätä tutkimusta myös sisältövalidiuden näkökulmasta, kun arvioidaan yhteiskuntaopin ja historian tehtäväsarjojen luotettavuutta eli tehtävien sisältöjen vastaavuutta suhteessa Perusopetuksen opetussuunnitelman perusteisiin (2004).

5.1.1 Ulkoinen validius

Yhteiskunnallisten aineiden seuranta-arviointi järjestettiin otantaperustaisena. Otannassa otettiin huomioon alueellinen edustavuus ja tilastollinen kuntaryhmä. Osituksessa perusjoukko, jonka muodostivat suomenkieliset ja ruotsinkieliset perusopetusta antavat koulut ja niiden 9. luokan oppilaat, jaettiin osiin maakunnittain ja tilastollisten kuntatyyppien mukaan. Tällä tavalla saatiin mukaan kaupunkimaisten, taajamien ja maaseutumaisten kuntien kouluja ja oppilaita oikeassa suhteessa ympäri Suomea. Menettelyllä voitiin varmistaa, että otos edusti luotettavasti kaikkia kouluja ja oppilaita. Otantaa tehtäessä oli otettava huomioon se, että keväällä 2011 Opetushallitus suoritti yhteiskunnallisten aineiden seuranta-arvioinnin lisäksi matematiikan ja luonnontieteiden seuranta-arvioinnit. Otannassa oleva

koulu saattoi olla mukana kevään 2011 aikana vain yhdessä edellä mainituista seuranta-arvioinneista. Ruotsinkielisten koulujen oppilaiden otannassa oli yliedustus siksi, että oppilaita haluttiin otokseen riittävän monesta koulusta, jotta voitiin tehdä tilastollisia johtopäätöksiä.

Seuranta-arvioinnin piiriin tulleet oppilaat valittiin kouluissa systemaattisella tasaväliotannalla. Kouluista, joiden 9. luokkien oppilasmäärä oli korkeintaan 40, otokseen kuuluivat kaikki oppilaat. Jos oppilaita oli 41–70, otokseen otettiin aakkosjärjestyksessä 2/3 oppilaista. Kouluissa, joiden 9.-luokkalaisten oppilasmäärä oli 71–150, otokseen poimittiin joka toinen oppilas ja sitä suuremmista kouluista aakkoslistasta poimittiin otokseen tasavälein joka kolmas oppilas. Mikäli oppilas oli poissa koulusta seuranta-arviointipäivänä, hänen tilalleen otettiin aakkosista seuraava oppilas muuttamatta poimintajärjestystä muulta osin alkuperäisestä. Myös ne oppilaat, joille on tehty päätös erityisestä tuesta tai joilla oli yksilöllistetty opetussuunnitelma jossakin oppiaineessa, osallistuivat yhteiskunnallisten aineiden seuranta-arviointiin. Oppilaat joiden äidinkieli oli muu kuin suomi tai ruotsi, kuuluivat myös tämän seuranta-arvioinnin piiriin. Rehtorilla oli kuitenkin mahdollisuus poiketa järjestelyistä perustellusta syystä.

Tästä väitöstutkimusaineistosta on poistettu Opetushallituksen aineistossa mukana olleet 423 oppilasta, jotka saivat erityistä tukea tai jotka puhuivat muuta kuin suomea tai ruotsia äidinkielenään. Koska tutkimusongelmana ovat yleisopetuksen oppilaiden saamat arvosanat sekä ne normit, jotka määrittävät arvosanan antamista yleisopetuksessa, erityistä tukea saaneiden oppilaiden tai muuta kuin suomea tai ruotsia äidinkielenään puhuneiden oppilaiden poistaminen tutkimusaineistosta ei huononna ulkoista validiutta. Kyse ei myöskään ole kadosta, koska tutkittavana perusjoukkona ovat yleisopetuksen suomen- ja ruotsinkieliset oppilaat. Näin ollen tämän väitöstutkimuksen aineisto on eri kuin Opetushallituksen seuranta-arviointiaineisto.

Tämän väitöstutkimuksen otos käsittää 3 613 suomenkielisten ja 545 ruotsinkielisten koulujen oppilasta eli yhteensä 4 158 oppilasta. Suomenkielisten oppilaiden osuus on 87 % ja ruotsinkielisten koulujen oppilaiden osuus on puolestaan 13 %. Otoksen 109 koulusta oli suomenkielisiä kouluja 94 ja ruotsinkielisiä kouluja oli 15. Otoksesta poistettiin lisäksi yksi oppilas, jonka historian sekä yhteiskuntaopin arvosanat olivat 8, mutta joka ei ollut vastannut seuranta-arvioinnissa kuin yhteen historian tehtävään eikä yhteenkään yhteiskuntaopin seuranta-arviointitehtävään.

Osa oppilaista oli poissa koulusta sinä päivänä, jolloin seuranta-arviointi järjestettiin, tai he tekivät vain osan tehtävistä. Pienissä kouluissa poissaolijoita ei voitu korvata aakkosissa seuraavalla oppilaalla, ja joissakin suuremmissa kouluissa rehtorit eivät korvanneet poissaollutta oppilasta seuraavalla. Joissain kouluissa muutamien, yksittäisten oppilaiden suorituksia ei arvioitu lainkaan tai vastauslomakkeet oli täytetty puutteellisesti. Yksi koulu ei palauttanut oppilaiden tehtävävihkoja

ollenkaan. Opetushallituksessa tarkastettiin kaikkien otosoppilaiden tehtävävihkot ja siinä yhteydessä aineistosta poistettiin viisi tehtävävihkoa, joiden vastaajat eivät olleet vastanneet tehtäviin asiallisesti tai lainkaan. Ulkoinen kato oli suuruudeltaan 4,8 %. Se jakautui satunnaisesti eikä vinouttanut arvioinnin tuloksia.

Sisäistä katoa oli yksittäisten vastaamatta jääneiden kysymysten osalta joko oppilastietolomakkeessa tai tehtävälomakkeissa. Tämän väitöstutkimuksen otoksessa oli 49 oppilasta, joilta puuttui tieto historian arvosanasta (1,2 %) ja 55 oppilasta, joilta vastaava tieto puuttui yhteiskuntaopista (1,3 %). Jompikumpi tai molemmat arvosanatiedot puuttuivat 70 oppilaalta, mikä on 1,7 % lopullisen otoksen oppilaista. Näiden keskeisten muuttujien kato oli hyvin pieni. Edellä mainitun lisäksi katoa oli joissain taustamuuttujiksi luokiteltavissa muuttujissa, esimerkiksi jatko-opintosuuntautumisvaihtoehdossa (1,8 %), mutta kysymys oli silloinkin niin pienistä määristä, etteivät ne vaikuta tulosten tulkintaan tai johtopäätöksiin. Vanhempien koulutus aiheutti eniten sisäistä katoa. Siinä kato oli 2,3 % eli 97 oppilasta ei tiennyt jommankumman tai kummankaan vanhemman koulutusta. Täydelliset tiedot saatiin kaikkiaan yhteensä 3 909 oppilaalta. Yhteensä 249 oppilasta oli jättänyt vastaamatta johonkin taustatietoja kartoittavaan kohtaan, mikä on 6 % väitöstutkimuksen otoksesta.

Mikäli oppilas oli jättänyt vastaamatta yhteiskuntaopin tai historian tehtäväosioon, hänen vastauksensa pisteitettiin nollassi, joten tällainen puuttuva tieto ei lisännyt katoa.

5.1.2 Sisäinen validius

Tässä tutkimuksessa on käytetty Opetushallituksen yhteiskunnallisten aineiden seuranta-arviointiin laatimaa mittaria. Yhteiskunnallisten aineiden seuranta-arvioinnin valmistelu ja mittarin laadinta alkoivat Opetushallituksessa keväällä 2010. Opetushallituksen tekemissä oppimistulosten seuranta-arvioinneissa keskeinen painopiste oli kyseisen oppiaineen tietojen ja taitojen osaamisen mittaamisessa sekä siinä, kuinka hyvin Perusopetuksen opetussuunnitelman perusteissa (2004) määritellyt oppiaineen tavoitteet, sisällöt ja kriteerit toteutuivat.

Tutkijatriangulaatio-periaatteen mukaan yhteiskuntaopin ja historian seuranta-arvioinnin tehtäväsarjojen laatijat edustivat laajaa oppiaineiden kokemusta ja ammattipätevyyttä. Seuranta-arvioinnin mittaristo laadittiin yhteistyössä Opetushallituksen ja Helsingin yliopiston Koulutuksen arviointikeskuksen kanssa. Marraskuussa 2010 koottiin asiantuntijaryhmä, johon kuului historian ja yhteiskuntaopin opetuksen sekä arviointimenetelmien asiantuntijoita. Asiantuntijaryhmään kuului historian ja yhteiskuntaopin didaktiikan asiantuntijoita Helsingin yliopiston opettajakoulutuslaitokselta, historian ja yhteiskuntaopin opettajia Helsingin normaalilyseosta ja Pohjois-Tapiolan koulusta sekä oppimistulosten arvioinnin asiantuntijoita

Opetushallituksesta ja Helsingin yliopiston Koulutuksen arviointikeskuksesta. Edellisen ryhmän lisäksi Opetushallituksessa muodostettiin syksyllä 2010 oma yhteiskunnallisten aineiden työryhmä, johon kuuluivat projektipäällikkö, erikoistutkija, joka vastasi metodeista, sekä kaksi tutkimussihteeriä.

Yhteiskuntaoppiin ja historiaan laaditut tehtävät perustuivat Perusopetuksen ope-
tussuunnitelman perusteissa (2004) määriteltyihin tavoitteisiin, sisältöihin ja oppi-
aineille asetettuihin kriteereihin. Tehtävien tarkoituksena oli mitata monipuolisesti
sitä, miten historialle ja yhteiskuntaopille asetetut tavoitteet on saavutettu. Tehtä-
väsarjat esitettiin laajasti kahdella erilaisella tehtäväsarjalla tammi-helmikuussa
2011. Esikokeeseen osallistui yhteensä noin 450 perusopetuksen 9. luokan oppilasta
11 eri koulusta ympäri Suomea. Koska otoksessa oli myös ruotsinkielisten koulu-
jen oppilaita, käännetty esikoetehtävät testattiin yhdessä ruotsinkielisessä koulussa.
Esikokeeseen osallistuneiden suomen- ja ruotsinkielisten opetusryhmien opettajia
pyydettiin kommentoimaan tehtäväsarjojen tehtäviä.

Opetushallituksen ohjeistuksen mukaan kahdesta esikoetehtäväsarjasta koostettiin
yksi yhteiskuntaopin ja historian yhteinen tehtäväsarja siten, että 50–60 % oppi-
laista oli ratkaissut noin 50 % valituista tehtävistä oikein. Osa tehtäväsarjaan vali-
tuista tehtävistä oli vaikeita (ratkaisuosuus noin 20 %), osa keskivaikeita (ratkaisuo-
suus 40–60 %) ja osa helppoja (ratkaisuosuus yli 80 %).

Esikokeen jälkeen varsinaiseen seuranta-arvioinnin tehtäväsarjaan valittiin parhaat
osiot ja niihin tehtiin tarkennuksia esikokeeseen osallistuneiden ryhmien opetta-
jilta saadun palautteen perusteella sekä esikoetehtävien tarkastusvaiheessa tehtyjen
muiden havaintojen pohjalta. Varsinaisen yhteiskunnallisten aineiden seuranta-
arvioinnin tehtävien valinnassa käytettiin perinteisiä menetelmiä, kuten vaikeusta-
son arviointia ja reliaabeliuden tarkistamista. Yksittäisten tehtävien osalta kriteerinä
pidettiin, että niiden tulee korreloida positiivisesti ja kohtalaisen voimakkaasti (r_{xy}
> 0,20) kokonaisratkaisuosuuteen. Tehtävien tuli mitata johdonmukaisesti oppi-
aineen yleistä osaamistasoa. Tavoitteena oli saada arviointiin vaikeustasoltaan eri-
laisia tehtäviä. Tehtävien erottelukyky varmistettiin. Osan tehtävistä tuli erotella
heikosti suoriutuvat oppilaat, osan tuli erotella keskimääräisesti suoriutuvia ja osan
piti erotella hyvin suoriutuvat oppilaat. Erottelukyvyltään heikot osiot poistettiin
lopullisesta seuranta-arvioinnista. Tehtävien reliaabeliuden tarkistamisessa käytettiin
sisäistä johdonmukaisuutta mittaavaa Cronbachin menetelmää ja tunnuslukuna
 α -kerrointa.

Yhteiskunnallisten aineiden asiantuntijat laativat esikokeen pohjalta yhden yhteiskunnallisten aineiden tehtäväsarjan pisteitysohjeineen. Jokaiselle tehtävälle oli selkeät vastaus- ja pisteitysohjeet. Tehtäväsarjassa oli monipuolisesti historian ja yhteiskuntaopin osaamista mittaavia kirjallisia tehtäviä. Tehtävistä noin 30 % oli niin sanottuja tuottamistehtäviä eli avotehtäviä ja 70 % oli monivalintatehtäviä. Kummassakin oppiaineessa niiden tehtäviltä edellytettiin sitä, että ne mittasivat oppiaineen yleistä osaamistasoa.

Yhteiskunnallisten aineiden seuranta-arviointitietoa kerättiin vakiintuneeseen tapaan useammanlaisten tehtävien avulla ja yhdellä tehtäväsarjalla. Jo aiemmin tämän tutkimuksen johdantoluvussa on esitetty malli, jota Opetushallituksen toteuttamissa oppimistulosten arvioinneissa noudatetaan. Jotta tehtävät olisivat valideja, niiden tuli mitata oikeita asioita ja perustua Perusopetuksen opetussuunnitelman perusteissa (2004) määriteltyihin yhteiskuntaopin ja historian tietoihin ja taitoihin. Tehtävien tuli mitata kyseisten oppiaineiden osaamista riittävän laajalaisesti.

Seuranta-arvioinnin sisältö määriteltiin niin, että arvioinnilla selvitettiin oppilaiden tietoja ja taitoja yhteiskuntaopin ja historian keskeisillä alueilla. Tehtäviä laadittaessa kiinnitettiin huomiota siihen, että ne liittyivät yhteen tai useampaan niistä tavoitelauseista, joilla yhteiskuntaopin ja historian tavoitteet ja keskeiset sisällöt on määritelty Perusopetuksen opetussuunnitelman perusteissa (2004). Yhteiskuntaopin ja historian luokittelusta ja tehtävien ratkaisuosuuksista suhteessa opetussuunnitelman perusteissa määriteltyihin tavoitteisiin ja sisältöihin on tarkemmin kerrottu Opetushallituksen seuranta-arviointiraportissa (Ouakrim-Soivio & Kuusela 2012, 138–140).

Koska Opetushallitus käyttää seuranta-arvioinnissa sille laadittuja ja esitettuja historian ja yhteiskuntaopin tehtäviä jatkossa niin sanottuina ankkuritehtävinä, joiden avulla voidaan tarkastella oppilaiden osaamistasoa, tehtävät eivät ole kaikilta osin julkisia. Samasta syystä seuranta-arviointiraportissa on julkaistu vain esimerkinomaisesti erityyppisiä tehtäviä, jotta lukija saa käsityksen, minkälaisilla tehtäväsarjoilla oppilaiden osaamista ja opetussuunnitelman toteutumista on mitattu. Seuranta-arvioinnissa käytettyjen tehtävien kuvaukset löytyvät seuranta-arviointiraportista yhteiskuntaopissa (Ouakrim-Soivio & Kuusela 2012, 89–100) ja historiassa (Ouakrim-Soivio & Kuusela 2012, 78–88). Taulukoissa 6 ja 7 on esitetty tiivistetysti, miten Perusopetuksen opetussuunnitelman perusteiden (2004) tavoitteet ja kriteerit toteutuivat lopullisissa yhteiskuntaopin ja historian seuranta-arvioinnin tehtäväsarjoissa ja kuinka suuri osuus prosentteina ilmaistuna kertyi kyseisestä tehtäväluokituksesta.

TAULUKKO 6. Yhteiskuntaopin tehtävien luokittelu, tehtävien enimmäispistemäärät ja niiden ratkaisuosuudet.

Tehtävien luokittelu opetussuunnitelman perusteissa määriteltyjen tavoitteiden ja kriteerien mukaan:	Median välittämien tietojen, tilastojen ja graafien kriittinen tulkinta	Yhteis-kunnallisten asioiden perustelu	Yhteiskunnallisen päätöksenteon ja taloudellisten ratkaisujen useat erilaiset vaihtoehdot ja niiden seuraukset	Yhteiskunnallisen ja taloudellisen toiminnan eettiset kysymykset. Raportoitiin erikseen, sillä tehtävissä mitattiin ennen kaikkea sitä, osaako oppilas soveltaa vastuullisen kuluttamisen periaatteita käytännön tilanteessa.
Enimmäispistemäärä /prosenttiosuus (%) kokonaispisteistä	Yhteensä pisteitä enimmillään 20 p/ 30 % kokonaispisteistä	Yhteensä pisteitä enimmillään 24 p/ 37 % kokonaispisteistä	Yhteensä pisteitä enimmillään 22 p/ 33 % kokonaispisteistä	Raportoitiin erikseen
Ratkaisuosuus (%)	Ratkaisuosuus 58 %	Ratkaisuosuus 79 %	Ratkaisuosuus 53 %	Raportoitiin erikseen

TAULUKKO 7. Historian tehtävien luokittelu, tehtävien enimmäispistemäärät ja niiden ratkaisuosuudet.

Tehtävien luokittelu opetussuunnitelman perusteissa määriteltyjen tavoitteiden ja kriteerien mukaan:	Lähteiden käyttö ja tulkinta, jossa faktan erottaminen mielipiteestä, lähteet ja niistä tehdyt tulkinnat, tapahtumien erilaiset tulkinnat, erilaisten lähteitten lukeminen ja niiden tulkinta, menneisyyttä koskeviin kysymyksiin vastaaminen eri lähteiden avulla	Ajan hahmottaminen, jossa eri aikakaudet ja niiden nimeämiseen liittyvät olennaiset piirteet, ilmiöiden jatkuvuuden tunnistaminen eri aikakaudesta toiseen, tapahtumien sijoittaminen ajallisiin yhteyksiin	Historiallinen empatia, historiallisten tapahtumien syy- ja seuraussuhteet sekä ilmiön selittäminen eri toimijoiden kautta, jossa menneen ajan ihmisen asemaan eläytyminen ja ymmärrys siitä, miksi ennen toimittiin toisin kuin nyt	Erilaiset historian osaamiseen liittyvät taidot, kuten tapahtuman tai ilmiön selittäminen joidenkin toimijoiden kautta, asiaa selittävien tärkeiden tekijöiden erottaminen vähemmän tärkeistä, historiallisten tapahtumien syyt ja seuraukset, tapahtumista ja ilmiöistä muodostetut omat perustellut käsitykset
Enimmäispistemäärä/prosenttiosuus (%) enimmäispisteistä	Yhteensä pisteitä enimmillään 35 p/ 48 % enimmäispisteistä	Yhteensä pisteitä enimmillään 25 p / 34 % enimmäispisteistä	Yhteensä pisteitä enimmillään 4 p/ 5 % enimmäispisteistä	Yhteensä pisteitä enimmillään 9 p/ 12 % enimmäispisteistä
Ratkaisuosuus (%)	Ratkaisuosuus 48 %	Ratkaisuosuus 58 %	Ratkaisuosuus 51 %	Ratkaisuosuus 43 %

Yhteiskuntaopin ja historian tehtäväsarjassa oli sekä tuottamista vaativia avotehtäviä että monivalintatehtäviä. Monivalintatehtävissä on aina arvausmahdollisuus. Rantasen (2003) tutkimuksen mukaan monivalintatehtävissä kolmesta neljään vaihtoehtoa on riittävä määrä luotettavuuden kannalta ja miellyttävä määrä oppilaalle vastata. Mittarin mittaustarkkuus nousee, mikäli monivalintatehtävien rakenne on sellainen, että jokainen vastausvaihtoehto on itsenäinen väite. Samassa ajassa vastataan huomattavasti useampaan oikein–väärin-tehtävään kuin perinteiseen monivalintatehtävään. (Rantanen 2003, 5, 60, 191.) Yhteiskuntaopin ja historian monivalintatehtävät rakennettiin siten, että ne noudattivat edellä mainittuja Rantasen tutkimuksen tuloksia. Yhteiskunnallisten aineiden tehtävissä monivalinta- ja tuottamistehtävät sekä niiden pisteitykset jakautuivat seuraavan taulukon 8 mukaisesti.

TAULUKKO 8. Yhteiskuntaopin ja historian tehtäväsarjojen monivalinta- ja tuottamistehtävien lukumäärä sekä niistä kertyneet enimmäispisteet prosentteina.

Lähde: Ouakrim-Soivio & Kuusela 2012, 77.

	Yhteiskuntaopin tehtäväsarja	Historian tehtäväsarja
Tehtävien lukumäärä/ enimmäispistemäärä	29 tehtävää/ 66 pistettä	27 tehtävää/ 73 pistettä
Monivalintatehtäviä/ enimmäispistemäärä	56 tehtäväosiota/ 47 pistettä	19 tehtäväosiota/ 41 pistettä
Tuottamis- eli avotehtäviä/ enimmäispistemäärä	12 tehtäväosiota/ 19 pistettä	8 tehtäväosiota/ 32 pistettä

Yhteiskuntaopin tehtävät laadittiin arvioimaan oppilaiden osaamista perusopetuksen päättövaiheessa opetussuunnitelman perusteiden määrittämässä yhteiskuntaopin keskeisissä tiedoissa ja taidoissa (POPS 2004, 226–227). Tehtäviä laadittaessa kiinnitettiin huomiota siihen, että niillä mitatut tiedot ja taidot olivat keskeistä yhteiskuntaopin ydinosaamista, jota oppilailta voitiin edellyttää perusopetuksen päättövaiheessa. Yhteiskuntaopin tehtävien aihealueet olivat keskeisiä perusopetuksessa, ja niiden valinnassa otettiin huomioon suomalainen yhteiskunta ja talouselämä, Euroopan unioni ja yhteiskunnallis-taloudelliseen toimintaan liittyvät kysymykset.

Monivalintatehtävissä oli kysymyksiä yhteiskuntaopista, taloustiedosta, lakitiedosta, yrittäjyydestä ja Euroopan unionista. Yhteiskuntaopin keskeisten sisältöjen, tietojen ja käsitteiden, kuten hyvinvoinnin, vaikuttamisen ja päätöksenteon, kansalaisen turvallisuuden ja taloudenpidon sekä talouspolitiikan ja kansantalouden lisäksi kokeen avulla arvioitiin myös erilaisia yhteiskuntaoppiin keskeisesti liittyviä taitoja, kuten tilastojen ja graafien lukemista sekä medialukutaitoa. Tuottamistehtävissä oppilaan tuli osata tulkita, päätellä, yhdistellä ja perustella sekä tuottaa kirjallinen vastaus annettuihin kysymyksiin. Jotkut kysymyksistä edellyttivät asian, tapahtuman tai ilmiön pohdintaa, syy–seuraussuhteiden ymmärtämistä tai ilmiön tarkastelua monesta eri näkökulmasta ja perustellen.

Tehtävien laadinnassa kiinnitettiin erityistä huomiota siihen, että ne mittasivat fakta- ja käsitetiedon lisäksi erityisesti yhteiskuntaopin taitojen hallintaa ja osaamista. Seuranta-arvioinnin tehtävät luokiteltiin Perusopetuksen opetussuunnitelman perusteissa (2004) määriteltyjen yhteiskuntaopin tavoitteiden ja kriteereiden (POPS 2004, 226–227) mukaan kolmeen pääryhmään. Ne olivat

- median välittämien tietojen, tilastojen ja graafien kriittinen tulkinta
- yhteiskunnallisten asioiden perustelu sekä
- yhteiskunnallisen päätöksenteon ja taloudellisten ratkaisujen useat erilaiset vaihtoehdot ja niiden seuraukset.

Sekä monivalinta- että avotehtävissä kiinnitettiin huomiota siihen, että niissä oppilas joutui tulkitsemaan kriittisesti median välittämiä tietoja, tilastoja ja graafisia esityksiä, perustelemaan käsityksiään yhteiskunnallisista ja taloudellisista asioista sekä ottamaan kantaa yhteiskunnallisen ja taloudellisen toiminnan eettisiin kysymyksiin kuluttajana.

Koska historiaa käytetään tässä väitöstutkimuksessa tulosten vertailemiseksi ja yhteiskuntaopissa saatujen tulosten argumentoinnin perustana, on seuraavassa kuvattu lyhyesti myös historian seuranta-arvioinnin tehtävät. Myös historian osaamista mitattiin seuranta-arvioinnissa sekä monivalinta- että tuottamistehtävillä, jotka sisälsivät historian keskeisten sisältöjen, eri aikakausien, tapahtumien ja ilmiöiden osaamista sekä erilaisten historiaan liittyvien taitojen, kuten kartan-, dokumentin-, kuvan- ja tilastonlukutaitojen hallintaa. Historian tehtäviä laadittaessa kiinnitettiin huomiota siihen, että niissä mitatut tiedot ja taidot olivat keskeistä historian ydinosaamista, jota oppilailta voitiin edellyttää siitä huolimatta, että oppilaat opiskelevat yleensä historian oppimäärän perusopetuksen vuosiluokilla 5–8. Tehtävien aihealueet olivat keskeisiä perusopetuksessa, ja niiden valinnassa otettiin huomioon Suomen ja yleisen historian aikakausia, käsitteitä, ilmiöitä, tapahtumia ja keskeisimpiä henkilöitä.

Myös historian tehtävien laadinnassa kiinnitettiin erityistä huomiota siihen, että ne mittasivat faktatiedon lisäksi historiaan liittyvien taitojen hallintaa. Tehtävät perustuivat perusopetuksen opetussuunnitelman perusteissa määriteltyihin tavoitteisiin ja kriteereihin (POPS 2004, 217, 222), ja ne luokiteltiin historian taitojen osalta neljään pääryhmään:

- lähteiden käyttö ja tulkinta
- ajan hahmottaminen
- menneen ajan ihmisen asemaan eläytyminen ja historiallinen empatia sekä
- erilaisten historian osaamiseen liittyvien taitojen, kuten tapahtumien syy- ja seuraussuhteet, ilmiön selittäminen eri toimijoiden kautta ja omien perustelujen mielipiteiden muodostaminen. (Ouakrim-Soivio & Kuusela 2012, 77.)

Suurin osa seuranta-arviointia järjestäneistä historian ja yhteiskuntaopin opettajista piti yhteiskunnallisten aineiden seuranta-arviointia kokonaisuudessaan 9.-luokkalaisille sopivana. Opettajien käsitysten mukaan tehtävät vastasivat opetussuunnitelman tavoitteita ja sopivat vaikeustasoltaan kohtalaisesti tai hyvin perusopetuksen päättövaiheen opilaille.

Koulujen opettajia pyydettiin seuranta-arvioinnin jälkeen arvioimaan käytettyjen historian ja yhteiskuntaopin tehtävien vaikeustasoa ja sitä, miten ne olivat vastanneet opetussuunnitelman tavoitteita. Vastanneista opettajista (N = 165) 27 % ilmoitti, että seuranta-arvioinnin tehtävät vastasivat hyvin opetussuunnitelman tavoitteita, 53 % vastanneista opettajista oli sitä mieltä, että tehtävät vastasivat kohtalaisesti ja 7 % vastanneista oli sitä mieltä, että tehtävät vastasivat huonosti opetussuunnitelman tavoitteita. Kyselyyn vastanneista opettajista 13 % ei vastannut tähän kysymykseen lainkaan. Suurin osa koulujen opettajista, jotka vastasivat kyselyyn (76 %), piti tehtäviä vaikeustasoltaan sopivana. Vastanneista opettajista 42 % piti seuranta-arviointia kokonaisuudessaan sopivana, 40 % piti vaikeana, 5 % piti erittäin vaikeana ja 13 % ei vastannut tähän kysymykseen.

Opettajat esittivät arvioinnin kehittämistä koskevia ehdotuksia, kommentteja ja toiveita. Osa kommentteista (N = 146) liittyi yksittäisiin tehtäviin (22 %), osa seuranta-arvioinnin pituuteen ja liian pitkään koeaikaan (19 %) ja osa avotehtävien pisteityksen viemään aikaan (16 %). Muut kommentit koskivat seuranta-arvioinnin ajankohtaa (16 %), pisteitysohjeita (11 %), historian ja yhteiskuntaopin arvioimista samanaikaisesti ja samalla tehtävävihkolla (10 %) sekä monivalintatehtävien korjaamista nopeuttavan vastaussabluunan puuttumista (3 %). (Ouakrim-Soivio & Kuusela 2012, 130.)

5.2 Tutkimuksen reliabelius

Tässä väitöstutkimuksessa reliabeliutta tarkastellaan tyypillisten tilastollisten menetelmien, Cronbachin α -kertoimen, seuranta-arvioinnin ohjeistuksen, kuten sen osalta, että aineiston keruu ja tehtävien pisteitys on tehty samalla tavalla ja yhdenmukaisin ohjein, sekä sensoroinnin tulosten avulla.

Seuranta-arvioinnin luotettavuuteen liittyy se, antavatko mittavälineet tietoa, joka ei ole sattumanvaraista. Koko seuranta-arvioinnin tehtävien sisäistä yhdenmukaisuutta ilmaiseva Cronbachin α -kerroin oli 0,92. Osa-alueittain tarkasteltuna Cronbachin α oli yhteiskuntaopissa 0,87 ja historiassa 0,86. Cronbachin α -kertoimet osoittavat osioiden ja seuranta-arvioinnissa käytettyjen kysymysten mitanneen arviointivälineiden ja seuranta-arvioinnissa käytettyjen kysymysten mitanneen arviointivälineiden olleita kohteita riittävän johdonmukaisesti. (Ouakrim-Soivio & Kuusela 2012, 131.)

Kouluille tiedotettiin yhteiskunnallisten aineiden arvioinnista tammikuussa 2011, jolloin rehtorit saivat vastattavakseen koulukohtaisen taustakyselylomakkeen. Kouluille tiedotettiin arvioinnista tarkemmin vielä maaliskuussa 2011, ja viikkoa ennen varsinaista arviointipäivää kouluille toimitettiin aineisto, johon kuuluivat tehtävähkot, pisteitys- ja arviointiohjeet yhteiskunnallisten aineiden opettajille sekä kokeen toteutusohjeet rehtorille, opettajille ja kokeen valvojille. Seuranta-arviointi järjestettiin torstaina 14.4.2011 yhtenäisten järjestely- ja valvontaohjeiden mukaisesti. Tehtävien tekemiseen oli varattu aikaa yhteensä kolme 45 minuutin oppituntia eli yhteensä 135 minuuttia. Seuranta-arviointi oli kokonaisuudessaan toteutettava koulussa kuitenkin niin, että se päättyi viimeistään klo 14.00.

Rehtorin ohjeissa selvitettiin seuranta-arvioinnin käytännön järjestelyt ja otoksen toteuttaminen. Ohjeissa painotettiin, että rehtorin tuli huolehtia arviointitehtävien säilymisestä luottamuksellisina. Rehtori tai hänen valtuuttamansa henkilö sai antaa arviointiaineiston seuranta-arviointia valvovalle henkilölle vasta seuranta-arviointipäivänä. Rehtorin ohjeissa korostettiin, että aiemmista, muiden oppiaineaineiden seuranta-arvioinneista poiketen myös ne oppilaat, joille oli tehty päätös erityisestä tuesta, osallistuivat yhteiskunnallisten aineiden seuranta-arviointiin. Lisäksi ne oppilaat, joilla oli esimerkiksi lukivaikeus tai jotka puhuivat äidinkielenä muuta kuin suomea tai ruotsia, kuuluivat tämän arvioinnin piiriin. Rehtorilla oli mahdollisuus poiketa järjestelyistä perustellusta syystä, ja poikkeava menettely pyydettiin kirjaamaan oppilaan tehtävähkoon. Rehtoria pyydettiin vielä huolehtimaan siitä, että oppilaille ei missään vaiheessa kerrottu, ketkä kuuluivat otokseen. Kenellekään oppilaalle ei saanut antaa arvioinnin jälkeen tehtävähkoa. Vaikka oppilas ei saanut tehtävähkoa tai vastauksiaan itselleen, hänellä oli kuitenkin mahdollisuus saada omalta opettajaltaan tieto menestymisestään seuranta-arvioinnissa.

Oppilailta kerättiin myös tietoja heidän läksyjenteostaan, historian ja yhteiskuntaopin arvosanoista, käsityksistään itsestään oppijana sekä siitä, harrastivatko he yleensä historiaan tai yhteiskuntaoppiin liittyviä asioita. Oppilailta ei ole koskaan aikaisemmin oppimistulosten seuranta-arviointien yhteydessä kysytty heidän vanhempiansa koulutustaustasta. Oppilaat vastasivat taustakysymyksiin tehtävähkon alussa olleeseen optiseen osioon, jonka he täyttivät ennen varsinaisiin seuranta-arviointitehtäviin vastaamista.

Kun oppilaat olivat vastanneet yhteiskunnallisten aineiden seuranta-arviointiin, koulujen historian ja yhteiskuntaopin opettajat tarkastivat aineistot Opetushallituksen pisteitysohjeiden mukaan. Historian ja yhteiskuntaopin opettajat arvioivat oppilaiden tuottamistehtävien vastaukset, joiden arvioimiseksi oli laadittu täsmälliset pisteitysohjeet ja tyypillisiä eri pistemääriä osoittavia mallivastauksia pisteitysohjeineen. Opettajilla oli mahdollisuus saada Opetushallituksesta sekä sähköpostitse että puhelimitse ohjeita vastausten pisteitykseen. Aikaa tuottamistehtävien korjaustyöhön oli varattu kolme viikkoa. Korjaustyön yhteydessä opettajilla oli mahdollisuus saada tietoa omien oppilaittensa tuloksista. Koulun oppilaiden tehtävähkot

palautettiin Opetushallitukseen, jossa jokainen tehtävävihko tarkistettiin ennen kuin ne skannattiin ja monivalintatehtävät luettiin optisesti. Monivalintatehtävät tarkistettiin ja pisteitettiin Opetushallituksessa optisesti.

Monivalintatehtävistä saatua tulosta oli mahdollista verrata avotehtävistä saatuun tulokseen ja saada tällä tavoin vertailutietoa kouluissa noudatetun pisteityskäytännön yhtenäisyydestä. Yhteiskuntaopissa opettajan pisteittämien tuottamistehtävien tulosten korrelaatio monivalintatehtävien tuloksiin oli keskimäärin 0,62. Historiassa oppilaiden tuottamistehtävien tulosten korrelaatio monivalintatehtävien tulotasoon oli keskimäärin 0,59. (Ouakrim-Soivio & Kuusela 2012, 131.)

Tuottamistehtävien sensorointi toteutettiin Opetushallituksessa kesäkuussa 2011. Sensoroinnin tehtävänä oli varmistaa, ettei opettajien pisteityksissä ollut systemaattisia tai satunnaisia virhetulkintoja. Systemaattinen yhdenmukaisuus voitiin varmistaa vertaamalla opettajien ja sensoreiden antamien pisteiden keskiarvoja toisiinsa. Satunnaisia virheitä esiintyy vain vähän, jos sensoreiden ja opettajien antamat pistemäärät korreloivat voimakkaasti keskenään. Tuottamistehtävien arvioinnin yhdenmukaisuutta tutkittiin valtakunnallisesta otosaineistosta siten, että joka koulun vastauksista poimittiin satunnaisesti sekä tyttöjen että poikien tehtävävihkoja. Arvioitavaksi poimittiin kaikkiaan 694 oppilaan tehtävävihkot, joka oli lähes 15 % koko otoksesta. Ulkopuolisina arvioijina oli neljä perusopetuksen historian ja yhteiskuntaopin opettajaa eri puolilta maata sellaisista kouluista, jotka eivät kuulu neet otokseen. Yksi arvioitsija oli ruotsinkielinen ja loput kolme olivat suomenkielisiä. Tämän väitöstutkimuksen tekijä toimi Opetushallituksen yhteiskunnallisten aineiden seuranta-arvioinnin projektipäällikkönä ja osallistui tuottamistehtävien sensorointiin neljän muun sensorin apuna. Sensorit arvioivat tehtävät saman pisteitysohjeen perusteella, jota oli käytetty koulussa oppilaiden tuottamistehtävien arvioimiseen, ja he pisteyttivät tuottamistehtävät erilliselle lomakkeelle. Tämän jälkeen koulujen opettajien antamia pistemääriä verrattiin sensoreiden antamiin pistemääriin. Yhteiskuntaopin pisteityksen yleistaso ilmeni siten, että koulujen opettajien antamat pisteet olivat keskimäärin kaksi prosenttiyksikköä matalammat kuin ulkopuolisten sensoreiden antamat pisteet. Opettajien antamien pisteiden ja ulkopuolisten sensoreiden antamien pisteiden korrelaatio yhteiskuntaopin tuottamistehtävissä oli 0,90. Historian vastausten yleistason yhdenmukaisuus näkyi siten, että opettajien ja sensoreiden antamat pisteet olivat hyvin linjassa keskenään. Koulujen opettajien antamien pisteiden ja ulkopuolisten sensoreiden antamien pisteiden korrelaatio historian tuottamistehtävissä oli 0,90. (Ouakrim-Soivio & Kuusela 2012, 132.) Koko seuranta-arvioinnissa, niin historian kuin yhteiskuntaopin tuottamistehtävissä, opettajien antamat summapisteen olivat 18 ja sensoreiden 19 pistettä. Summapisteiden korrelaatio oli 0,90. Koska opettajien ja sensoreiden pisteitykset olivat niin yhdenmukaiset, oppilaiden vastauksista saamiin pistemääriin ei tehty muutoksia.

Tässä väitöstutkimuksessa käytetyt yhteiskuntaopin ja historian arvosanatiedot perustuvat oppilaiden omaan ilmoitukseen. Yhteiskunnallisten aineiden seuranta-arvioinnin osana olleessa oppilastietokyselyssä oppilaita pyydettiin kertomaan viimeisin todistusarvosana sekä yhteiskuntaopissa että historiassa. Keväällä 2011 yhdeksännellä vuosiluokalla olevien oppilaiden historian arvosana oli yleensä sama kuin heille keväällä 2010 annettu historian päättöarvosana. Yhteiskuntaopissa arvosana oli joko syyslukukauden 2010 päätteeksi annettu arvosana tai huhtikuuhun 2011 mennessä annettu oppilaan saama viimeisin jaksoarvosana. Koska oppilaiden arvosanatiedoissa saattoi olla satunnaisvirheitä siksi, että oppilaat eivät ehkä muistaneet arvosanaansa oikein, tai siksi, että yhteiskuntaopissa kyse ei ollut päättöarvosanasta vaan viimeisimmästä todistusarvosanasta, oppilailta saatuja arvosanatietoja verrattiin vuoden 2011 yhteishakurekisterissä olleiden noin 65 000 oppilaan yhteiskuntaopin ja historian arvosanatietoihin. Yhteishakurekisterin mukaan yhteiskuntaopin arvosanojen keskiarvo oli 7,8. Tyttöjen yhteiskuntaopin arvosanojen keskiarvo oli 7,8 ja poikien 7,7. Yhteishakurekisteritietojen mukaan historian arvosanojen keskiarvo oli 7,6. Tyttöjen historian arvosanojen keskiarvo oli 7,8 ja poikien 7,5. Oppilaiden yhteiskuntaopin ilmoittama arvosanojen keskiarvo oli vain hieman korkeampi (0,1 arvosanaa) kuin yhteisvalintarekisterissä olleen oppilasjoukon yhteiskuntaopin arvosanojen keskiarvo. Samoin historian arvosanojen keskiarvo oli oppilailla hieman korkeampi (0,2 arvosanayksikköä) kuin yhteisvalintarekisterissä olleen oppilasjoukon historian arvosanojen keskiarvo.

6 TUTKIMUKSEN TULOKSET: KOULUJEN VÄLISET EROT ARVOSANOIHIN SUHTEUTETUISSA SEURANTA-ARVIOINNIN RATKAISUOSUUKSISSA

Sekä yhteiskuntaopin että historian oppimistulosten seuranta-arviointiaineistot antoivat vahvoja viitteitä siitä, että oppilaiden saamien yhteiskuntaopin ja historian arvosanojen sekä näiden oppiaineiden seuranta-arvioinnissa osoitetun osaamisen välillä oli suuria koulujen välisiä eroja (Ouakrim-Soivio & Kuusela 2012, 110–111). Arvosanojen ja ratkaisuosuuksien väliset erot esiintyivät kaikissa annetuissa arvosanoissa yhteiskuntaopissa arvosanasta 4 arvosanaan 10 ja historiassa päättöarvosanasta 5 arvosanaan 10 (Ouakrim-Soivio & Kuusela 2012, 106–108). Näin siitä huolimatta, että Opetushallituksen seuranta-arvioinnit ovat yleensä vastanneet hyvin oppilaiden saamia arvosanoja, kun oppilaita tarkastellaan yhtenä ryhmänä. Vastaavuutta kuvaava korrelaatiokerroin on yleensä ollut yli 0,6.

Jo aikaisemmin Kuusela (2006) on tutkimuksessaan osoittanut, että keskimääräisen seuranta-arviointimenestyksen ja keskimääräisten arvosanojen yhteys alkaa koulutasolla heiketä. Kyse ei ole satunnaisesta vaan systemaattisesta vaihtelusta. ”Kun koulun oppilaat ovat menestyneet seuranta-arvioinnissa keskimääräistä heikommin, he ovat menestyneet heikosti myös keskimääräisiin arvosanoihin nähden ja kun on osattu hyvin, on osattu hyvin myös keskimääräisiin arvosanoihin nähden”. (Kuusela 2006; ks. myös Jakku-Sihvonen & Kuusela 2012b, 59.)

Koulujen arvosanakäytänteiden ja seuranta-arvioinnissa osoitetun osaamistason välinen ero ei ole uusi ilmiö Opetushallituksen seuranta-arvioinneissakaan. Äidinkielen ja matematiikan lisäksi näitä koulujen välisiä eroja on kuvattu koko 2000-luvun ajan myös muiden oppiaineiden seuranta-arvioinneissa (Hannén 2001, 36; Korkeakoski 2001, 84–85; Mattila 2002, 90–91; 2005, 108–112; Silverström 2002, 101–104; 2004, 123–128; 2006, 60; 2008, 76–77; Toropainen 2002, 7, 102–104; Kuusela 2006, 67–98; Lappalainen 2006, 67; 2010, 85; 2010, 58–68; Hirvonen 2012, 69–71; Kärnä, Hakonen, Kuusela 2012, 93; 148–150; Ouakrim-Soivio & Kuusela 2012, 110–112).

Yhteiskuntaopin ja historian tutkimusaineisto mahdollistaa ensimmäistä kertaa kahden erillisen oppiaineen välisen vertailun samojen oppilaiden ja koulujen tuotamalla aineistolla ja kahdella rinnakkaismittauksella, arvosanoilla ja seuranta-arvioinnissa osoitetulla osaamisella. Kerätty tutkimusaineisto ja kahden oppiaineen, yhteiskuntaopin ja historian, arvosanojen sekä ratkaisuosuuksien vertailu samojen oppilaiden tuloksilla tarjoaa mahdollisuuden tutkia lähemmin siitä, kuinka luotettava osaamisen mitta arvosana on.

Jotta arvosana ja seuranta-arviointi olisivat luotettavia osaamisen mittaamisessa, ne eivät saisi olla alttiita sattumanvaraisille eivätkä ainakaan systemaattisille virheille. Reliaabelius voidaan arvioida rinnakkaismittauksella, joka tässä tapauksessa muodostuu toisaalta yhteiskuntaopin ja historian seuranta-arvioinnista, toisaalta näiden oppiaineiden arvosanoista. Validius voidaan määritellä systemaattisten virheiden puuttumiseksi. Jos systemaattisia eroja on, ainakin toisen mittauksen validius voidaan kyseenalaistaa.

Tutkimuskysymykseen, miten oppimistulosten seuranta-arviointi ja arvosanat vastaavat toisiaan, pyritään vastaamaan luvussa 6.1 esitetyn tutkimusaineiston ja siitä johdettujen tulosten perusteella. Luvussa 6.2 on tarkoitus vastata kysymykseen esiintyykö kahden arvioinnin välillä mahdollisia systemaattisia eroja ja jos esiintyy, miten ne ilmenevät. Luvuissa 6.3 ja 6.4 esitetyn empiirisen aineiston pohjalta vastataan tutkimuskysymyksiin, voidaanko seuranta-arvioinnin perusteella tehdä johtopäätöksiä arvosanojen validiudesta vai jääkö kahden mittauksen validiusergelma ratkaisematta, sekä kysymykseen, voidaanko kahden oppiaineen samanaikaisella arvioinnilla ratkaista kahden mittauksen eli arvosanojen ja seuranta-arvioinnin validiutta koskeva ongelma.

6.1 Oppilaiden arvosanat ja seuranta-arvioinnin ratkaisuosuudet

Seuraavassa tarkastellaan oppilaiden arvosanoja ja seuranta-arvioinnissa saavutettuja ratkaisuosuuksia⁶⁶ sekä tutkimusongelman kannalta keskeisiä muuttujia. Lisäksi selvitetään, tuottavatko molemmat rinnakkaismittaukset, oppilaiden saamat arvosanat ja seuranta-arvioinnissa saavutetut ratkaisuosuudet, samansuuntaisia ja taustamuuttujiin nähden johdonmukaisia tuloksia. Seuranta-arvioinnin luotettavuuden tarkastelemiseksi tutkitaan ensin kouluarvosanaa, jota käytetään tässä seuranta-arvioinnin luotettavuuden rinnakkaismittarina ja sen jälkeen vertaillaan näitä kahta Perusopetuksen opetussuunnitelman perusteissa (2004) määritellyn osaamisen mittaria.

66 Pistemäärien sijasta yksittäisen tehtävän tai tehtäväosion tai koko seuranta-arvioinnin tulos ilmaistaan ratkaisuosuuksina tai ratkaisuprosentteina, koska seuranta-arvioinnin eri osa-alueiden enimmäispistemäärät vaihtelevat ja pistekeskisarvojen tulkinta ja vertailu olisi muuten hankalaa. Ratkaisuosuus on laskettu jakamalla saavutettu pistemäärä teoreettisella enimmäispistemäärällä ja muuntamalla saatu suhdeluku prosenteiksi. (Ouakrim-Soivio & Kuusela 2012, 28.)

6.1.1 Oppilaiden arvosanoista

Oppilaiden omaan ilmoitukseen perustuvien tietojen mukaan heidän (N = 4 103) viimeisimmän lukukausi- tai jaksotodistuksensa yhteiskuntaopin arvosanojensa keskiarvo oli 7,9. Poikien (N = 2 006) yhteiskuntaopin arvosanojen keskiarvo oli 7,6 ja tyttöjen (N = 2 097) 8,1. Historiassa viimeisin kouluarvosana oli oppilaiden keväällä 2010 saama päättöarvosana. Historian päättöarvosanojen keskiarvo oli kaikilla oppilailta (N = 4 110) 7,8. Pojilla (N = 2 011) historian päättöarvosanojen keskiarvo oli 7,8 ja tytöillä (N = 2 099) 7,9.

Samoilta oppilailta tiedusteltiin myös heidän viimeisintä äidinkielen ja matematiikan arvosanaansa, joka oli niin ikään joko lukukausi- tai jaksotodistukseen merkitty kouluarvosana. Oppilaiden (N = 4 120) matematiikan arvosanojen keskiarvo oli 7,6 (pojilla 7,5 ja tytöillä 7,7) ja oppilaiden (N = 4 119) äidinkielen arvosanojen keskiarvo oli 7,9 (pojilla 7,5 ja tytöillä 8,3).

Yhteiskuntaopin ja historian arvosanojen välinen korrelaatio oli vahva ($r_{xy} = 0,76$). Äidinkielen, matematiikan, historian ja yhteiskuntaopin arvosanojen keskinäiset korrelaatiot olivat vahvoja, kuten taulukosta 9 ilmenee. Yhteiskuntaopin ja historian arvosanat korreloivat vahvasti äidinkielen ja matematiikan arvosanojen kanssa. Oppilaat, jotka ovat saaneet hyviä arvosanoja yhdessä aineessa, ovat saaneet niitä toisissakin oppiaineissa.

TAULUKKO 9. Yhteiskuntaopin, historian, äidinkielen ja matematiikan arvosanojen keskinäinen korrelaatio.

	r_{xy}			
	Matematiikan arvosana	Äidinkielen arvosana	Historian arvosana	Yhteiskuntaopin arvosana
Matematiikan arvosana	1,00	0,62	0,59	0,61
Äidinkielen arvosana	0,62	1,00	0,61	0,65
Historian arvosana	0,59	0,61	1,00	0,76
Yhteiskuntaopin arvosana	0,61	0,65	0,76	1,00

Ristiintaulukoitaessa oppilaiden (N = 4 089) yhteiskuntaopin ja historian arvosanat saatiin tulokseksi, että yli 50 % oppilaista oli saanut saman arvosanan yhteiskuntaopissa ja historiassa. Kyseinen tulos koski kaikkia yhteiskuntaopin arvosanoja 5–10, kuten taulukosta 10 ilmenee.

TAULUKKO 10. Oppilaiden yhteiskuntaopin ja historian arvosanat.

		Oppilaiden viimeisin jakso- tai lukukausitodistukseen merkitty yhteiskuntaopin arvosana						
		4	5	6	7	8	9	10
Historian päättö-arvosana keväältä 2010	5	60 %	51 %	12 %	2 %			
	6	40 %	33 %	53 %	17 %	4 %		
	7		14 %	24 %	51 %	21 %	5 %	1 %
	8		2 %	10 %	24 %	52 %	25 %	6 %
	9			1 %	5 %	20 %	53 %	34 %
	10					2 %	15 %	59 %
Yhteensä		100 %	100 %	100 %	100 %	100 %	100 %	100 %

6.1.2 Oppilaiden ratkaisusuuksista ja arvosanoista

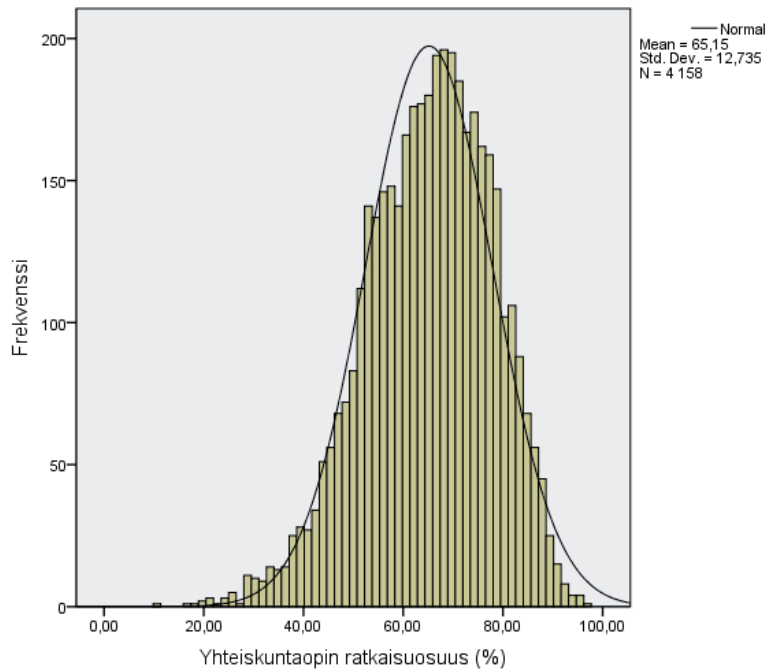
Taulukossa 11 on esitetty oppilaiden keskimääräiset arvosanat yhteiskuntaopissa (N = 4 103) ja historiassa (N = 4 109) sekä arvosanojen keskihajonta. Lisäksi taulukkoon on merkitty oppilaiden (N = 4 158) saavuttamat keskimääräiset ratkaisusuudet yhteiskuntaopissa (65 %) ja historiassa (51 %). Taulukosta 11 näkyy myös, että yhteiskuntaopin ratkaisusuuksien keskihajonta oli 12,7 ja historiassa se oli keskimäärin 13,8. Yhteiskuntaopin seuranta-arvioinnissa oppilaiden saavuttamat ratkaisusuudet vaihtelivat 11 %:sta 97 %:iin ja historiassa 6 %:sta 95 %:iin.

TAULUKKO 11. Yhteiskuntaopin ja historian arvosanojen keskiarvot, keskimääräiset ratkaisusuudet ja arvosanojen sekä ratkaisusuuksien hajonnat.

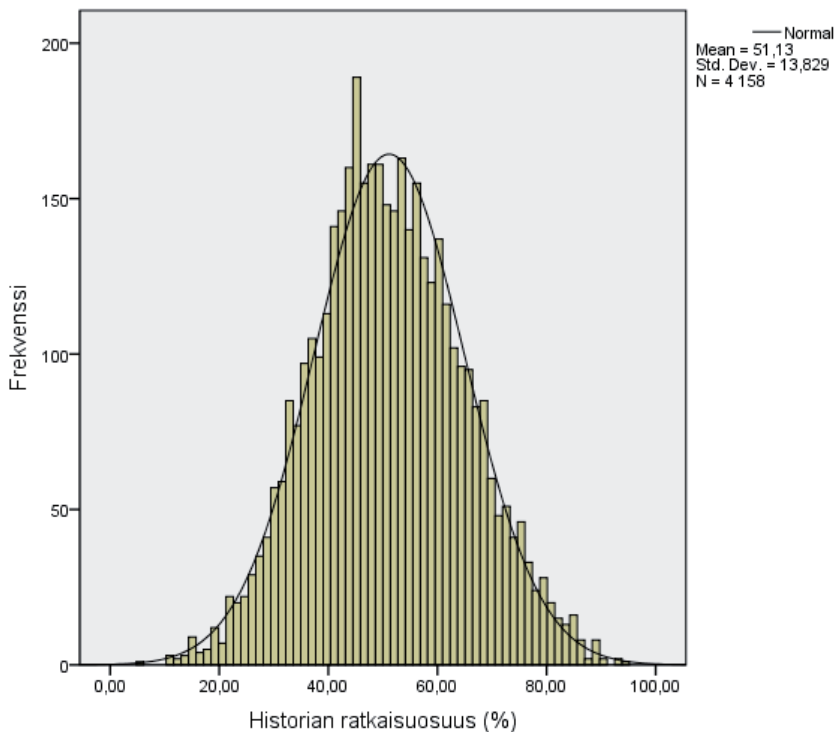
	N	Minimi	Maksimi	Keskiarvo	Hajonta
Yhteiskuntaopin arvosana	4103	4	10	7,8	1,2
Historian arvosana	4109	5	10	7,8	1,2
Yhteiskuntaopin ratkaisuosuus (%)	4158	10,6	97,0	65,2	12,7
Historian ratkaisuosuus (%)	4158	5,5	94,5	51,1	13,8

Kuviossa 4 on esitetty yhteiskuntaopin ratkaisuosuus ja kuviossa 5 on historian ratkaisuosuus histogrammeina. Yhteiskuntaopin keskimääräinen ratkaisuosuus oli 13 prosenttiyksikköä korkeampi kuin historian keskimääräinen ratkaisuosuus. Yhteiskuntaopin ja historian osaamista mittaavat tehtävät olivat erilaisia, sillä kummankin oppiaineen tehtäväsarjoilla pyrittiin mittaamaan Perusopetuksen opetussuunnitelman perusteissa (2004) näille kahdelle eri oppiaineille asetettuja tavoitteita, tietoja

ja taitoja. On kuitenkin mahdollista, että historiassa seuranta-arvioinnin tehtävät olivat joko vaativampia kuin yhteiskuntaopin tehtävät ja tästä syystä historian ratkaisuosuus jäi yhteiskuntaopin ratkaisuosuutta matalammaksi tai oppilaat osasivat viimeksi opiskelemiaan asioita eli yhteiskuntaoppia paremmin ja historiaa heikommien, koska historiaa oppilaat olivat opiskelleet viimeksi kahdeksannen vuosiluokan päättyessä eli vuotta aiemmin.

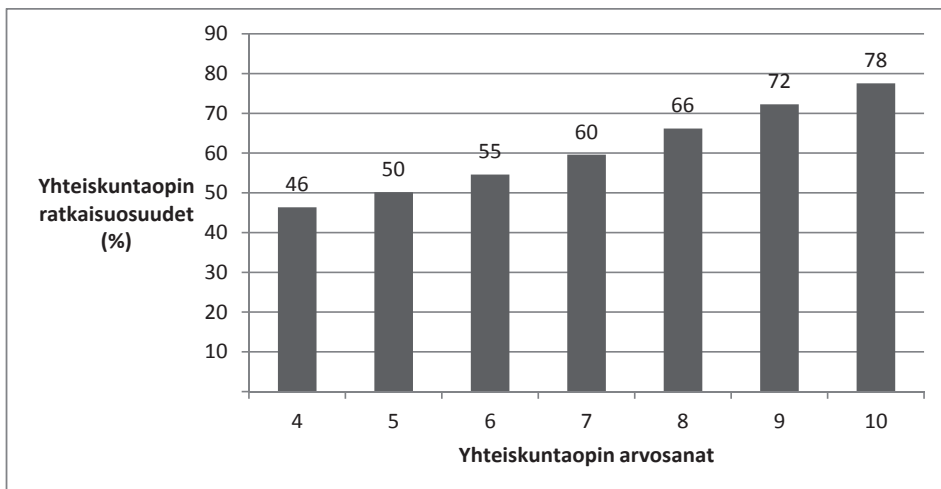


KUVIO 4. Yhteiskuntaopin osaamistaso ratkaisuprosentteina.



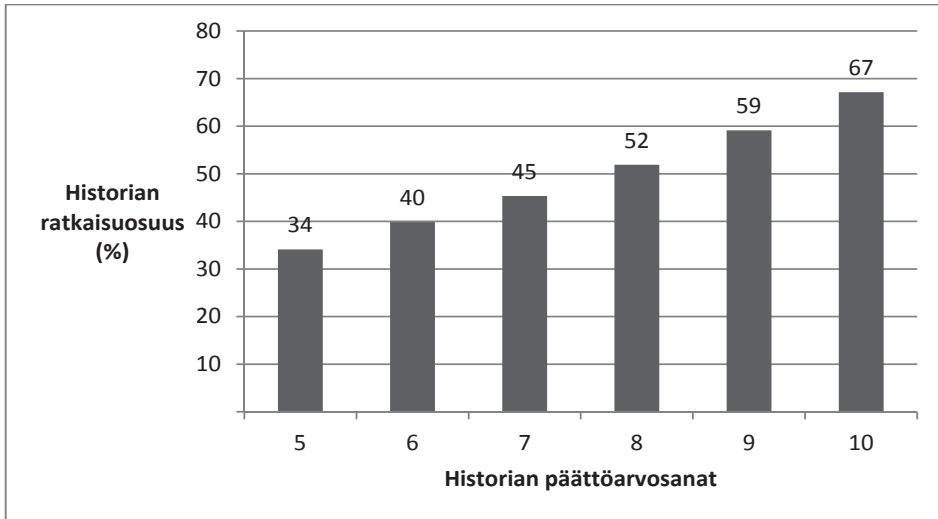
KUVIO 5. Historian osaamistaso ratkaisuprosenteina.

Kuviossa 6 on esitetty oppilaiden keskimääräiset yhteiskuntaopin seuranta-arvioinnissa saavutetut ratkaisuosuudet kouluarvosanoittain. Yhteiskuntaopin seuranta-arvioinnin ja arvosanojen välinen korrelaatio oli ($r_{XY} = 0,55$) ja sitä vastaava tilastollinen selitysosuus ($r^2 = 0,30$), jolla kuvataan yhteisen vaihtelun osuutta, oli 30 %.



KUVIO 6. Oppilaiden yhteiskuntaopin seuranta-arvioinnissa saamat ratkaisuprosentit arvosanoittain.

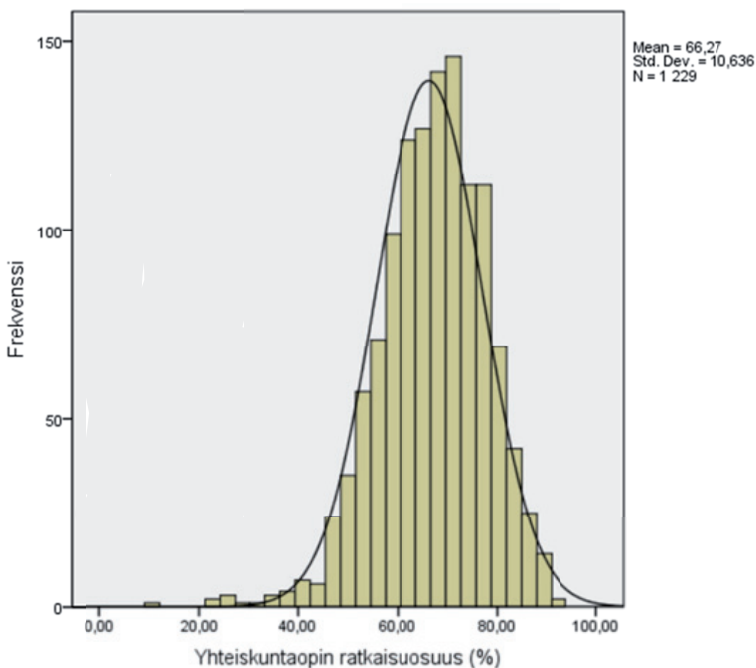
Kuviossa 7 on esitetty päättöarvosanoittain oppilaiden saamat keskimääräiset ratkaisuosuudet historian seuranta-arvioinnissa. Historiassa seuranta-arvioinnin ja arvosanojen välinen korrelaatio oli ($r_{xy} = 0,60$) ja sitä vastaava selitysosuus ($r^2 = 0,36$) oli 36 %:a yhteisestä vaihtelusta.



KUVIO 7. Oppilaiden historian seuranta-arvioinnissa saamat ratkaisuprosentit päättöarvosanoittain.

Kuvioista 6 ja 7 ilmenee, että oppilaiden seuranta-arvioinnista saavuttamansa keskimääräiset ratkaisuosuudet ja heidän saamansa yhteiskuntaopin ja historian arvosanat olivat johdonmukaisesti yhteydessä toisiinsa. Sekä yhteiskuntaopissa että historiassa oppilaat saivat keskimäärin sitä parempia ratkaisuosuuksia, mitä parempia arvosanoja he saivat yhteiskuntaopissa tai historiassa. Liitetaulukoista 3c ja 3d ilmenee, että kummassakin oppiaineessa kaikkien annettujen arvosanojen ratkaisuosuuksien keskihajonnat olivat samaa suuruusluokkaa.

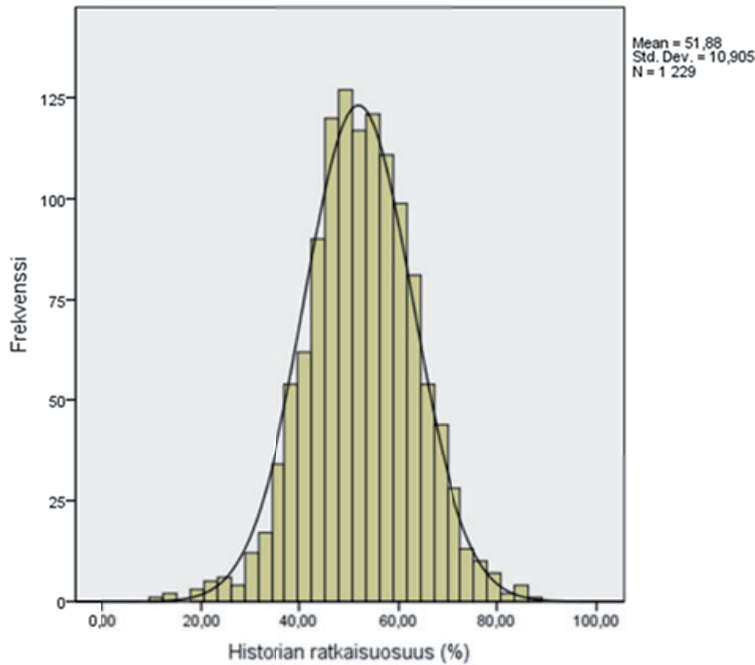
Kuviossa 8 on esitetty yhteiskuntaopin keskimääräiset ratkaisuosuudet arvosanalle 8. Perusopetuksen opetussuunnitelman perusteissa (2004) on kuvattu kriteerit arvosanalle 8, jonka on ajateltu toimivan opettajan mittatikkuna päättöarvosanoja annettaessa. Mikäli päättöarvioinnin kriteerit yhdenmukaistaisivat päättöarvosanoja, seuranta-arviointiin osallistuneiden oppilaiden, joiden viimeisin yhteiskuntaopin arvosana oli 8 ennen huhtikuun 2011 seuranta-arviointia, tulisi asettua kutakuinkin ratkaisuosuuksien 62,0 ja 69,2 välille kuten luvussa 6.3.3 myöhemmin esitetään ja perustellaan.⁶⁷ Kuitenkin kuviosta 13 on nähtävissä, että yhteiskuntaopin arvosanan 8 saaneiden oppilaiden seuranta-arvioinnissa saavuttamat ratkaisuosuudet vaihtelivat alle 20 %:sta yli 90 %:iin. Vaihtelua oli paljon eikä yhteiskuntaopin arvosana 8 näytä olleen poikkeus, kun sitä verrataan muihin oppilaiden saamiin yhteiskuntaopin arvosanoihin. Yhteiskuntaopissa arvosanan 8 ja historian päättöarvosanan 8 saaneiden oppilaiden seuranta-arvioinnissa saavuttamien ratkaisuosuuksien suuri hajonta voi viitata reliabelius- eli luotettavuusongelmaan, mutta taustalla voi olla jokin systemaattinen tekijä, joka aiheuttaa sen, että arvosanassa 8 ratkaisuosuuksien keskihajonta oli yhtä suurta kuin muissakin annetuissa arvosanoissa.



KUVIO 8. Yhteiskuntaopin arvosanan 8 saaneiden oppilaiden yhteiskuntaopin seuranta-arvioinnissa saavuttamat ratkaisuosuudet.

67 Tässä on kuvattu niiden oppilaiden ratkaisuosuuksia, jotka olivat saaneet viimeisimmässä arvioinnissa ennen huhtikuun 2011 seuranta-arviointia, yhteiskuntaopissa arvosanan 8.

Vertailun vuoksi kuviossa 9 on esitetty historian keskimääräiset ratkaisuosuudet päättöarvosanalle 8. Kuvioista 9 ilmenee, että historian päättöarvosanan 8 saaneiden oppilaiden historian ratkaisuosuudet vaihtelivat 10 %:sta 90 %:iin. Vaihtelua oli paljon myös historiassa eikä historian päättöarvosana 8 ollut poikkeus, kun sitä verrattiin muihin oppilaiden saamiin historian päättöarvosanoihin.



KUVIO 9. Historian päättöarvosanan 8 saaneiden oppilaiden jakauma suhteessa historian seuranta-arvioinnissa saatuihin ratkaisuosuuksiin.

Kun koulujen selitysosuus laskettiin arvosanan 8 saaneilla oppilailla, sisäkorrelaatio⁶⁸ tuotti tulokseksi 22 %. Se tarkoittaa, että kun taustamuuttujaksi valittiin koulu, arvosanan 8 saaneiden oppilaiden ratkaisuosuuksien eroista koulu selitti yli viidenneksen. Tämän tutkimuksen kannalta on kiinnostavaa, mikä koulun ominaisuus tuottaa 22 %:n vaihtelun koulujen välille.

68 Sisäkorrelaatio on koulujen eroja kuvaava mitta. Malin määrittelee sisäkorrelaation siten, että sillä tarkoitetaan ryhmien välisen varianssin osuutta kokonaisvariانسista. Muuttujan sisäkorrelaatio lasketaan jakamalla muuttujan kokonaisvariانسin kahteen osaan, ryhmien väliseen variانسin ja ryhmien sisäiseen variانسin. Sisäkorrelaatio vaihtelee yleensä välillä $0 \leq p \leq 1$. Määritelmä löytyy sivustolta: <https://www.jyu.fi/erillis/agoracenter/tutkimus/.../Monitasomallit.ppt>

6.1.3 Oppilaiden ratkaisuosuuksista ja arvosanoista eri taustamuuttujien mukaan luokiteltuna

Edellisessä luvussa esitettyjen tulosten perusteella näyttää siltä, että oppilaiden saamat arvosanat ja seuranta-arvioinnissa osoitettu osaaminen korreloivat oppilastasolla vahvasti ja kummatkin osaamisen mittarit näyttävät samaan suuntaan. Saatujen tulosten perusteella näyttäisi siltä, että kummatkin mittarit, seuranta-arviointi ja arvosanat, ovat valideja oppilastasolla, kun ne validiuden määritelmän mukaan konvergoivat keskenään (Kane 2006, 36–37; ks. Ketokivi 2009, 67–68.)

Seuraavaksi tarkastellaan taustamuuttujien yhteyttä oppilaiden saamiin arvosanoihin ja seuranta-arvioinnissa osoitettuun osaamiseen. Yleensä taustamuuttujat ovat johdonmukaisesti yhteydessä arvosanoihin ja seuranta-arvioinnissa osoitettuun osaamiseen. Esimerkiksi oppilaat, jotka tekevät säännöllisesti kotitehtävät, saavat keskimäärin parempia arvosanoja kuin ne oppilaat, jotka eivät oman ilmoituksensa mukaan tee kotitehtäviä koskaan tai melko harvoin. Osa taustamuuttujista on oppilaista riippumattomia, osa taas oppilaiden omista valinnoista johtuvia. Kun kyse on jälkimmäisistä, niin mahdollisen syy-yhteyden suunta ei aina ole selvä.

Taulukossa 12 on kuvattu eetan neliön (η^2) avulla tiivistetysti liitetaulukoiden 5a–7j sisältämä tieto siitä, kuinka paljon kyseinen taustamuuttuja selittää ratkaisuosuuksien ja arvosanojen eroista yhteiskuntaopissa ja historiassa. Tutkittuja taustamuuttujia olivat muun muassa sukupuoli, koulun opetuskieli sekä vanhempien koulustausta. Oppilaiden keskimääräiset arvosanat ja ratkaisuosuudet tarkastettiin entisen läänijaon ja tilastollisen kuntatyyppin mukaan luokiteltuna, mutta niissä tilastollinen selitysosuus oli 0 %. Tämän lisäksi tutkittiin, ovatko sellaiset taustamuuttujat kuten oppilaiden ensisijainen jatko-opintosuuntautumisvaihtoehto, annettujen kotitehtävien tekeminen, oppilaan harrastuneisuus ja se, miten huoltajat seuraavat läksyjen tekemistä, johdonmukaisessa yhteydessä oppilaiden ratkaisuosuuksiin ja arvosanoihin.

Yhteiskuntaopin ratkaisuosuuksien ja arvosanojen vertailemiseksi tutkimuksessa tarkasteltiin myös historian ratkaisuosuuksia ja arvosanoja edellä mainittujen taustamuuttujien mukaan luokiteltuna.

TAULUKKO 12. Yhteiskuntaopin ja historian selitysosuudet ratkaisuosuuksien ja arvosanojen suhteen eri taustamuuttujien mukaan tarkasteltuna.⁶⁹

Taustamuuttuja	Selitysoisuus (η^2) ratkaisuosuuksissa	Selitysoisuus (η^2) arvosanoissa
Sukupuoli	Yhteiskuntaopissa 1 % *** Historiassa 0 % n.s.	Yhteiskuntaopissa 3 % *** Historiassa 0 % *
Koulun opetuskieli	Yhteiskuntaopissa 3 % *** Historiassa 2 % ***	Yhteiskuntaopissa 0 % n.s. Historiassa 0 % n.s.
Vanhempien koulutustausta	Yhteiskuntaopissa 4 % *** Historiassa 7 % ***	Yhteiskuntaopissa 7 % *** Historiassa 9 % ***
Ensisijainen jatko-opintosuuntautumismuutos	Yhteiskuntaopissa 16 % *** Historiassa 18 % ***	Yhteiskuntaopissa 24 % *** Historiassa 26 % ***
Annettujen kotitehtävien tekeminen	Yhteiskuntaopissa 7 % *** Historiassa 7 % ***	Yhteiskuntaopissa 11 % *** Historiassa 9 % ***
Huoltajat seuraavat annettuja kotitehtäviä	Yhteiskuntaopissa 1 % *** Historiassa 1 % ***	Yhteiskuntaopissa 1 % *** Historiassa 1 % ***
Historian ja yhteiskuntaopin aihealueiden harrastaminen	Yhteiskuntaopissa 2 % *** Historiassa 8 % ***	Yhteiskuntaopissa 3 % *** Historiassa 6 % ***

Molemmat mittaukset, arvosanat ja seuranta-arviointien ratkaisuosuudet, olivat johdonmukaisessa yhteydessä sellaisiin taustamuuttujiin, joihin niiden aikaisempien tutkimusten perusteella saattoi olettaakin olevan yhteydessä. Seuraavassa on kuitenkin nostettu esiin kaksi taustamuuttujaa: opetuskieli ja annettujen kotitehtävien tekeminen. Koulun opetuskielen ollessa taustamuuttujana huomio kiinnittyy taulukossa 12 siihen, että koulun opetuskieli selittää seuranta-arvioinnissa saavutettua keskimääräistä ratkaisuosuutta, mutta ei lainkaan keskimääräisiä arvosanoja.

Koulun opetuskielen 3 %:n selitysoisuus yhteiskuntaopin ratkaisuosuuksissa oli melko suuri. Suomenkielisten koulujen oppilaiden ($N = 3\ 613$) ja ruotsinkielisten koulujen oppilaiden ($N = 545$) välillä ei ollut eroja annetuissa yhteiskuntaopin arvosanoissa.

Suomenkielisten koulujen oppilaiden yhteiskuntaopin arvosanojen keskiarvo oli 7,9 ja ruotsinkielisten koulujen oppilaiden arvosanojen keskiarvo oli 7,8, eivätkä erot olleet tilastollisesti merkitseviä ($F = 0,74$, $df = 1, 4\ 102$; $p > 0,05$). Suomenkielisten koulujen oppilaiden yhteiskuntaopin ratkaisuosuus (66 %) oli 7 prosenttiyksikköä ruotsinkielisten koulujen oppilaiden (ratkaisuosuus 59 %) ratkaisuosuutta korkeampi. Suomen- ja ruotsinkielisten koulujen oppilaiden ratkaisuosuuksien erot olivat tilastollisesti erittäin merkitseviä ($F = 141,25$; $df = 1, 4\ 159$; $p < 0,001$).

69 Taulukkoon 12 on merkitty tähdityksellä merkitsevyydestä tulos seuraavasti: (***) tarkoittaa tilastollisesti erittäin merkitsevää, jolloin $p < 0,001$, (**) tarkoittaa tilastollisesti merkitsevää, jolloin $p < 0,01$, (*) tarkoittaa tilastollisesti melkein merkitsevää, jolloin $p < 0,05$ ja merkintä (n.s.) tarkoittaa, ettei tilastollista merkitsevyyttä ole, jolloin $p > 0,05$.

Taulukosta 12 on nähtävissä, että annettujen kotitehtävien säännöllinen tekeminen selitti taustamuuttujana enemmän yhteiskuntaopin ja historian arvosanoja kuin kyseisten oppiaineiden ratkaisuosuuksia. Suurimmat ratkaisuosuuksien erot yhteiskuntaopissa (10 %) muodostuivat niiden oppilaiden (N = 1 234) välille, jotka ilmoittavat tekevänsä aina annetut kotitehtävät ja niiden oppilaiden (N = 245) välille, jotka ilmoittivat, etteivät he tee juuri koskaan annettuja kotitehtäviä. Yhteiskuntaopin ratkaisuosuuksien erot olivat tilastollisesti erittäin merkitseviä (F = 101,89, df = 3; 4 148, p < 0,001). Myös yhteiskuntaopin arvosanoissa oli tilastollisesti erittäin merkitsevät erot (F = 176,33, df = 3; 4093, p < 0,001) niiden oppilaiden (N = 1 226) hyväksi, jotka ilmoittivat tekevänsä aina annetut kotitehtävät, verrattuna niihin, jotka ilmoittivat, etteivät he tee juuri koskaan annettuja kotitehtäviä.

Yhteiskuntaopin arvosanojen ero (1,2 arvosanaa) muodostui suurimmaksi aina läksynsä tekevien oppilaiden ja silloin tällöin annetut kotitehtävät tekevien oppilaiden (N = 745) välille. Post hoc -testin (Tukeyn b)⁷⁰ tuloksena tilastollisesti merkitsevät erot muodostuivat sekä yhteiskuntaopin arvosanoissa että ratkaisuosuuksissa sen mukaan, kuuluivatko oppilaat joihinkin seuraavista kolmesta ryhmästä: aina läksynsä tekevät, usein kotitehtävänsä tekevät tai ei juuri koskaan tai silloin tällöin kotitehtävänsä tekevät.

Historiassa ratkaisuosuuksien ero oli täsmälleen samansuuruinen (10 %) aina tekevänsä tekevien oppilaiden eduksi kuin yhteiskuntaopissa ja historian arvosanojen keskiarvoissa oli 1,1 arvosanan verran eroa aina kotitehtävänsä tekevien oppilaiden (N = 1 225) ja ei juuri koskaan läksynsä tekevien oppilaiden (N = 242) välillä. Oppilasryhmien väliset erot olivat myös historian ratkaisuosuuksissa tilastollisesti erittäin merkitseviä (F = 99,11, df = 3, 4 148, p < 0,001). Samoin näiden oppilasryhmien väliset erot historian arvosanoissa olivat tilastollisesti erittäin merkitseviä (F = 142,16, df = 3; 4 099, p < 0,001). Post hoc testin tuloksena syntyivät tilastollisesti merkitsevät erot sekä historian arvosanoissa että ratkaisuosuuksissa sen mukaan, tekivätkö oppilaat annetut tehtävät aina, usein tai vain silloin tällöin tai ei koskaan.

Tutkituista taustamuuttujista ja niiden yhteydestä oppilaiden saamiin yhteiskuntaopin ja historian arvosanoihin ja ratkaisuosuuksiin huomio kiinnittyi koulun opetuskieleen ja kotitehtävien tekemiseen. Suomen- ja ruotsinkielisten koulujen oppilaiden välillä oli eroja yhteiskuntaopin ja historian seuranta-arvioinneissa osoitettuun osaamiseen sikäli, että kummassakin oppiaineessa opetuskieli selitti seuranta-arvioinnissa osoitettua osaamista, mutta ei kyseisten oppiaineiden arvosanoja. Oppilaiden väliset erot annettujen kotitehtävien tekemisen mukaan luokiteltuna selittivät puolestaan enemmän oppilaiden välisiä yhteiskuntaopin ja historian arvosanaeroja kuin näiden oppiaineiden ratkaisuosuuksia.

70 Tehtäessä SPSS-ohjelmalla post hoc -testausta, ohjelma antaa toistakymmentä eri vaihtoehtoa tehdä parittaisvertailutesti. Tukeyn b-testi on yksi yleisimmistä parittaisvertailutesteistä, joka järjestää ryhmäkeskiarvot osaryhmiksi, jotka eivät eroa toisistaan. Määritelmä Tukeyn b-testiin löytyy sivustolta: <http://www.uku.fi/~mauranen/spss/jatko/anova18.html>

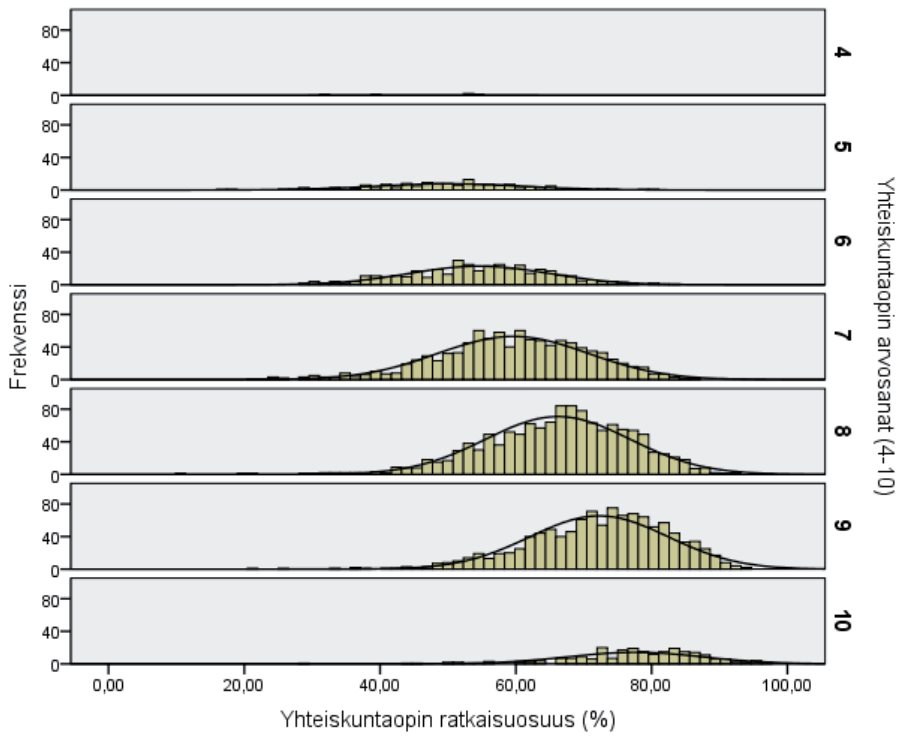
6.2 Seuranta-arvioinnissa menestyminen sekä arvosanat yhteiskuntaopissa ja historiassa

Seuraavassa tarkastellaan oppilaiden saamia yhteiskuntaopin ja historian ratkaisuosuuksia ja arvosanoja oppilastasolla, koulujen sisällä sekä koulujen välillä, jotta voidaan selvittää, esiintyykö näiden kahden rinnakkaisen osaamisen mittarin, arvosanojen ja seuranta-arvioinnin välillä systemaattisia eroja ja jos niitä ilmenee, niin miten.

6.2.1 Oppilastasolla

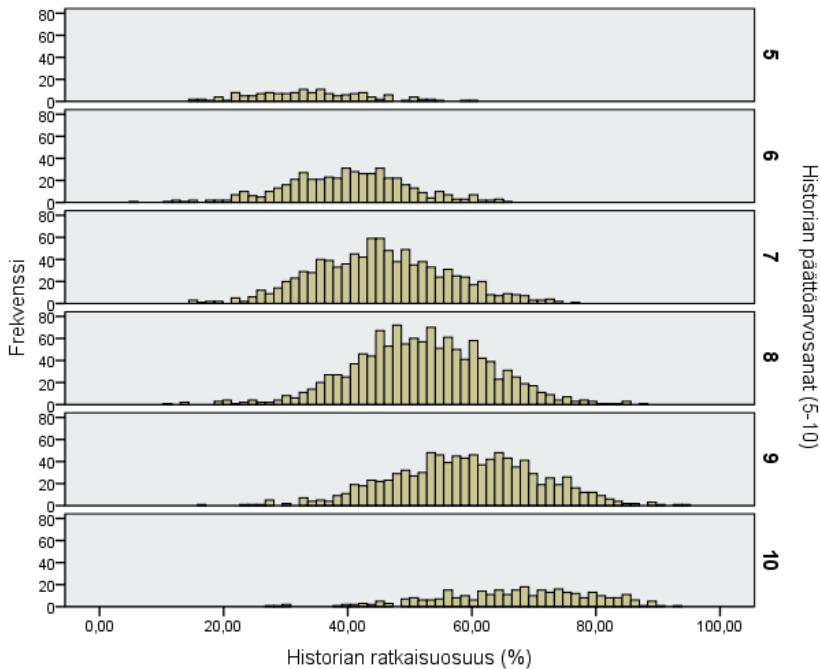
Kuten luvussa 6.1 on jo esitetty, oppilaiden yhteiskuntaopin keskimääräiset arvosanat ja seuranta-arvioinnissa saavutetut keskimääräiset ratkaisuosuudet korreloivat keskenään melko vahvasti ($r_{xy} = 0,55$), samoin oppilaiden historian päättöarvosanat ja historian ratkaisuosuudet, jossa korrelaatio oli 0,60. Myös yhteiskuntaopin ja historian ratkaisuosuudet korreloivat vahvasti keskenään ($r_{xy} = 0,74$), samoin yhteiskuntaopin ja historian arvosanat ($r_{xy} = 0,76$).

Oppilaiden seuranta-arvioinnissa saavuttamat ratkaisuosuudet suhteessa annettuihin kouluarvosanoihin on esitetty yksityiskohtaisemmin yhteiskuntaopissa kuviossa 10 ja historiassa kuviossa 11. Kuviosta 10 näkyy, että niiden oppilaiden, jotka saivat yhteiskuntaopin arvosanaksi esimerkiksi 7, saavuttamat ratkaisuosuudet yhteiskuntaopin seuranta-arvioinnissa vaihtelivat alle 20 %:sta hieman yli 80 %:iin. Vastaa vasti yhteiskuntaopissa kiitettävän arvosanan 9 saaneiden oppilaiden ratkaisuosuudet vaihtelivat noin 30 %:sta lähes 90 %:iin. Myös arvosanan 10 saaneiden oppilaiden ratkaisuosuuksissa oli vaihtelua noin 50 %:sta lähes 90 %:iin. Yhteiskuntaopin arvosanan 4 saaneiden oppilaiden ratkaisuosuuksien vaihtelu näkyy heikosti kuviossa 10, koska kyseisen arvosanan saaneita oppilaita oli vain viisi, kuten liitetäulukossa 3c on esitetty.



KUVIO 10. Oppilaiden yhteiskuntaopin ratkaisuosuuksien jakaumat suhteessa heidän saamiinsa yhteiskuntaopin arvosanoihin.

Kuviosta 11 havaitaan, että kaikissa oppilaille annetuissa historian päättöarvosanoissa historian ratkaisuosuudet vaihtelivat suuresti, myös päättöarvosanassa 8. Historian päättöarvosanassa 8 oppilaiden seuranta-arvioinnissa saavuttamat ratkaisuosuudet vaihtelevat keskimäärin 20 %:sta 85 %:iin ja esimerkiksi arvosanan 7 saaneiden ratkaisuosuudet vaihtelevat 18 %:sta 78 %:iin. Kiitettävän historian päättöarvosanan saaneiden oppilaiden ratkaisuosuudet vaihtelevat puolestaan arvosanassa 9 noin 35 %:sta hieman yli 85 %:iin ja päättöarvosanassa 10 noin 40 %:sta hieman yli 90 %:iin.



KUVIO 11. Oppilaiden historian ratkaisuosuuksien jakaumat suhteessa heidän saamiinsa historian päättöarvosanoihin.

Jo aiemmin luvussa 6.1.2 kiinnitettiin huomiota siihen, että oppilaiden ratkaisuosuuksien keskihajonnat olivat suuria. Taulukkoon 13 on kerätty tiivistetysti yhteiskuntaopin ja historian arvosanat, niitä vastaavat keskimääräiset ratkaisuosuudet sekä niiden keskihajonnat.

TAULUKKO 13. Yhteiskuntaopin ja historian keskimääräiset ratkaisuosuudet arvosanoittain, kyseisen arvosanan saaneiden oppilaiden määrä ja ratkaisuosuuksien keskihajonnat.

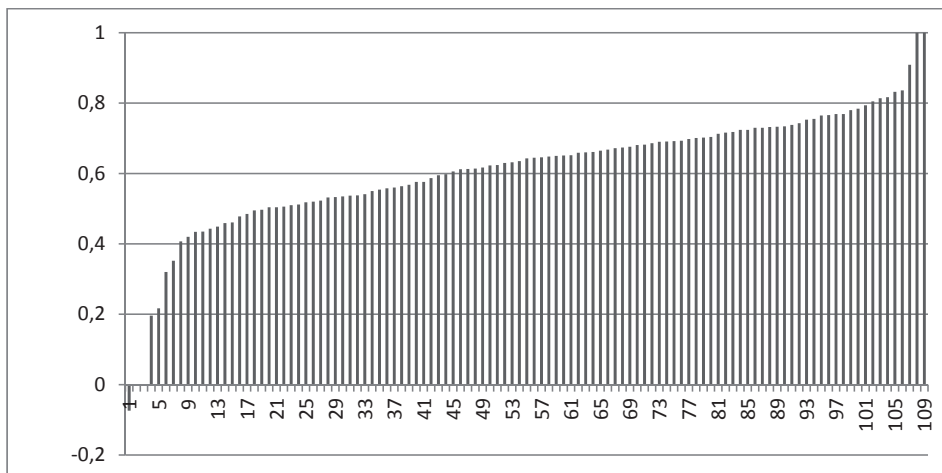
Yhteiskuntaopin arvosana	Keskimääräinen ratkaisuosuus (%)	N	Keskihajonta
4	46,4	5	10,2
5	50,2	140	11,1
6	54,6	393	10,4
7	59,6	972	11,1
8	66,2	1 252	10,7
9	72,3	1 104	10,2
10	77,5	237	10,1
Historian arvosana	Ratkaisuosuus (%)	N	Keskihajonta
5	34,1	152	9,4
6	39,8	481	9,9
7	45,3	943	10,6
8	51,9	1 229	10,9
9	59,1	974	11,7
10	67,1	330	12,4

Taulukosta 13 ilmenee yksityiskohtaisemmin, sama kuin edellä esitetystä kuvioista 10 ja 11, että oppilaiden saavuttamien ratkaisuosuuksien keskihajonta yhteiskuntaopin arvosanalla 8 tai historian päättöarvosanalla 8 on vähintäänkin yhtä suurta kuin muiden oppilaiden saamien arvosanojen ja niitä vastaavien keskimääräisten ratkaisuosuuksien suhteen.

Oppilaiden saavuttamissa keskimääräisissä ratkaisuosuuksissa keskihajonta oli suurta kaikissa annetuissa yhteiskuntaopin ja historian kouluarvosanoissa.

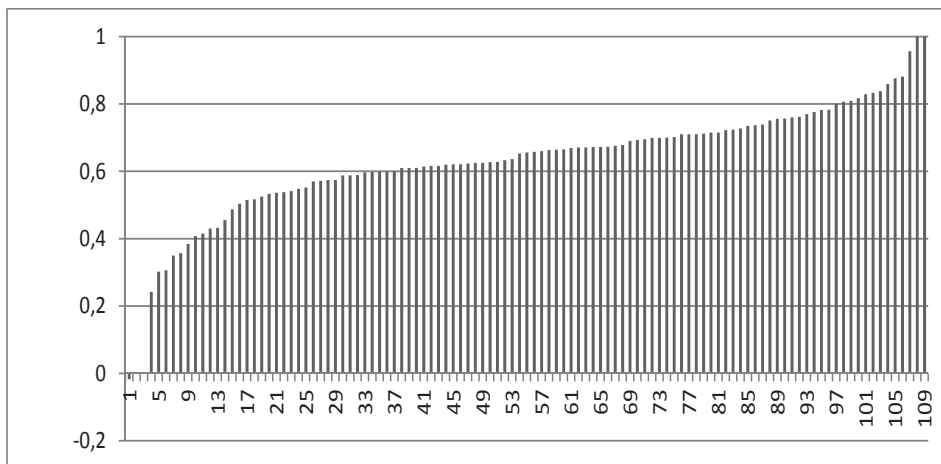
6.2.2 Koulujen sisällä

Kuviossa 12 on esitetty kouluittain (N = 109) koulujen sisäiset korrelaatiot yhteiskuntaopin seuranta-arvioinnissa osoitetun osaamistason eli ratkaisuosuuksien ja kouluissa annettujen yhteiskuntaopin arvosanojen välillä.



KUVIO 12. Yhteiskuntaopin arvosanojen ja ratkaisuosuuksien korrelaatiot kouluittain laskettuna.

Kuviossa 13 on esitetty kouluittain (N = 109) niiden sisäiset korrelaatiot historian seuranta-arvioinnissa osoitetun osaamistason ja kouluissa annettujen historian arvosanojen välillä.



KUVIO 13. Historian arvosanojen ja ratkaisuosuuksien korrelaatiot kouluittain laskettuna.

Koulujen sisäiset korrelaatiot sekä yhteiskuntaopin että historian arvosanojen ja seuranta-arvioinnissa osoitetun osaamisen välillä näyttävät kuvioiden 12 ja 13 mukaan asettuvan 0,40 ja 0,80 välille. Koulujen sisäisten korrelaatioiden keskiarvot olivat varsin korkeat: yhteiskuntaopissa ($r_{XY} = 0,61$) ja historiassa ($r_{XY} = 0,63$).

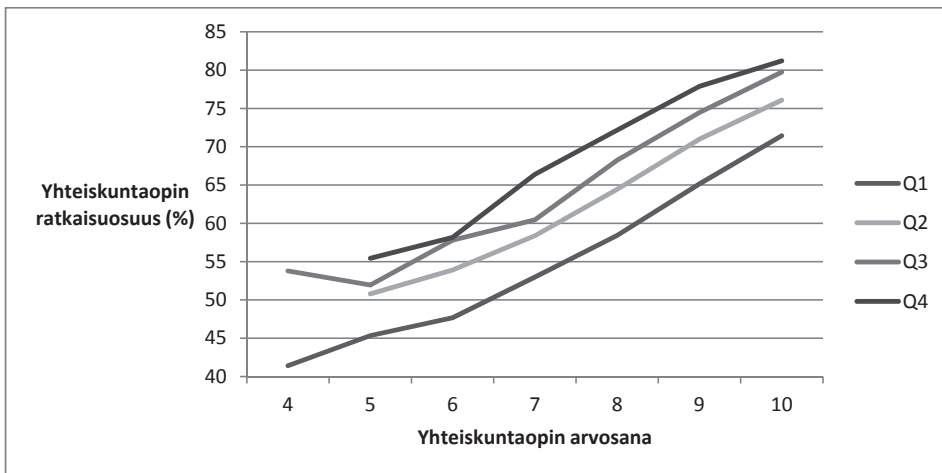
Koulujen sisäiset korrelaatiot osoittavat, että kun on osattu seuranta-arvioinnissa hyvin, on myös saatu hyviä arvosanoja ja kääntäen. Vain yhdessä koulussa korrelaatio oli negatiivinen sekä yhteiskuntaopissa ($r_{XY} = -0,07$) että historiassa ($r_{XY} = -0,02$). Kahdessa koulussa korrelaatio oli yksi sekä yhteiskuntaopissa että historiassa, mutta kysymys oli kahden oppilaan tiedoista kummassakin koulussa, jolloin korrelaatioksi voi tulla yksi tai miinus yksi.

6.2.3 Koulujen välillä

Opetushallituksen julkaisema yhteiskunnallisten aineiden seuranta-arviointi osoitti, että kun koulut luokiteltiin neljään osaamiskvartiiliin (Q1–Q4) niiden yhteiskuntaopissa ja historiassa osoittaman seuranta-arviointimenestyksen mukaan, kummasakin oppiaineessa oli havaittavissa, että kouluissa (Q4), joissa yhteiskuntaopin tai historian seuranta-arvioinneissa menestytettiin keskimäärin hyvin, annettiin yhteiskuntaopissa ja historiassa samoilla ratkaisuosuuksilla alempia arvosanoja kuin niissä kouluissa (Q1), joissa keskimääräiset ratkaisuosuudet yhteiskuntaopissa ja historiassa jäivät matalammiksi (Ouakrim-Soivio & Kuusela 2012, 110–111).

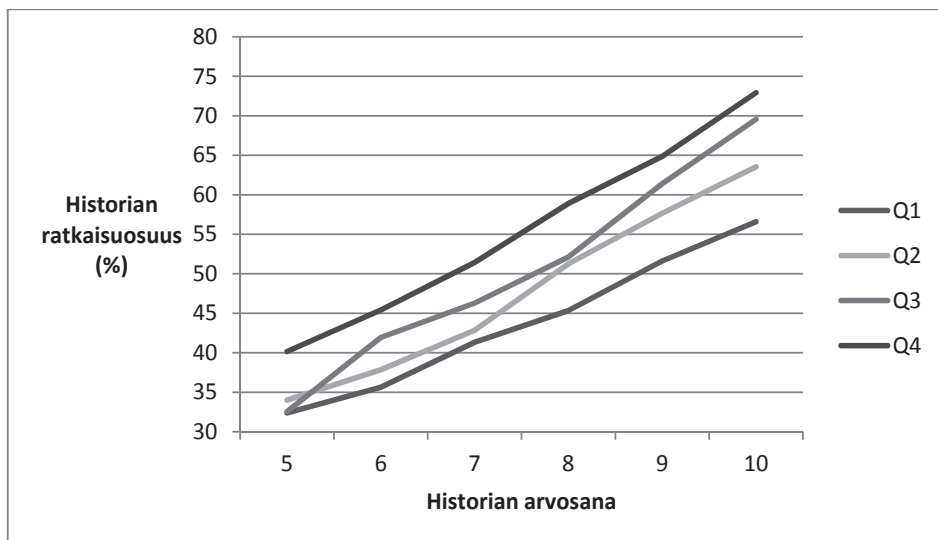
Koska tästä tutkimusaineistosta poistettiin erityistä tukea saaneiden oppilaiden ja muuta kuin suomea tai ruotsia äidinkielenään puhuvien oppilaiden seuranta-arviointitulokset, on kuvioissa 14 ja 15 havainnollistettu tällä tutkimusaineistolla koulujen ($N = 109$) yleisopetuksen oppilaiden ($N = 4\,158$) yhteiskuntaopin ja historian arvosanat sekä heidän saamansa ratkaisuosuudet yhteiskuntaopissa ja historiassa. Kuviossa 14 on esitetty koulujen väliset arvosana- ja osaamistasoerot niin, että jokaisesta koulusta on laskettu yleisopetuksen oppilaiden keskimääräinen osaamistaso yhteiskuntaopissa eli oppilaiden yhteiskuntaopin ratkaisuosuuksien keskiarvo. Tämän jälkeen koulut on luokiteltu järjestyksessä neljään ryhmään osaamiskvartiileittain (Q1–Q4) sen mukaan, mikä koulun keskimääräinen osaaminen oli yhteiskuntaopin seuranta-arvioinnissa. Kuvioon merkitty ryhmä Q1 tarkoittaa yhteiskuntaopin seuranta-arvioinnissa heikoimmin menestyneeseen neljännekseen kuuluvia kouluja ja Q4 tarkoittaa kouluja, jotka menestyivät seuranta-arvioinnissa parhaiten. Kuvioissa 14 ja 15 on x-akseleille merkitty oppiaineiden arvosanat ja y-akseleille niitä vastaavat keskimääräiset ratkaisuosuudet.

Kuvioista 14 ja 15 on havaittavissa, että kouluissa (Q4), joissa yhteiskuntaopin tai historian seuranta-arvioinneissa menestyttiin keskimäärin hyvin, annettiin yhteiskuntaopissa ja historiassa samoilla ratkaisuosuuksilla alempia arvosanoja kuin niissä kouluissa (Q1), joissa keskimääräiset ratkaisuosuudet yhteiskuntaopissa ja historiassa jäivät matalammiksi. Käytännössä tämä tarkoittaa sitä, että esimerkiksi yhteiskuntaopissa (kuvio 14) ratkaisuosuudella 55 % sai sellaisessa koulussa, joka kuului keskimääräisen osaamisensa puolesta parhaimpaan neljännekseen (Q4), keskimäärin arvosanan 5 ja heikoimmin menestyneeseen neljännekseen kuuluvassa koulussa (Q1) arvosanan 7.



KUVIO 14. Koulujen osaamistaso arvosanoittain yhteiskuntaopissa.

Historiassa koulut jaettiin samalla tavalla kuin yhteiskuntaopissakin seuranta-arvioinnissa keskimäärin osoitetun osaamistason mukaan neljään osaamiskvartiiliin (Q1–Q4). Kuvioista 15 näkyy, että kyseessä ei ollut ainoastaan yhteiskuntaopin ratkaisuosuuksien ja annettujen arvosanojen välisestä erosta, vaan sama ilmiö on havaittavissa myös historiassa. Kuvion 15 mukaan esimerkiksi historiassa ratkaisuosuudella 45 % Q4-ryhmään kuuluvassa koulussa historian päättöarvosana oli keskimäärin 6 ja sellaisessa koulussa, joka kuului Q1-ryhmään, päättöarvosana oli samalla 45 %:n ratkaisuosuudella keskimäärin 8. Sekä yhteiskuntaopissa että historiassa seuranta-arvioinnin ratkaisuosuuksien erot olivat suuria kaikissa kouluarvosanoissa ja kummassakin oppiaineessa.



KUVIO 15. Koulujen osaamistaso arvosanoittain historiassa.

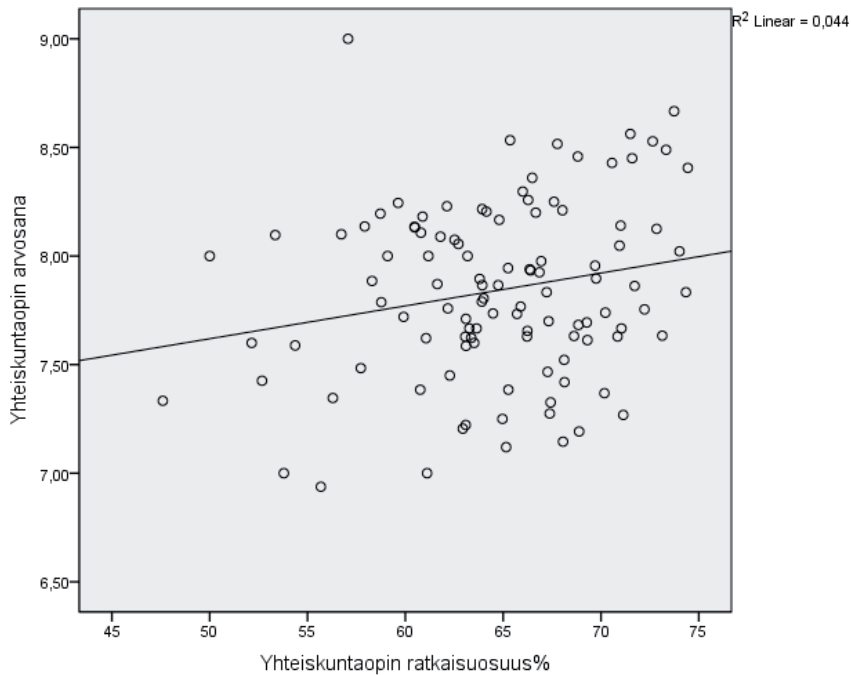
Kuten aiemmin tässä tutkimuksessa on esitetty, Perusopetuksen opetussuunnitelman perusteiden (2004) mukaan määritellyt osaamisen mittarit, arvosanat ja seuranta-arviointi, näyttävät antavan viitteitä sekä yhteiskuntaopissa että historiassa siitä, että kahdessa oppiaineessa samalla osaamistasolla eli seuranta-arvioinnissa saavutetuilla ratkaisuosuuksilla oppilaat saivat erilaisia arvosanoja, kuten kuvioissa 14 ja 15 on esitetty. Saman asian voi ilmaista toisin niin, että tietyllä arvosanalla oppilaiden saamat keskimääräiset ratkaisuosuudet vaihtelivat suuresti, yhteiskuntaopissa arvossana 8 ero oli jopa 14 prosenttiyksikköä. Historiassa puolestaan arvossanan 10 ratkaisuosuuksien ero oli suurimmillaan ryhmiin Q1 ja Q4 kuuluvien koulujen välillä 16 prosenttiyksikköä. Kuvioiden 14 ja 15 avulla voidaan päätellä, että koulujen välillä on eroja.

Kuvioista 14 ja 15 ei voi kuitenkaan päätellä sitä, kumpi rinnakkaisista osaamisen mittareista, arvosanat vai seuranta-arviointi, kuvaa validimmin koulujen osaamistasoa. Yhteiskunnallisten aineiden tutkimusaineisto antoi mahdollisuuden laskea, kuinka paljon koulujen kokonaisvaihtelusta oli koulujen sisäistä vaihtelua ja kuinka paljon kyse oli koulujen välisestä vaihtelusta. Liitetaulukoissa 8a ja 8b on esitetty koulujen väliset erot yhteiskuntaopissa ja historiassa sekä seuranta-arvioinnista että arvosanoista laskettuina. Koulujen väliset erot olivat seuranta-arvioinnissa osoitetussa yhteiskuntaopin osaamisessa 13,9 %. Historiassa koulujen väliset erot olivat seuranta-arvioinnissa osoitetussa osaamisessa puolestaan 12,5 %. Kun sisäkorrelaatio laskettiin yhteiskuntaopin arvosanoista, koulujen väliset erot olivat enää 6,7 % ja historiassa vain 5,1 %. Sisäkorrelaatiolla laskettuna sekä yhteiskuntaopissa että historiassa koulujen välinen vaihtelu oli suurta seuranta-arvioinnissa osoitetussa osaamisessa ja selvästi pienempää, kun tarkasteltiin koulujen keskimääräisiä arvosanoja.

Näyttää siltä, että arvosanat taipuvat kohti keskimääräistä. Koulujen väliset erot arvosanoista laskettuna olivat noin puolet pienempiä kuin seuranta-arviointien ratkaisuosuuksista laskettuna. Kuuselan selvitys tukee edellä esitettyä yhteiskuntaopin ja historian sisäkorrelaatiolla laskettua tulosta siitä, että koulujen erot näyttäivät olevan suurempia seuranta-arviointituloksissa kuin annettujen arvosanojen perusteella (Kuusela 2012,18–19).

Oppilaiden saamat yhteiskuntaopin ja historian arvosanat korreloivat vahvasti seuranta-arvioinnissa saavutettujen ratkaisuosuuksien kanssa, kuten jo aiemmin luvussa 6.2.1 ja 6.2.2 on osoitettu. Koulujen sisäiset korrelaatiot esimerkiksi yhteiskuntaopissa olivat kahta poikkeusta lukuun ottamatta korkeammat kuin koulujen keskimääräisten arvosanojen korrelaatiot suhteessa keskimääräiseen seuranta-arviointimenestykseen. Koulujen tasolla yhteiskuntaopin keskimääräiset ratkaisuosuudet korreloivat enää heikosti koulujen keskimääräisiin arvosanoihin nähden. Korrelaatio oli yhteiskuntaopissa enää 0,21 ja historiassakin vain 0,38.

Kuviossa 16 on esitetty kouluissa annettujen yhteiskuntaopin arvosanojen keskiarvon ja koulujen yhteiskuntaopin seuranta-arvioinnissa saavuttamien keskimääräisten ratkaisuosuuksien välinen suhde. Kuvioista 16 on nähtävissä, että yhteiskuntaopissa yhteistä vaihtelua oli vain 4,4 %. Kuvion perusteella näyttää siltä, että koulujen yhteiskuntaopin seuranta-arvioinnissa osoitetun osaamisen ja annettujen yhteiskuntaopin arvosanojen välillä ei olisi juuri yhteyttä. Esimerkiksi koulut, joiden keskimääräiset yhteiskuntaopin arvosanat olivat lähellä arvosanaa 7, ovat saaneet yhteiskuntaopin ratkaisuosuuksiksi 53 %:sta lähes 70 %:iin. Vastaavasti koulut, joiden antamien yhteiskuntaopin arvosanojen keskiarvo on ollut 8, ovat saaneet seuranta-arvioinnissa keskimääräisiksi ratkaisuosuuksiksi 52 %:sta 74 %:iin.

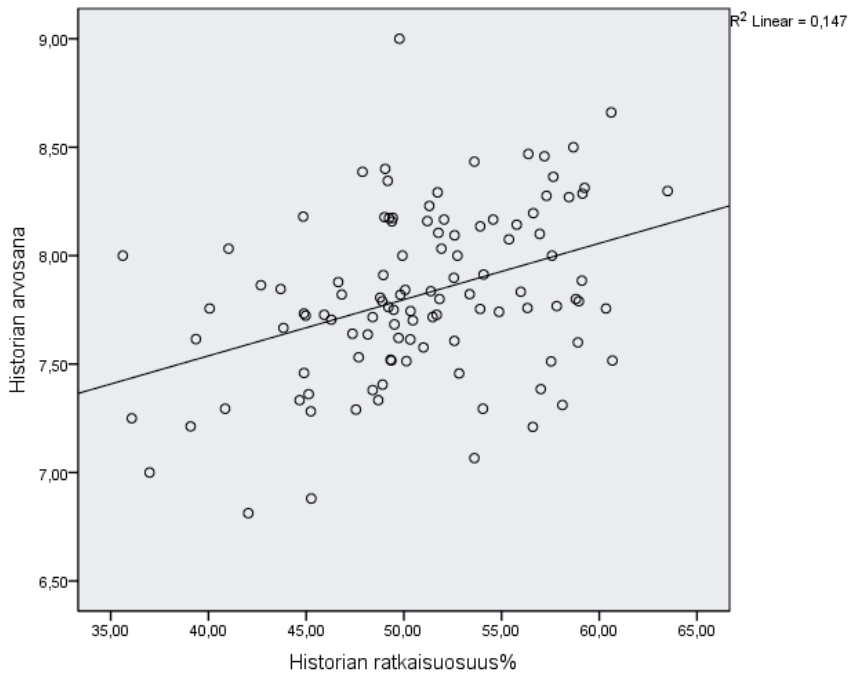


KUVIO 16. Kouluissa annettujen yhteiskuntaopin arvosanojen keskiarvon suhde koulujen keskimääräisiin yhteiskuntaopin ratkaisuosuuksiin.

Kuvion 16 tulos on odottamaton, koska yhteiskuntaopin seuranta-arviointi ja arvosanat konvergoivat oppilastasolla ja koulujen sisällä, kun konvergenssin mittana pidetään korrelaatioita. Seuranta-arviointi korreloi positiivisesti oppilaiden arvosanojen kanssa yli koko aineiston ($r_{XY} = 0,55$) sekä kouluittain laskettuihin arvosanoihin keskimäärin ($r_{XY} = 0,61$), kuten jo aiemmin on esitetty. Mutta koulujen tasolla kahden rinnakkaisen mittauksen, yhteiskuntaopin seuranta-arvioinnin ja yhteiskuntaopin arvosanojen, konvergenssi heikkenee huomattavasti ($r_{XY} = 0,21$).

Historiassa tulokset olivat samanlaiset kuin yhteiskuntaopissa. Oppilastasolla seuranta-arviointi ja arvosanat konvergoivat vahvasti ($r_{XY} = 0,60$), samoin koulujen sisällä keskimäärin ($r_{XY} = 0,63$), mutta koulujen välillä yhteys heikkeni, kun konvergenssin mittana käytetään korrelaatiokerrointa ($r_{XY} = 0,38$).

Kuvioon 17 on laskettu sama koulujen välinen ero seuranta-arvioinnissa osoitetun osaamisen ja arvosanojen suhteen myös historiassa. Kuvion 17 perusteella näyttää siltä, että yhteys koulujen historian seuranta-arvioinnissa osoitetun osaamistason ja annettujen historian keskimääräisten arvosanojen välillä oli heikko. Myös historiassa yhteistä vaihtelua oli vähän (14,7 %). Esimerkiksi koulut, joiden keskimääräiset historian arvosanat olivat 8, saivat ratkaisuosuuksiksi historian seuranta-arvioinnissa 35 %:sta noin 60 %:iin. Arvosanan 7 saaneiden koulujen keskimääräiset ratkaisuosuudet vaihtelivat puolestaan 37 %:sta 54 %:iin.



KUVIO 17. Kouluissa annettujen historian arvosanojen keskiarvon suhde koulujen keskimääräisiin historian ratkaisuosuuksiin.

Vaikka historiassakin yhteistä vaihtelua oli vähän, oli se kuitenkin hieman yli 10 prosenttiyksikköä enemmän kuin yhteiskuntaopissa.

Jo yhteiskunnallisten aineiden seuranta-arviointiraportin perusteella tiedetään, että tällaisia koulujen välisiä eroja on sekä yhteiskuntaopissa että historiassa, ja ne vaikuttavat olevan yhteydessä koulun yleiseen seuranta-arvioinnissa osoitettuun osaamistasoon (Ouakrim-Soivio & Kuusela 2012, 110–11). Sama koulujen välisten erojen yhteys koulujen yleiseen osaamistasoon eli seuranta-arvioinnissa saavutettuun keskimääräiseen ratkaisuosuuteen on myös osoitettu tämän tutkimuksen aineistolla luvussa 6.2.2 kuvioissa 14 ja 15. Esitettyjen tulosten perusteella näyttää siltä, että ainakin toisen osaamisen mittarin, arvosanan tai seuranta-arvioinnin validius voidaan asettaa kyseenalaiseksi. Itsestään selvänä lähtökohtana on yleensä pidetty sitä, että juuri opettajien antamat arvosanat eivät ole valideja (Cizek 1997, 29; Loyd & Loyd 1997, 481).

Saatu tutkimustulos, jonka mukaan kummatkin osaamisen mittarit, seuranta-arviointi ja arvosanat, korreloivat vahvasti kummassakin oppiaineessa, yhteiskuntaopissa ja historiassa oppilastasolla ja koulujen sisällä, mutta eivät enää koulujen välisessä vertailussa, antaa aiheita pohtia, onko kyseessä yksi ongelma, jossa on kaksi

ilmenemismuotoa vai kaksi kokonaan eri ongelmaa. Onko niin, että koulujen välillä keskimääräisten arvosanojen ja ratkaisuosuuksien korrelaatio pienenee, koska arvosanat taipuvat kohti keskimääräistä vai onko kysymys myös siitä, että Perusopetuksen opetussuunnitelman perusteiden (2004) arviointia koskevat normit antavat sekä opettajille että seuranta-arvioinnin tehtävät laatineelle ryhmälle mahdollisuuden tulkita perusteita varsin väljästi.

On mahdollista, että yhteiskuntaopin ja historian opettajat ovat tulkinneet eri tavoin eri kouluissa Perusopetuksen opetussuunnitelman perusteissa (2004) määrittäjä normeja oppilaan opintojen aikaisesta arvioinnista ja päättöarvioinnista, kun he ovat arvioineet oppilaitaan ja antaneet heille arvosanoja. On myös mahdollista, että eri kouluissa yhteiskuntaopin ja historian opettajat ovat tulkinneet eri tavalla opetussuunnitelman perusteissa määriteltyjä opetuksen tavoitteita ja sisältöjä, sillä suomalaisessa koulutusjärjestelmässä opettajan pedagoginen vapaus antaa siihen mahdollisuuden, kuten tämän tutkimuksen luvuissa 2 ja 3 on jo aiemmin esitetty. Tällöin kyse olisi siitä, että eri kouluissa yhteiskuntaopin ja historian opettajat painottaisivat opetuksessa ja oppilaan arvioinnissa eri asioita. Lisäksi on mahdollista, että yhteiskuntaopin ja historian seuranta-arvioinnin tehtäväsarjat laatinut asiantuntijaryhmä on tulkinnut eri tavalla Perusopetuksen opetussuunnitelman perusteissa (2004) yhteiskuntaopin ja historian oppiaineiden tavoitteita, sisältöjä ja arviointikriteereitä kuin yhteiskuntaopin ja historian opettajat. Tällöin kyse olisi siitä, että yhteiskuntaopin ja historian seuranta-arviointi olisi kohdannut eri tavalla eri koulujen opetuksen ja sen arvioinnin painotuksia.

Mikäli koulujen väliset erot keskimääräisissä ratkaisuosuuksissa annettujen keskimääräisten arvosanojen suhteen johtuisivat siitä, että kouluissa olisi tulkittu eri tavalla opetuksen tavoitteita, sisältöjä ja arviointiohjeita kuin seuranta-arvioinnin tehtäväryhmä on tulkinnut, päädyttäisiin siihen, että koulujen väliset erot seuranta-arvioinnissa osoitetun osaamisen ja annettujen arvosanojen suhteen johtuisivat Perusopetuksen opetussuunnitelman perusteista (2004) tehdyistä erilaisista tulkinnoista, joita tehdään toisaalta arvosanoja annettaessa ja toisaalta seuranta-arvioinnin tehtäväsarjoja laadittaessa. Opetuksen erilaiset painotukset eri kouluissa eivät kuitenkaan kykene selittämään sitä, miksi seuranta-arviointi näyttää olevan validi mittari oppilastasolla ja koulujen sisällä, mutta ei enää koulujen välisessä vertailussa, varsinkaan kun yhteiskunnallisten aineiden seuranta-arvioinnissa käytetyt tehtäväsarjat eivät muuttuneet oppilastasolta koulujen tasolle siirryttäessä.

6.3 Koulujen väliset erot keskimääräisissä ratkaisuosuuksissa ja arvosanojen keskiarvoissa

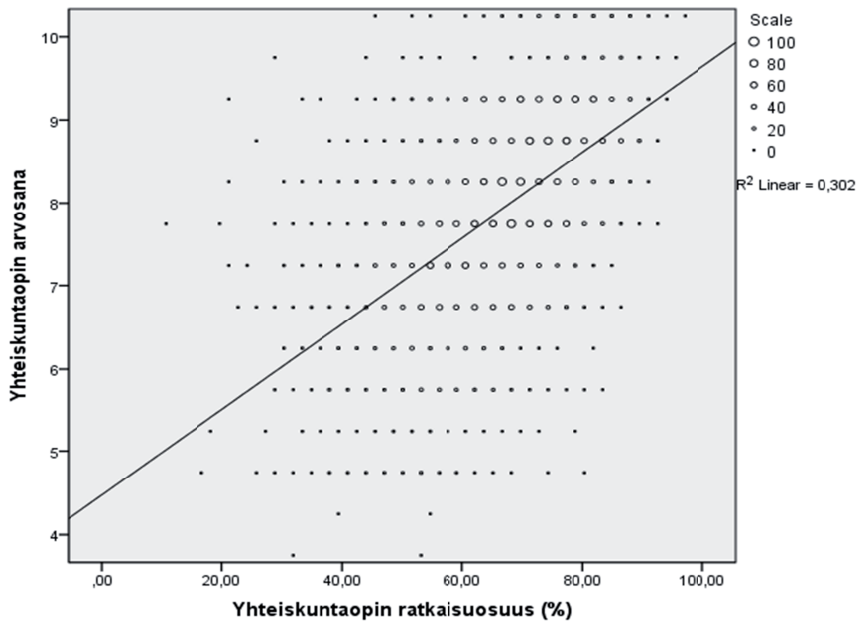
Jakku-Sihvonen ja Kuusela esittivät jo 2000-luvun alussa tehdyn koulutuksen tasa-arvoon keskittyneen meta-arvioinnin yhteydessä, että ”kouluissa tulisi kiinnittää huomiota oppilasarvioinnin yhdenmukaisuuteen ja oikeudenmukaisuuteen” (Jakku-Sihvonen 2002a, 13). Kymmenen vuotta myöhemmin Jakku-Sihvonen ja Kuusela toteavat selvityksessään (2012), että ”oppimistulosten arviointien jatkoanalyysit ovat osoittaneet, että arvosanatiedoista laskettuna koulujen erot ovat olennaisesti pienempiä kuin arviointikokeiden perusteella laskettuna. Analyysit ovat osoittaneet, että keskimääräistä paremmin suoriutuneissa kouluissa oppilaan arviointi on ankarampaa ja keskimääräistä heikommin suoriutuneissa kouluissa vaatimustasoa lasketaan arvosanaa annettaessa”. (Jakku-Sihvonen & Kuusela 2012b, 76.) Yhteiskuntaopin ja historian aineisto tukee edellä esitettyä Jakku-Sihvosen ja Kuuselan tulosta, kuten tämän tutkimuksen luvussa 6.2 osoitettiin. Jotta koulujen välisiä eroja annettujen arvosanojen ja seuranta-arvioinnissa osoitetun osaamisen suhteen voitaisiin tutkia tarkemmin, kouluja tulee voida verrata keskenään.

6.3.1 Koulujen välisten yhteiskuntaopin ratkaisuosuuksien ja arvosanaerojen laskeminen

Perinteinen menetelmä edellisen luvun lopussa esitetyn koulujen välisten ratkaisuosuuksien ja arvosanaerojen laskemiseksi olisi lineaarinen regressioanalyysi ja sen perusteella laskettavat residuaalit.⁷¹ Kuvioista 18 käy ilmi, että hajontaa on jokaisen annetun yhteiskuntaopin arvosanan kohdalla paljon. Toinen kuvioista 18 ilmenevä seikka on, että regressioanalyysi ei sovi lähtökohdaksi, sillä arvosanojen ja arvioinnissa osoitetun osaamisen eli ratkaisuosuuksien yhteys ei ole lineaarinen.

Kuvion 18 mukaan vaikuttaisi siltä, että heikoimmat arvosanat saaneet oppilaat olisivat menestyneet yhteiskuntaopin seuranta-arvioinnissa arvosanaansa nähden paremmin ja hyvän arvosanan saaneet oppilaat taas odotettua heikommin. Kuvion osoittama lopputulos johtaisi puolestaan myöhemmissä analyyseissäkin systemaattisiin keskimääräisistä arvosanoista johtuviin vääristymiin ja virhetulkintoihin, jos residuaaleja käytettäisiin analyyysien perusteena. Tällöin vaikuttaisi siltä, että koulut, joissa on annettu usein heikkoja arvosanoja, menestyisivät seuranta-arvioinnissa arvosanoihin nähden oletettua paremmin, ja koulut, joissa on annettu usein hyviä arvosanoja, menestyisivät seuranta-arvioinnissa arvosanoihin nähden oletettua heikommin.

71 Residuaalilla tarkoitetaan havaitun todellisen arvon ja ennustetun arvon erotusta. Määritelmä löytyy Helsingin yliopiston sivustolta: <http://www.helsinki.fi/atk/tilasto/tao/regresiduals.html>



KUVIO 18. Regressioanalyysi, jossa x-akselilla yhteiskuntaopin ratkaisuosuudet ja y-akselilla yhteiskuntaopin arvosanat.

Mahdollisten systemaattisten erojen selvittämiseksi voidaan oppilaan saamasta yhteiskuntaopin arvosanasta ja hänen yhteiskuntaopin seuranta-arvioinnissa osoittamastaan osaamisesta muodostaa yhdistetty muuttuja.

6.3.2 Odotusarvojen laskeminen

Koska regressioanalyysi ei sovellu tässä yhteydessä käytettäväksi, hyödynnetään menetelmää, jota on käytetty jo aikaisemminkin epälineaarisuudesta johtuvan ongelman ratkaisemiseksi (Kuusela 2006). Tässä tutkimuksessa menetelmää on kehitetty vielä siten, että kullekin oppilaan saamalle yhteiskuntaopin ja historian arvosanasta lasketulle odotusarvolle on laskettu vielä raja-arvot eli vaihteluväli.

Ensimmäisessä vaiheessa sekä yhteiskuntaopin että historian arvosanoille lasketaan ratkaisuprosenttien odotusarvo, joka on kutakin arvosanaa vastaava keskimääräinen ratkaisuprosentti.

Näin laskettuna päädytään odotusarvoihin, jotka on kuvattu taulukossa 14. Jokaista yhteiskuntaopin arvosanaa kohden on laskettu odotusarvo seuranta-arvioinnissa saatujen ratkaisuosuuksien perusteella. Kuten taulukosta 14 ilmenee, esimerkiksi yhteiskuntaopissa arvosanan 7 saaneiden oppilaiden keskimääräinen ratkaisuosuus

ja samalla odotusarvo on 59 %, arvosanan 8 saaneiden oppilaiden odotusarvo on 66 % jne. Odotusarvo on siis se ratkaisuosuus, jonka kunkin arvosanan saanut oppilas olisi saanut, mikäli arvosana ja seuranta-arviointi toimisivat täysin rinnakkaisina mittareina.

TAULUKKO 14. Esimerkki odotusarvojen laskemisesta oppilaiden saamien yhteiskuntaopin arvosanojen ja seuranta-arvioinnissa osoitetun osaamisen eli ratkaisuosuuksien perusteella.

Yhteiskuntaopin arvosana	Keskimääräinen ratkaisuosuus (%)	Hajonta	N
4	46,4	10,2	5
5	50,2	11,1	140
6	54,6	10,4	393
7	59,4	11,4	979
8	66,2	10,8	1 256
9	72,3	10,2	1 104
10	77,5	10,1	237
Yhteensä	65,1	10,6	4 114

Toisessa vaiheessa jokaisen oppilaan todellista seuranta-arvioinnissa saamaa ratkaisuprosenttia verrataan hänen yhteiskuntaopin arvosanansa perusteella saatuun odotusarvoonsa laskemalla näiden erotus. Jos oppilaan todellinen ratkaisuosuus ylittää odotusarvon, hän on osannut yhteiskuntaopin seuranta-arvioinnissa odotettua paremmin. Jos taas erotus on ollut negatiivinen, hänen arvosanaan suhteutettu menestyksensä seuranta-arvioinnissa on ollut odotettua heikompi. Kääntäen tämä voidaan tulkita myös niin, että olennainen positiivinen ero merkitsee, että oppilas on saanut odotettua heikomman arvosanan, ja negatiivinen ero, että syystä tai toisesta oppilaan arvosana on odotettua parempi. Keskiarvon määritelmän perusteella näin muodostetun yhdistetyn muuttujan keskiarvo on tasan nolla jokaiselle arvosanalle ja siis myös koko aineistolle.

Kuten taulukosta 14 käy ilmi, olennaisena erona voidaan pitää noin viiden prosenttiyksikön eroa, sillä kahden arvosanan välinen keskimääräisten ratkaisuprosenttien ero vaihtelee 3,8 ja 6,7 prosenttiyksikön välillä. Samasta taulukosta havaitaan, että kunkin arvosanan keskiarvon taustalla on noin 10 prosenttiyksikön suuruinen standardipoikkeama. Uuden muuttujan avulla on mahdollista tutkia, kuvaako standardipoikkeama pelkkää satunnaisvaihtelua, vai löydetäänkö poikkeamiin sellaisia tilastollisia yhteyksiä, jotka voidaan tulkita arvosanoihin systemaattisesti yhteydessä oleviksi tekijöiksi.

On mahdollista, että jokin oppilasryhmä liikkuu useimmiten arvosanojen alarajoilla ja jokin toinen oppilasryhmä ylärajoilla ilman, että olisi perusteltua puhua arvosanojen systemaattisista vääristymisistä. Kuitenkin jo ratkaisuosuuksien standardipoikkeamat, jotka olivat yli 10 prosenttiyksikköä, osoittavat, ettei kysymys voi olla siitä, että käsiteltäisiin vain kunkin arvosanan sisäistä ja siksi hyväksyttävää vaihtelua. Myös seuranta-arviointiin liittyy satunnaisvaihtelua. Jos oppilas on epäonnistunut seuranta-arvioinnissa, hän on saanut arvosanaansa nähden heikommän ratkaisuosuuden ja jos hän taas on onnistunut, on hän saanut paremman ratkaisuosuuden kuin hänen arvosanansa on antanut odottaa.

Jos oletetaan, että jokaisella arvosanalla on odotusarvonsa, siihen liittyy myös jonkinlainen raja-arvo, koska arvosanaluokitus on paljon karkeampi osaamisen mitta kuin ratkaisuosuus. Seuraavassa on laskettu esimerkkinä yhteiskuntaopin arvosanalle 8 lasketut raja-arvot. Yhteiskuntaopin arvosanojen 8 ja 7 ero on noin 6,6 prosenttiyksikköä. Puolet siitä on noin 3,3. Kun arvosanan 8 keskimääräinen ratkaisuosuus on noin 66,2, kuten taulukossa 14 on jo aiemmin esitetty, päästään siihen, että yhteiskuntaopin arvosanan 8 alaraja on 66,2–3,3 eli 62,9. Yhteiskuntaopin arvosanan 8 yläraja saadaan samalla tavalla arvosanan 8 ja 9 erotuksesta, joka on noin 6,1 ja puolet siitä on noin 3,0. Kun pyöristysvirheet otetaan huomioon, yläraja on 66,2 + 3,0 eli 69,2.

Edellisen esimerkin mukaan yhteiskuntaopin arvosanaa 8 vastaava odotusarvo sisältää kaikki ratkaisuosuudet väliltä noin 62,9–69,2. Arvosanan 4 ratkaisuosuuden alaraja on alin ratkaisuosuus, jonka arvosanan 4 saanut oppilas sai seuranta-arvioinnissa eli 31,8 %. Arvosanan 10 ratkaisuosuuden yläraja on puolestaan arvosanan 10 saaneiden oppilaiden ylin ratkaisuosuus eli 97 %. Taulukossa 15 on esitetty lopputulos, jossa yhteiskuntaopin arvosanoille 4–10 on laskettu seuraavat raja-arvot edellisen esimerkin mukaisesti.

TAULUKKO 15. Yhteiskuntaopin raja-arvot arvosanoille 4–10. Taulukkoon on merkitty ratkaisuosuuden ala- ja yläraja kullekin kouluarvosanalle.

Ratkaisuosuuden alaraja (%)	Yhteiskuntaopin arvosana	Ratkaisuosuuden yläraja (%)
31,8	4	48,3
48,4	5	52,4
52,5	6	57,0
57,1	7	62,8
62,9	8	69,2
69,3	9	74,9
75,0	10	97,0

Kaikki laskennallisten rajojen sisällä olevat arvot on koodattu nolliksi. Kullekin arvosanalle on laskettu todennäköinen ala- ja yläraja kuten edellä esitetystä taulukossa 15 on kuvattu. Tämän jälkeen tutkimusaineisto on koodattu uudelleen siten, että jokaista yhteiskuntaopin ja historian arvosanaa vastaavat ala- ja ylärajat siirrettiin SPSS-tilukoon. Näin saadaan kloonatuksi kutakin arvosanaa vastaavat raja-arvot jokaiselle oppilaalle. Jokaiselle oppilaalle laskettiin alarajan alitus vähentämällä alaraja hänen saamastaan yhteiskuntaopin (tai historian) ratkaisuosuudesta. Jos oppilas saa negatiivisen luvun, jäädään alarajan alle. Loput koodataan nolliksi. Samalla tavalla jokaiselle oppilaalle on laskettu myös ylärajan ylitys vähentämällä yläraja ratkaisuosuudesta. Jos oppilas saa positiivisen luvun, yläraja ylittyy. Loput koodataan nolliksi. Tämän jälkeen lasketaan näiden kahden muuttujan summa, jolloin on saatu uusi oppilastason muuttuja, joka on tässä tutkimuksessa nimetty *eroksi odotusarvosta*. Muuttujan nimellä, *ero odotusarvosta*, viitataan siihen, ylittääkö tai alittaaako oppilas hänen arvosanaansa nähden lasketun odotusarvon. Muuttujalla kuvataan koulujen keskimääräistä poikkeamaa arvosanojen perusteella lasketusta ennusteesta.

Oppilastason muuttujan avulla on laskettu kouluittain keskiarvot sekä yhteiskuntaopissa että historiassa. Ero on laskettu suunnassa todellinen ratkaisuosuus-odotusarvo. Näin ollen, kun ero on positiivinen, on osattu hyvin arvosanoihin nähden ja kun erotus on negatiivinen, on osattu heikosti arvosanoihin nähden. Uusi muuttuja, koulun keskimääräinen poikkeama odotusarvosta, kuvaa todellista tilannetta. Korjaamaton odotusarvo, kuten taulukossa 14 on esitetty, antaa keskiarvoksi 0 ja keskihajonnaksi 10,6. Uuden muuttujan, yhteiskuntaopin ero odotusarvosta tai historian ero odotusarvosta, standardipoikkeama pienenee 10,6:sta 8,4:ään. Se on odotettua, koska nyt arvosanojen sisäinen, sallittu vaihtelu on suodatettu pois. Jos taas olennaisia eroja on, analyysin seuraava vaihe on pyrkiä löytämään selityksiä koulujen väliselle vaihtelulle.

Kaikki validiuden määritelmät liittyvät mittaamiseen. Konstruktio- eli käsitevalidiuden näkökulma liittyy vielä erityisesti siihen, vastaako mittaus sitä teoreettista käsitettä, jota sen on tarkoituskin mitata (Messick 1980, 1015; 1989b, 13; 1995, 741). Tässä tapauksessa ydinkäsite on yhteiskuntaopin osaaminen, joka on jo aikaisemmin määritelty luvussa 3.3.⁷² Haasteena on, ettei yhteiskuntaopin tai historian osaamiselle ole olemassa mitään absoluuttista kolmatta mittaa, johon voisi tehdä vertailuja. Käytettävissä on ainoastaan kaksi Perusopetuksen opetussuunnitelman perusteissa (2004) määritellyyn osaamiseen kohdistuvaa mittaria, arvosanat ja seuranta-arviointi. Näiden kahden yhteiskuntaopin osaamista mittaavan mittarin validiuden kannalta on olennaista tarkastella sitä, voidaanko yhteiskuntaopin osaamista yleensäkin arvioida yhdellä seuranta-arvioinnilla tai yhdellä arviointikerralla, ja jos voidaan, onko nyt käytetty seuranta-arviointi sellainen, että se mittaa kohtalaisen

72 Tässä tutkimuksessa yhteiskuntaopin ja historian osaaminen ja sen määrittely perustuvat Perusopetuksen opetussuunnitelman perusteisiin (2004), sillä muuta kansallisesti yhdenmukaiset sisällöt, tavoitteet ja kriteerit määrittävää toimijaa, asiakirjaa tai kriteereitä ei tällä hetkellä suomalaisessa koulutusjärjestelmässä ole käytettävissä.

harhattomasti yhteiskuntaopin osaamista. Kuten tutkimuksen johdantoluvussa jo esitettiin, kyse on ennen kaikkea siitä, mittaavatko seuranta-arviointi ja arvosanat oppilaiden osaamista Perusopetuksen opetussuunnitelman perusteissa (2004) määriteltyjen päättöarvioinnin kriteereiden mukaisesti, yhdenvertaisin periaattein.

Korkeakosken mukaan on ”harhaista väittää, että kansallisissa seuranta-arvioinneissa mitataan sitä, mitä kouluissa opetetaan” (Korkeakoski 2005, 188). Hänen mukaansa myöskään ”yksi seuranta-arviointikerta ei riitä pitkälle menevien päätelmien tekemiseen” (Korkeakoski 2005, 182). Kuitenkin perusopetuksen tulee suomalaisessa koulutusjärjestelmässä perustua opetussuunnitelman perusteissa opetukselle määriteltyihin tavoitteisiin ja sisältöihin. Tilanne, jossa kouluissa pääsääntöisesti opetettaisiin eri asioita kuin mitä Perusopetuksen opetussuunnitelman perusteissa (2004) on määritelty, johtaisi ennen pitkää siihen, että normit eivät ohjaisi opetusta sen enempää kuin oppilaan arviointiakaan.

Samalla tavalla kuin muitakin oppiaineita, myös yhteiskuntaopin osaamista ja yhteiskuntaopin opetukselle asetettuja tavoitteita on pakko voida arvioida. Täysin mahdollista on kuitenkin se, että esimerkiksi yhteiskuntaopin tavoitteiden saavuttamista on vaikea arvioida yhdellä ainoalla seuranta-arvioinnilla. Kyse ei kuitenkaan ole siitä, mittaako seuranta-arviointi kaikkea sitä yhteiskunnallista osaamista, jota oppilaalla mahdollisesti on perusopetuksen päättyessä ja jota ei ehkä millään mittarilla kokonaisuudessaan tavoiteta. Kyse on ennen kaikkea siitä, indikoiko seuranta-arviointi riittävän kattavasti ja monipuolisesti sitä yhteiskuntaopin osaamista, jota oppilaalla Perusopetuksen opetussuunnitelman perusteissa (2004) oletetaan olevan, kun hän siirtyy peruskoulusta jatko-opintoihin ja työelämään. Seuranta-arvioinnin validiuteen liittyvät tällöin kysymykset siitä, mittaako seuranta-arviointi oikeita asioita ja mittaako se oppilaiden yhteiskuntaopin osaamista riittävän laajasti. Pohdinta annettujen arvosanojen ja seuranta-arvioinnin validiudesta johtaa jo tutkimuksen alussa esitettyihin kysymyksiin siitä, voidaanko seuranta-arvioinnin perusteella tehdä johtopäätöksiä arvosanojen validiudesta, vai jääkö kahden mittauksen validiusongelma ratkaisematta sekä kysymykseen, voidaanko kahden oppiaineen samanaikaisella arvioinnilla ratkaista kahden mittauksen eli arvosanojen ja seuranta-arvioinnin validiutta koskeva ongelma.

Tämän tutkimuksen lähtökohdaksi on valittu kaksi rinnakkaista osaamisen mittaria, yhteiskuntaopin seuranta-arviointi ja yhteiskuntaopin arvosana. Kun lisäksi käytössä on samojen koulujen ja samojen oppilaiden historian seuranta-arvioinnin tulokset ja päättöarvosanat, voidaan historiasta kerättyä tutkimusaineistoa käyttää yhteiskuntaopissa saatujen tulosten vertailemiseksi. Kaksi mittausta antaa mahdollisuuden tarkastella niiden validiutta ja konvergenssia. Mahdollisia väitteitä on kaksi: joko oppilaan yhteiskuntaopin arvosana on validimpi mittari tai seuranta-arviointi mittaa yhteiskuntaopin osaamista validimmin. Kolmas, periaatteellinen väite voisi olla se, että kumpikaan tapa mitata yhteiskuntaopin osaamista ei ole validi. Tuolloin arvosanat ja seuranta-arviointi olisivat eri tavalla harhaisia. Tähän kolmanteen vaihtoehtoon ei kuitenkaan päästä empiirisesti käsiksi.

Arvosanojen validiustarkastelu etenee siten, että ensin laaditaan Toulminin argumentaatiomallin mukaisesti tutkimusaineiston lähtötietojen perusteella kaksi keskenään kilpailevaa väitettä, joita koetellaan validiusargumentoinnin avulla. Toulminin argumentaatiomallissa päättelyketju alkaa tosiasista tai lähtötiedosta, jonka perusteella muodostetaan väite tai loppupäätelmä (Toulmin 2003, 90–91). Vaikka Toulminin mukaan argumentaatio on alakohtaista, eri alojen argumentaatiota yhdistää se, että kaikkia johtopäätöksiä voidaan analysoida edellä esitetyn kaavan mukaisesti (Toulmin 2003, 92; Kane 2006, 28). Vaikka ketju, tosiasia – esitetty perustelu tai todiste – väite tai johtopäätös, on hyvin yleinen, perustelut tai todisteet voivat olla hyvinkin erityisiä tai täsmällisiä. Tämän vuoksi testituloksia voidaan tulkita erilaisista todisteista tai perusteluista käsin. Todisteet väitteille eivät ole itsestäänselvyyksiä ja siksi ne tulee aina perustella. Toulminin mukaan hänen argumentointimallissaan ei ole kyse siitä, että väitteen tueksi esitetään lisää tosiasiallista, faktuaalista, tietoa, vaan kyse on siitä, että osoitetaan, mihin johtopäätös tai väite perustuu. Toulminin mukaan esitettäessä tosiasioita tai lähtötietoja väitteen tai johtopäätöksen tueksi johtaa se argumentoinnin vaiheittaiseen etenemiseen, joka ei perustu tiedon lisäämiseen, vaan siihen, että jokainen esitetty väite tai johtopäätös saa vahvistusta tai tukea olemassa olevista säännöistä, laeista tai yleispätevästä päättelystä. Perusteet tekevät lähtötiedoista merkityksellisiä suhteessa johtopäätöksiin, ja ne toimivat ikään kuin siirtymävaiheena, kun edetään lähtötiedosta väitteeseen tai johtopäätökseen. (Toulmin 2003, 91.)

Toulminin päättelymallia hyödynnetään siten, että kerätyn aineiston perusteella on muodostettu lähtötieto siitä, että koulujen välillä on eroja annettujen keskimääräisten arvosanojen ja keskimääräisen seuranta-arvioinnissa osoitetun osaamisen suhteen. Edellisissä luvuissa 6.1 ja 6.2 on jo perusteltu ja annettu todisteita siitä, että toinen osaamisen mittari, seuranta-arviointi tai arvosana, on epävalidi. Lisäksi luvuissa 6.1–6.3 on jo annettu tarkennuksia ja varauksia, jonka pohjalta laaditaan kaksi kilpailevaa väitettä, joita koetellaan ja haastetaan validiusargumentoinnin avulla. Tässä tutkimuksessa luvuissa 6.4.1 ja 6.4.2 koeteltavat väitteet ovat: arvosanat ovat validi yhteiskuntaopin ja historian osaamisen mittari koulujen välisiä osaamiseroja vertailtaessa tai arvosanat eivät ole validi osaamisen mittari. Kuten tieteellisiä teorioita yleensäkin, tulkinta-argumentteja haastetaan tai kyseenalaistetaan siten, että tulkinta-argumentin perusteella laaditaan oletuksia, joita myös haastetaan. Validiusargumentointi onnistuu, mikäli esitetty tulkinta antaa riittävän taustan ja argumentaatiolla onnistutaan sulkemaan pois muut tulkinnat. (Kane 2006, 29.) Mikäli tulkinta-argumentit sisältävät tilastollisia yleistyksiä, otoksen perusteella tehdyistä havainnoista ja niiden syistä tulisi voida tehdä koko perusjoukkoa koskevia yleistyksiä. Jos tulkinnasta ei voi tehdä yleistystä, tulkinnan validiutta tulee epäillä. (Kane 2006, 23.) Validiusargumentoinnin tarkoituksena on tarjota esitettyjen tulkintojen avulla kokonaisarviointi havaintojen puolesta ja niitä vastaan (Kane 2006, 28). Seuraavassa luvussa 6.4 väitteitä arvosanat ovat validi mittari tai arvosanat eivät ole validi mittari, tuetaan ja testataan empiirisen aineiston perusteella ”jos–niin”-säännön avulla (ks. Kane 2006, 23). Esitettyjä väitteitä arvosanojen validiudesta

tai epävalidiudesta osaamisen mittarina haastetaan vasta-argumenteilla, joilla pyritään horjuttamaan esitettyjä väitettä tarjoamalla niille vaihtoehtoisia selityksiä tai osoittamalla niissä loogisia tai empiirisiä puutteita (Bachman 2005, 10; ks. Hildén 2008, 8–10).

6.4 Arvosanat ovat validi osaamisen mittari?

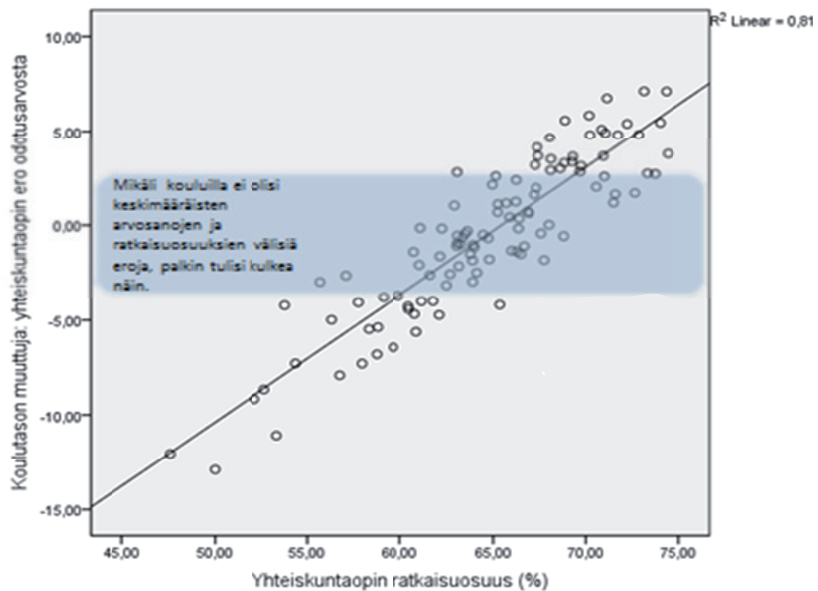
Tässä luvussa keskimääräisissä ratkaisuosuuksissa ja keskimääräisissä arvosanoissa ilmenneiden koulujen välisten erojen taustalla on oletus siitä, että arvosanat ovat validi mittari kuvaamaan oppilaiden osaamista yhteiskuntaopissa ja historiassa. Oletukseen sisältyy ajatus, että yhdellä seuranta-arvioinnilla ei voi mitata yhteiskuntaopin osaamista yhtä laajasti ja perustellusti, koska seuranta-arvioinnilla saadaan kuitenkin olennaisesti vähemmän tietoa kuin opettajalla on arvosanaa antaessa käytössään. Tämä on kaikkien seuranta-arviointien dilemma ja jopa sellainen, joka yleensä nostetaan esiin Opetushallituksen seuranta-arviointiraporteissa (ks. esim. Lappalainen 2006, 6, 75; Lappalainen 2011, 92).

On mahdollista, että seuranta-arvioinnissa syntyy oppilastason satunnaista vaihtelua. Esimerkkinä oppilastason satunnaisvaihtelusta voisi olla hieman harvinaisempi tilanne, jossa jotkut oppilaat onnistuvat seuranta-arvioinnissa poikkeuksellisen hyvin, tai todellisuudessa paljon yleisempi tilanne, jossa oppilaat menestyvät seuranta-arvioinnissa odotettua heikommin. Tämä on kuitenkin reliabeliuteen eli mittausprosessin luotettavuuteen eikä validiuteen liittyvä kysymys. Jos kysymys on vain satunnaisesta vaihtelusta, koulujen välille ei tulisi syntyä systemaattisia eroja.

Kuviossa 19 on ensinnäkin kuvattu empiirisellä aineistolla todellinen tilanne, jossa selittävänä muuttujana x-akselilla on koulun yhteiskuntaopin seuranta-arvioinnissa osoitettu keskimääräinen osaamistaso eli yhteiskuntaopin ratkaisuosuus ja y-akselilla on uusi, laskettu koulutason muuttuja eli ero yhteiskuntaopin odotusarvosta. Osittain yhteys on laskennallista, koska x-akselille kuvattu seuranta-arvioinnissa osoitettu osaaminen sisältyy osittain myös y-akselilla olevaan eroon odotusarvosta. Lineaarinen yhteys ratkaisuosuuden ja seuranta-arviointituloksen ero odotusarvoista välillä ei ole empiirinen löydös, vaan argumentoinnin tuki. Taustalla olevasta todellisesta kuvioista ilmenee, että yhteistä vaihtelua on 81 %. Käytännössä se tarkoittaa sitä, että mitä paremmin koulu on menestynyt seuranta-arvioinnissa, sitä paremmin on menestytty koulun antamiin keskimääräisiin arvosanoihin nähden. Kuvioista näkyy myös, että koulujen väliset erot ovat suuria. Tutkimusaineiston perusteella tiedetään, että äärimmillään koulujen väliset erot ovat 20 prosenttiyksikköä, joka vastaa noin neljää kouluarvosanaa, kuten aiemmin esitetystä taulukosta 14 ilmenee.⁷³

73 Taulukossa 14 esitetyn perusteella tiedetään, että olennaisena erona voidaan pitää noin viiden prosenttiyksikön eroa. Kahden arvosanan välinen keskimääräinen ratkaisuprosenttien ero vaihtelee 3,8 ja 6,7 prosenttiyksikön välillä. Näin ollen 20 prosenttiyksikön ero vastaisi neljää kouluarvosanaa.

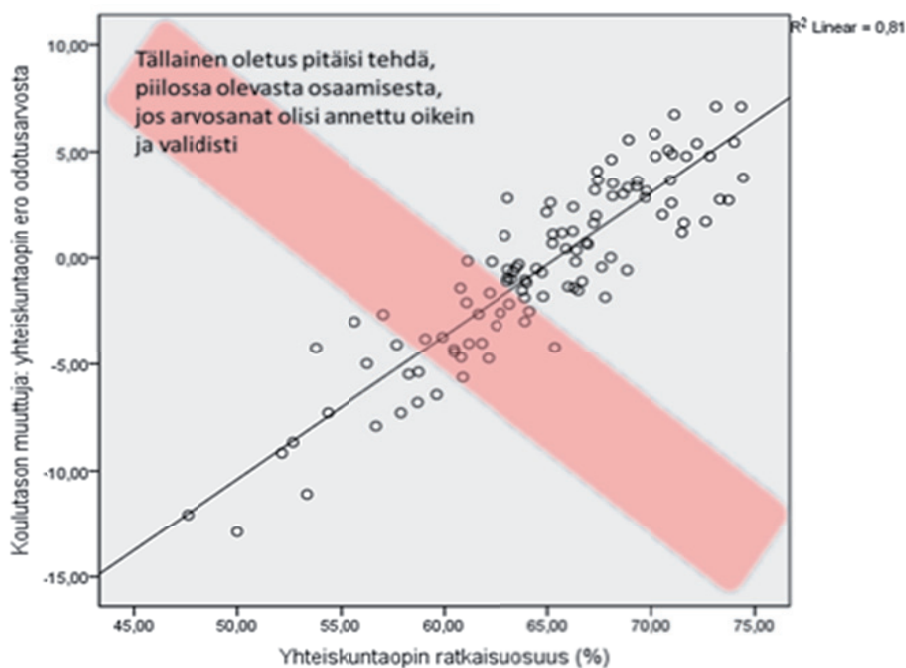
Todellisen tilanteen lisäksi kuvioon 19 on piirretty vaakatasoon kulkeva palkki, jolla kuvataan ideaalimallia, mikäli koulujen välillä ei olisi keskimääräisten arvosanojen ja ratkaisuosuuksien kesken eroja. Aiheellista on kuitenkin kysyä, miksi ei saada nollakorrelaatiota eli miksei regressiosuora kulje vaakasuoraan, kuten kuvioon 19 piirretty palkki osoittaa. Mikäli regressiosuora kulkisi kuvioon piirretyn palkin mukaisesti vaakatasossa, koulujen välillä ei olisi ratkaisuosuuksien ja arvosanojen välisiä eroja, vaan seuranta-arvioinnissa osoitettu osaaminen vastaisi kutakuinkin kouluissa annettuja arvosanoja. Tällöin validiusongelmia ei olisi, koska mittarit olisivat yhdenmukaisia. Kun on osattu seuranta-arvioinnissa paremmin, on myös saatu vastaavassa suhteessa parempia arvosanoja, tai kun on osattu seuranta-arvioinnissa huonommin, on myös saatu vastaavassa suhteessa heikompia arvosanoja. Näin vain satunnaisvaihtelu jäisi jäljelle. Nyt kuitenkin on selvästi olemassa hyvin voimakas koulujen välinen systemaattinen vaihtelu, kuten kuviossa 19 todellista tilannetta kuvaava regressiosuora osoittaa.



KUVIO 19. Koulujen keskimääräinen yhteiskuntaopin ratkaisuosuus suhteessa koulutason muuttuun sekä ideaalimallia kuvaava palkki.

6.4.1 Väitteen, arvosanat ovat validi mittari, koetteleminen

Kuviossa 20 on havainnollistettu todellista tilannetta yhteiskuntaopissa kuvaava regressiosuora, joka kulkee nousevasti vasemmalta oikealle. Samaan kuvioon piirretty vasemmalta oikealle laskevasti kulkeva pinkki kuvaa oletettua tilannetta, jossa oppilaiden osaamista jää seuranta-arvioinnin ulkopuolelle. Kuvion taustalla on oletus siitä, että oppilaan yhteiskuntaopin arvosana on validi, mutta oppilaiden osaamista jää seuranta-arvioinnin ulkopuolelle, ikään kuin piiloon. Tällä piiloon jäävällä yhteiskuntaopin osaamisella voidaan tarkoittaa joko oppilaiden muuta osaamista tai sellaista ”todellista osaamista”, joka on johdettu Perusopetuksen opetussuunnitelman perusteista (2004) ja joka on oleellista yhteiskuntaopin kriteereiden ja arvosanan antamisperusteiden kannalta.



KUVIO 20. Yhteiskuntaopin keskimääräinen ratkaisuosuus suhteessa koulutason muuttujaan sekä palkki, jolla kuvataan mahdollista seuranta-arvioinnin ulkopuolelle jäävää osaamista.

Mikäli kuvioon piirretty palkki kuvaisi todellista tilannetta ja oletettaisiin, että yhteiskuntaopin arvosanat ovat valideja, siitä seuraisi toisen osaamisen mittarin kannalta, että

a) seuranta-arviointi on mitannut vääriä asioita, kuten esimerkiksi oppilaiden luetun ymmärtämistä tai esimerkiksi detaljitiedon muistamista edellyttäviä asioita tai

b) seuranta-arviointi on ollut liian kapea-alainen

Koska seuranta-arvioinnissa saavutetut ratkaisusuudet ja oppilaiden arvosanat korreloivat vahvasti sekä oppilastasolla että koulujen sisällä, on epätodennäköistä, että oletus a) pitäisi paikkansa.

Voi olla, että yhteiskunnallisten aineiden seuranta-arvioinnin tehtäväsarjat ovat olleet oletuksen b) mukaan kapea-alaisia, eikä niillä ei ole kyetty mittaamaan riittävän monipuolisesti ja laaja-alaisesti erilaisia yhteiskuntaopin ja historian tietoja ja taitoja, joita Perusopetuksen opetussuunnitelman perusteiden (2004) mukaan kouluissa opetetaan ja oppilaat osaavat. Mikäli oletus yhteiskuntaopin tehtäväsarjojen kapea-alaisuudesta pitäisi paikkansa, se olisi voinut johtaa siihen, että opetussuunnitelman perusteissa määriteltyä osaamista olisi jäänyt piiloon kuviossa 20 esitetyllä tavalla. Jos yhteiskuntaopin osaamista olisi näin jäänyt piiloon, se tarkoittaisi joko sitä, että

1. Opetushallituksen laatimalla yhteiskunnallisten aineiden seuranta-arvioinneilla enempää kuin muidenkaan oppiaineiden oppimistulosten seuranta-arvioinneilla ei ole tavoitettu eikä tavoiteta todellista oppiaineen osaamista tai
2. yhteiskuntaoppi ja historia ovat sellaisia oppiaineita, joissa on poikkeuksellisen vaikea tavoittaa oppilaiden todellista osaamista tai
3. oppilaille olisi sitä enemmän muuta, piiloon jäänyttä osaamista, mitä heikommin he olisivat menestyneet seuranta-arvioinnissa.

Ensimmäistä ja toista oletusta ei tue se tieto, että yhteiskunnallisten aineiden, samoin kuin muidenkin seuranta-arviointien tehtäväryhmien jäsenet, pyritään kokoamaan niin, että he edustavat parasta mahdollista oppiaineen ja oppimistulosten mittaamisessa tarvittavaa osaamista, aivan kuten luvussa 5.1.2 on esitetty. Yhteiskuntaopin ja historian oppimistuloksia tulee myös voida mitata, aivan kuten muidenkin oppiaineiden oppimistuloksia. Esitetyistä oletuksista kolmas lienee todennäköisin ja sellainen, johon seuraavassa pureudutaan tarkemmin.

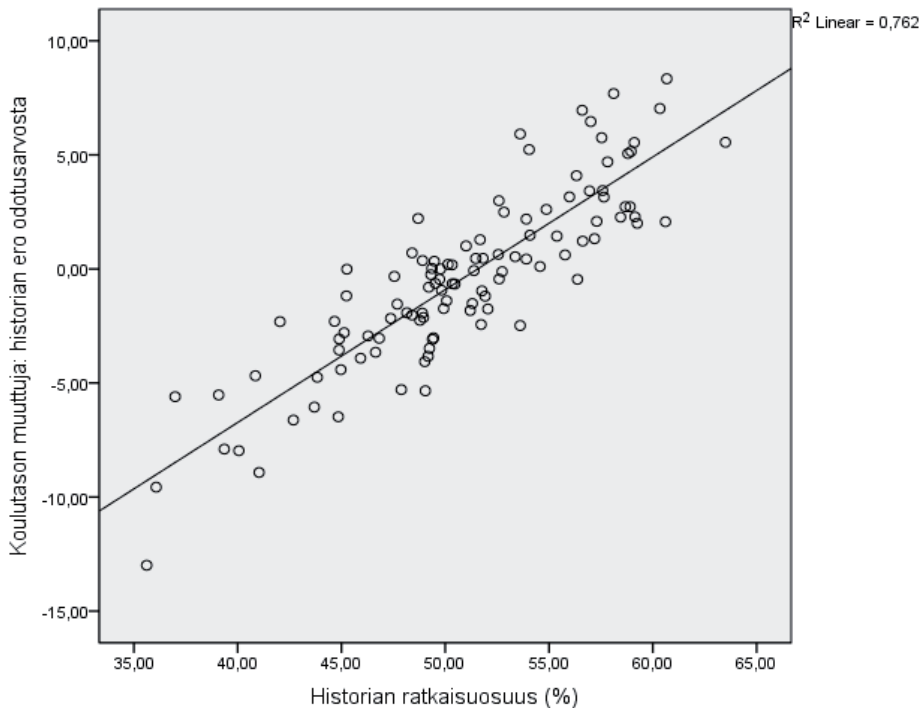
Jos oletuksen 3) mukaisesti oppilaiden muuta osaamista olisi jäänyt seuranta-arvioinnin ulkopuolelle, voisi kyse olla siitä, että seuranta-arviointi olisi osunut vaihtelevasti eri koulujen tai eri opettajien painotuksiin. Opettajan pedagoginen vapaus, samoin kuin se, että valintoja tehdään sekä opetuksessa että arvosanoja annettaessa, perustuu Perusopetuksen opetussuunnitelman perusteisiin (2004), jossa yhteiskuntaopin sisältöjen osalta on määritelty varsin väljästi ne opetukselle asetetut tavoitteet ja sisällöt, joita koulujen ja opettajien tulee opetuksessaan noudattaa (POPS 2004, 226–227). Korkeakoski on pukenut edellisen sanoiksi toteamalla, että ”Suomessa opettajan pedagoginen vapaus on laaja ja se johtaa oppimistulosten arvioinnissa siihen, että opettajien pedagogisesta ajattelusta muodostuu avainasema” (Korkeakoski 2005, 187), jolla hän viitanee siihen, että opettajan toiminta ja opettajien valitsemat opetuksen sisällöt sekä opetuksessa käytetyt menetelmät ohjaavat enemmän kuin valtakunnalliset opetussuunnitelman perusteet. On siis aivan perusteltua olettaa, että yhteiskuntaopin seuranta-arviointi on osunut paremmin kohdalleen joidenkin toisten koulujen pedagogisiin tai sisällöllisiin painotuksiin kuin joissain toisissa kouluissa asetettuihin painotuksiin. Yhteiskuntaopissa, samoin kuin kaikissa muissakin perusopetuksen oppiaineissa, opettajat tekevät opettavien aiheiden, tietojen ja taitojen välillä valintoja jatkuvasti. Mikään ei liioin tue oletusta, että opettajien pedagogiset valinnat olisivat vääriä tai vähemmän oikeita kuin seuranta-arvioinnin tehtäväryhmien tekemät valinnat. Kun tämä sallitaan, hyväksytään koulujen tasolla tapahtuva satunnaisvaihtelu. Tässä on kuitenkin kyse ennen kaikkea seuranta-arvioinnin reliiabeliudesta, ei validiudesta.

Mikäli piiloon jäävällä osaamisella tarkoitettaisiin puolestaan sellaista ”todellista osaamista”, joka on johdettu opetussuunnitelman perusteista, tarkoittaisi kuviossa 20 oleva palkki sitä, että seuranta-arvioinnin tehtäväsarjat yhteiskuntaopissa ja historiassa olisivat olleet kapea-alaisia, eikä niillä olisi kyetty mittaamaan riittävän monipuolisesti ja laaja-alaisesti erilaisia yhteiskuntaopin tietoja ja taitoja, joita Perusopetuksen opetussuunnitelman perusteiden (2004) mukaan kouluissa opetetaan ja oppilaat osaavat. Mikäli jälkimmäinen oletus yhteiskuntaopin tehtäväsarjojen kapea-alaisuudesta pitäisi paikkansa, se olisi johtanut siihen, että opetussuunnitelman perusteissa määriteltyä osaamista olisi jäänyt piiloon, kuviossa 20 esitetyllä tavalla. Kuvion 20 palkkia tulisi tulkita silloin niin, että sitä enemmän oppilailta on tavoiteltua ja oikeaa osaamista, mitä heikommin he ovat menestyneet seuranta-arvioinnissa. Tämä oletus johtaisi kuitenkin ristiriitaan, sillä tutkimusaineiston perustella tiedetään, että oppilaiden seuranta-arviointimenestys ja arvosanat korreloivat vahvasti, samoin koulujen sisällä. Oletus vaatisi jatko-oletuksen, jonka mukaan mitä huonompi seuranta-arviointimenestys ja mitä huonommat arvosanat oppilaalla on, sen enemmän oppilaalla olisi piiloon jäänyttä, ”todellista, opetussuunnitelmaan perustuvaa osaamista”.

Vaikka seuranta-arviointi menisi osittain kouluissa opettujen tietojen ja taitojen ohi ja siksi seuranta-arviointi osuisi vaihtelevasti koulujen painotuksiin, on edelleen todennäköistä, että hyviä yhteiskuntaopin arvosanoja saaneet oppilaat osaavat siitä osasta, joka osui seuranta-arvioinnissa kohdalleen, enemmän kuin ne oppilaat, joiden osaaminen yhteiskuntaopissa on heikkoa, kuten kuviossa 20 on esitetty. Aiemmin tässä tutkimuksessa osoitetut oppilastason ja koulujen sisäiset vahvat korrelaatiot selittyvät silloin näin, mutta koulujen väliset seuranta-arvioinnissa osoitetun osaamisen ja arvosanojen väliset erot jäisivät yhä ratkaisematta.

Kanen (2006, 23) mukaan tulkittujen tutkimustulosten validius edellyttää, että tulkittava argumentti on johdonmukainen, tulkinnat ovat järkeviä ja niistä tehdyt oletukset ovat uskottavia. Toulminin mukaan esitetyn väitteen tulee olla myös mahdollisimman hyvin perusteltu ja väitettä tulee tukea osoittamalla, mihin se perustuu (Toulmin 2003, 90).

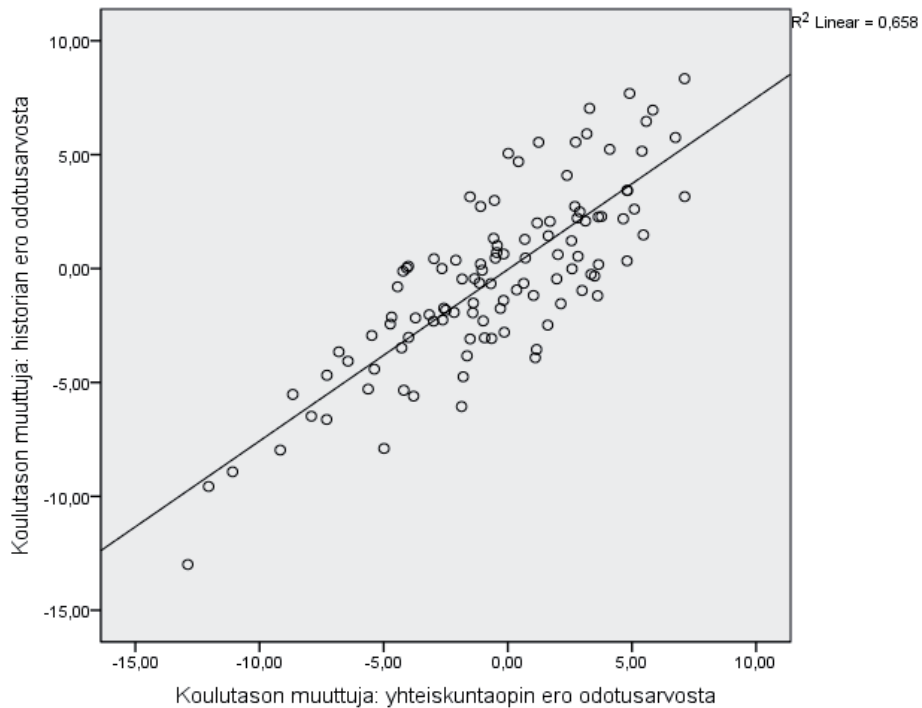
Yhteiskuntaopin seuranta-arvioinnissa osoitetun osaamistason vertailemiseksi historian osaamistason kanssa on seuraavassa laskettu uusi koulutason muuttuja, historian ero odotusarvosta, joka on samalla tavalla kuin yhteiskuntaopissa, koulun keskimääräinen poikkeama historian seuranta-arvioinnissa osoitetusta odotusarvosta. Kuvioista 21 ilmenee, että yhteistä vaihtelua historiassa on 76 %.



KUVIO 21. Koulujen keskimääräinen historian ratkaisuosuus suhteessa koulutason muuttajaan eli eroon odotusarvosta.

Oppilaiden seuranta-arvioinneissa osoitettu osaaminen ja oppilaiden arvosanojen mukaan laskettujen odotusarvojen ylitykset tai alitukset korreloivat vahvasti kummassakin oppiaineessa. Yhteiskuntaopin seuranta-arviointimenestys ja ero odotusarvosta yhteiskuntaopissa korreloivat koulutasolla erittäin vahvasti ($r_{XY} = 0,90$). Näin tapahtui myös historian seuranta-arviointimenestyksen ja historian eron odotusarvosta suhteen: korrelaatio oli ($r_{XY} = 0,87$). Samalla on nähtävissä se systematiikka, joka muodostuu koulujen osaamistason ja oppilasarviointin välillä: kun osaamistaso eli seuranta-arvioinnissa saavutettu keskimääräinen ratkaisuosuus on ollut hyvä, se on ollut hyvä myös keskimääräisiin arvosanoihin nähden. Sama systematiikka on havaittavissa yhteiskuntaopin lisäksi myös historiassa. Samankaltaisen huomion ovat tehneet yläluokkien oppimaan oppimista tutkineet Hautamäki ja muut, jotka huomasivat, että oppimaan oppimisen kannalta heikoiksi määritellyissä luokissa annettiin ”korotettuja arvosanoja” (Hautamäki ym. 2000, 241, 243). Opetushallituksen arvioitsijat ovat tehneet havainnon matematiikan seuranta-arvioinnissa siitä, että ”Koulun arvosanakäytäntö on yleensä ollut sitä ankarampi, mitä parempi on ollut koulun tulos. Toisin sanoen vaativa arvosanakäytäntö ja hyvä keskimääräinen osaaminen näyttävät kulkevan käsi kädessä. Koulujen arvosanalinjien ero matematiikassa on ollut suurimmillaan yli 20 prosenttiyksikköä.” (Mattila 2005, 110.) Matematiikan lisäksi myös äidinkielessä on ollut tyypillistä, että koulun vaativa arvosanalinja ja hyvä keskimääräinen osaaminen liittyivät toisiinsa (esim. Lappalainen 2004, 141).

Yhteiskuntaopin lisäksi samoilta oppilailta, samoista kouluista ja samaan aikaan kerättiin myös historian seuranta-arviointiaineisto. Kuviossa 22 on esitetty tutkimusaineistosta laskettu todellinen tilanne, jossa historian koulutason muuttuja eli ero odotusarvosta selittää tilastollisesti 66 % yhteiskuntaopin koulutason muuttujasta tai toisin päin. Kuvioista siis ilmenee aito empiirinen yhteys, jossa koulujen odotusarvoista lasketut ylärajan ylitykset ja alarajojen alitukset ovat samoissa kouluissa, joissa ovat samat oppilaat ja opettajat, mutta nyt tarkastelussa on kaksi eri oppiainetta ja niiden seuranta-arvioinneista saadut tulokset.



KUVIO 22. Erot odotusarvosta yhteiskuntaopissa ja historiassa.

Koska yhteiskuntaopin seuranta-arvioinnin lisäksi historian seuranta-arviointi suoritettiin samaan aikaan, samoissa kouluissa ja samoilla oppilailla, voidaan oletuksia

- a) seuranta-arvioinnissa on mitattu vääriä asioita tai
- b) seuranta-arviointi on ollut liian kapea-alainen

tarkastella vielä toisen oppiaineen eli historian seuranta-arvioinnin avulla. Mikäli edellä esitetyt oletukset a) ja b) pitäisivät paikkansa, siitä seuraisi joko

1. yhteiskuntaopin ja historian seuranta-arvioinnit mittaavat vääriä asioita tai
2. yhteiskuntaopin seuranta-arviointi mittaa vääriä asioita, mutta historian seuranta arviointi on kapea-alainen tai
3. historian seuranta-arviointi mittaa vääriä asioita, mutta yhteiskuntaopin seuranta-arviointi on kapea-alainen tai
4. molemmat seuranta-arvioinnit ovat kapea-alaisia.

Esitetyistä vaihtoehtoista ensimmäinen on jo aiemmin tässä luvussa osoitettu paik-kansapitämättömäksi, koska seuranta-arviointi ja oppilaiden arvosanat korreloivat vahvasti oppilastasolla ja koulujen sisällä. Esitetyistä vaihtoehtoista viimeinen eli se, että kummatkin seuranta-arvioinnit olisivat kapea-alaisia, lienee todennäköisin ja vahvin oletus. Mikäli oletetaan, että arvosanat mittaavat validisti oppilaiden osaa-mista, mutta seuranta-arvioinnin validius pettäisi oletuksen 4) mukaisesti siksi, että kummatkin seuranta-arvioinnin tehtäväsarjat olisivat kapea-alaisia, se tarkoittaisi, että seuranta-arvioinnin tehtäväryhmä olisi onnistunut laatimaan samalla tavoin kaksi epävalidia seuranta-arvioinnin tehtäväsarjaa kahteen eri oppiaineeseen, jolla on eri oppiainespesifit tavoitteet, sisällöt ja kriteerit. Tällöin yhteiskuntaopin lisäksi myös historian seuranta-arvioinnin ulkopuolelle olisi jäänyt piiloon oppilaiden osaamista. Mikäli näin olisi, tätä seuranta-arviointien ulkopuolelle jäävää osaamista tulisi olla samoissa kouluissa, samoilla oppilailla ja kahdessa eri oppiaineessa, joiden tavoitteet, sisällöt ja esimerkiksi opetuksessa käytetyt oppikirjat ovat kuitenkin eri-laisia. Johdonmukaisin ajatus olisi, että kummassakin oppiaineessa, yhteiskuntao-pissa ja historiassa, olisi jokin yhteinen eri opettajien painotuksia ohjaava elementti, jonka tulisi olla samalla tavalla mukana myös seuranta-arvioinnissa, koska muuten piiloon jäävää osaamista olisi eri tavalla eri kouluissa.

Nyt kuitenkin tiedetään kuvion 22 perusteella, että todellisuudessa yhteiskuntaopin ero odotusarvosta selittää tilastollisesti 66 % historian erosta odotusarvoista tai toi-sin päin. Kuviosta ilmenee aito empiirinen yhteys, jossa koulujen odotusarvoista lasketut ylärajan ylitykset ja alarajojen alitukset olivat samoissa kouluissa, joissa olivat samat oppilaat ja opettajat, mutta nyt tarkastelussa oli kaksi eri oppiainetta ja niiden seuranta-arvioinnista saadut tulokset. Jotta kuviossa 22 havainnollistettu tulos kahden seuranta-arvioinnin, yhteiskuntaopin ja historian, perusteella saavute-tuista koulujen odotusarvojen mukaan lasketusta poikkeavasta osaamisesta voidaan varmistaa ja voidaan olla varmoja siitä, että kaksi seuranta-arviointia yhteiskuntao-pissa ja historiassa eivät sisällä yhteistä seuranta-arvioinnin ulottumattomiin jäävää elementtiä, voidaan oletusta testata vielä ns. ulkoisen ankkurin avulla.

Yhteiskuntaopin ja historian seuranta-arviointia ei tarjottu keväällä 2011 maksu-palvelutuotteena koulutuksen järjestäjille niin, että halukkaat koulutuksen järjestä-jät tai koulut olisivat voineet tilata Opetushallitukselta yhteiskunnallisten aineiden seuranta-arvioinnin. Tästä syystä yhteiskunnallisissa aineissa ei ole ollut mahdollista saada käsitystä siitä, miten koulut erilaistuvat vanhempien koulutustason suhteen. Kuitenkin tiedetään, että vanhempien koulutustaustalla on vahva yhteys seuranta-arviointituloksiin ja oppilaiden arvosanoihin. Tässä tutkimuksessa ns. ulkoinen ankkuri on muuttuja, jonka tiedetään olevan yhteydessä keskimääräisiin seuranta-arviointituloksiin. Kyseinen muuttuja on siis vanhempien koulutustausta.

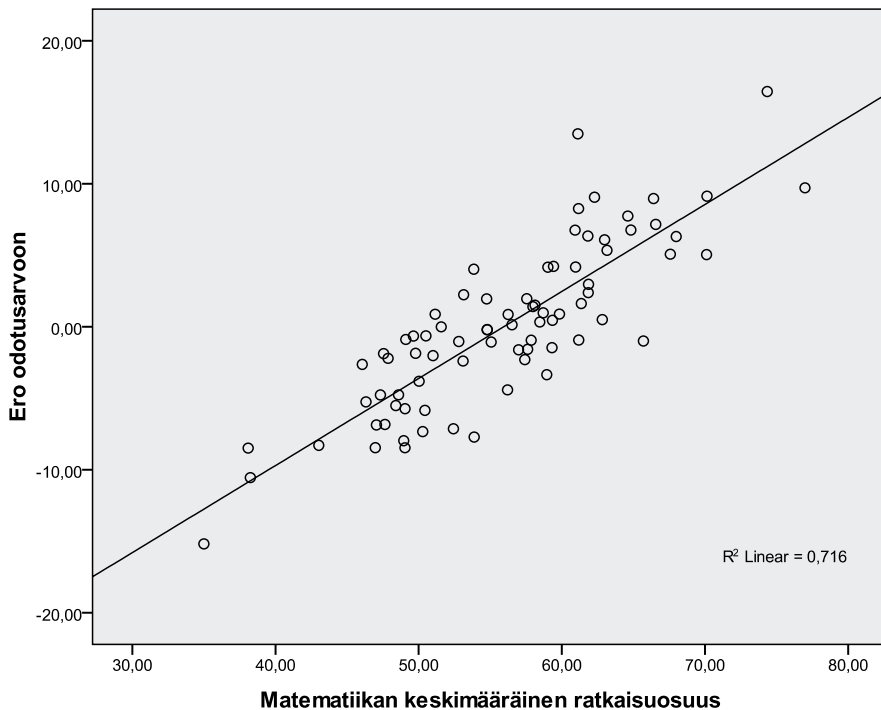
Pääkaupunkiseudun kunnat (Helsinki, Espoo, Kauniainen ja Vantaa) tilasivat yhteensä 79 koululle ja 6 231 oppilaalle matematiikan seuranta-arvioinnin. Mukana oli myös muutama matematiikan seuranta-arvioinnissa ollut otoskoulu. Matematiikka oppiaineena sisältää hyvin vähän mahdollisuuksia tulkita oppiaineessa määriteltyjä tietoja ja taitoja eikä matematiikan oppiaineen luonteeseen kuulu samalla tavalla tulkinnallisuus kuin yhteiskuntaoppiin tai historiaan. Matematiikan oppimistuloksia on lisäksi arvioitu varsin pitkään, vuodesta 1998 lähtien, joten matematiikan osaamisen mittaamiselle on jo pitkät perinteet ja vakiintuneet käytänteet.

Lisäksi tiedetään, että pääkaupunkiseudulla koulut erilaistuvat voimakkaasti sekä vanhempien koulutuksen että keskimääräisen osaamisen suhteen: mitä korkeampi vanhempien koulutustaso on, sen paremmat oppimistulokset ovat.⁷⁴ Yhteys vanhempien koulutustaustan ja keskimääräisen matematiikan osaamisen välillä pätee sekä oppilas- että koulutasolla.

Kuviossa 23 on esitetty koulujen matematiikan keskimääräisen osaamistason suhde keskimääräiseen eroon matematiikan odotusarvosta.⁷⁵ Matematiikassa ero odotusarvosta on laskettu samalla tavalla kuin yhteiskuntaopissa luvussa 6.3.3 on esitetty: jokaisen oppilaan todellista matematiikan seuranta-arvioinnissa saamaa ratkaisuprosenttia on verrattu hänen matematiikan arvosanansa perusteella saamaansa odotusarvoon laskemalla näiden erotus (oppilaan todellinen matematiikan ratkaisuprosentti – matematiikan arvosanan perusteella laskettu odotusarvo = ero odotusarvoon). Kuvioista 23 ilmenee sama matematiikassa kuin yhteiskuntaopissakin: kun on osattu hyvin matematiikan seuranta-arvioinnissa, on osattu hyvin myös keskimääräisiin arvosanoihin nähden.

74 Kaikki kevään 2011 seuranta-arvioinnit viittaavat siihen, että vanhempien koulutuksella on vahva tilastollinen yhteys oppimistuloksiin (Kuusela 2012, 11, 21–24). Jo aiemmin (2006) Kuusela on osoittanut yhteyden olevan koulutasolla, mutta vain suurissa asutuskeskuksissa, joissa koulut pääsevät erilaistumaan.

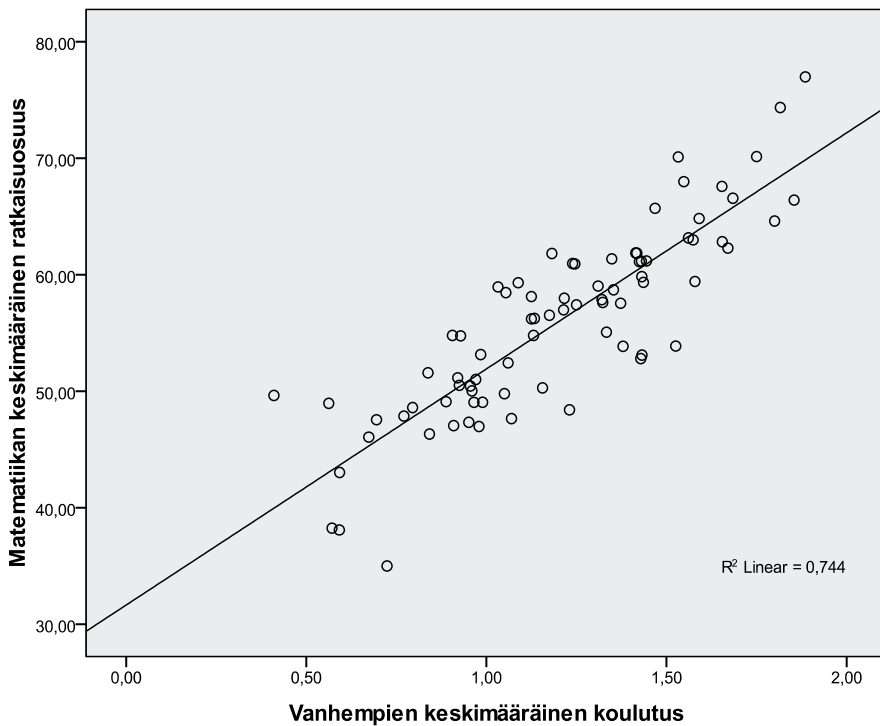
75 Matematiikan oppimistuloksiin liittyvät kuviot 23, 24 ja 25 ovat Kuuselan tekemiä.



KUVIO 23. Matematiikan keskimääräisen ratkaisuosuuden suhde keskimääräiseen eroon matematiikan odotusarvosta.

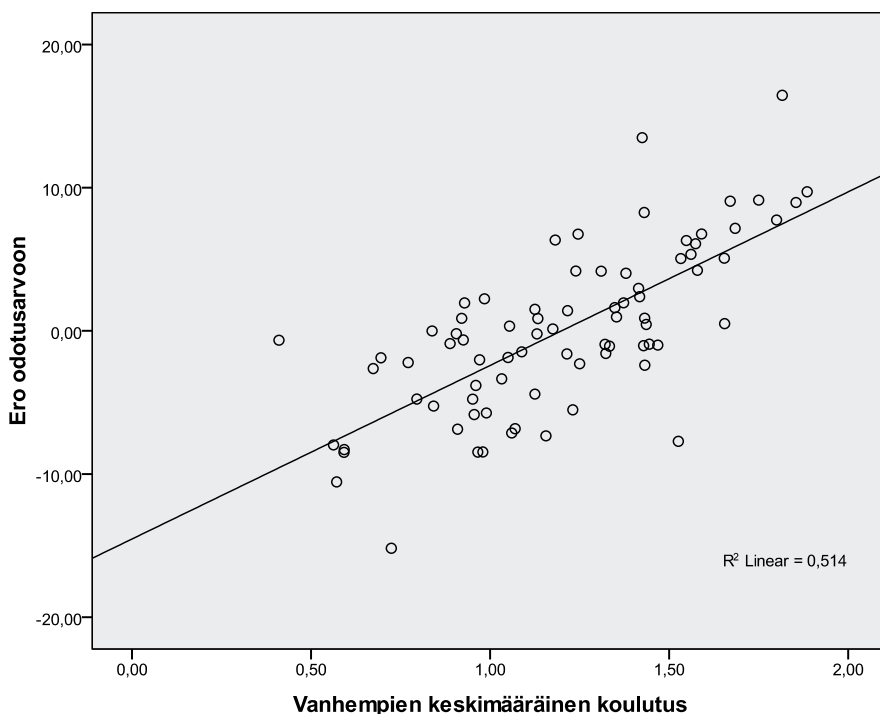
Pääkaupunkiseudun kuntien kaikille kouluilleen tilaaman matematiikan seuranta-arvioinnin avulla voidaan osoittaa, että sama säännönmukaisuus koulujen yleisen osaamistason yhteydestä oppilaiden osaamiseen pätee matematiikassa aivan samalla tavalla kuin yhteiskuntaopissa tai historiassa. Näin siitä huolimatta, että kysymyksessä ovat eri oppiaineet, eri opettajat, eri oppilaat ja eri koulut. Lisäksi arvosanojen paremman validiuden puolesta puhuva abstrakti ”todellisen piiloon jäävän osaamisen” oletus saisi seurakseen oletuksen, että matematiikan opetus osuu sitä huommin yhteen oikean osaamisen kanssa, mitä koulutetumpia ovat oppilaiden huoltajat keskimäärin.

Kuviossa 24 on esitetty vanhempien koulutustason yhteys matematiikan keskimääräiseen ratkaisuosuuteen, josta ilmenee, että mitä koulutetumpia vanhemmat ovat, sitä paremmin oppilaat ovat menestyneet matematiikan seuranta-arvioinnissa.



KUVIO 24. Vanhempien keskimääräisen koulutustason yhteys matematiikan keskimääräiseen osaamistason.

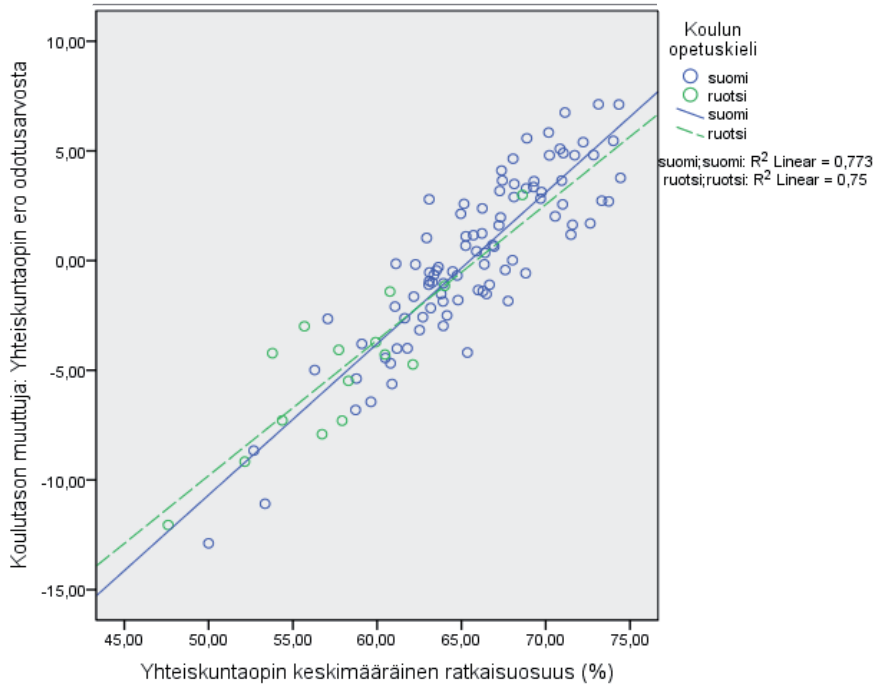
Kuviossa 25 on kuvattu vanhempien keskimääräisen koulutustason yhteys matematiikan eroon odotusarvosta, joka on siis laskettu vähentämällä matematiikan arvosanan perusteella lasketusta odotusarvosta oppilaan todellinen matematiikan ratkaisuosuus. Jos oletetaan, että matematiikan arvosanat ovat validimpi mitta, pitäisi kuvion 25 mukaan olettaa, että matematiikan seuranta-arviointi osuisi parhaiten kohdalleen niissä kouluissa, joiden oppilaiden huoltajat ovat korkeammin koulutettuja.



KUVIO 25. Vanhempien keskimääräisen koulutustason yhteys eroon matematiikan odotusarvoista.

Luvussa 6.1.3 kiinnitettiin huomio kahteen oppilastason yhteyteen. Opetuskielillä oli tilastollinen yhteys sekä yhteiskuntaopin että historian ratkaisusuuksiin, mutta ei arvosanoihin, ja läksyjen tekemisellä oli voimakkaampi yhteys oppilaiden saamiin arvosanoihin kuin ratkaisusuuksiin. Ensimmäinen havainto voidaan nyt tulkita, ja jälkimmäinen vaatii varmistuksen, ettei läksyjen tekeminen ja sen kouluittainen vaihtelu tarjoa selitystä koulujen vaihteleville arviointikäytänteille.

Ruotsinkielisten koulujen oppilaat saivat parempia yhteiskuntaopin arvosanoja kuin suomenkielisten koulujen oppilaat suhteessa seuranta-arvioinnissa osoitettuun osaamiseen. Tarkasteltaessa koulutasolla suomenkielisten ja ruotsinkielisten koulujen keskimääräistä ratkaisusuutta suhteessa yhteiskuntaopin eroon odotusarvosta, kuviosta 26 ilmenee, että suomenkielisten ja ruotsinkielisten koulujen arvosanakäytännöt eivät eroa toisistaan, sillä regressiosuorat kulkevat lähes päällekkäin. Ainoa syy suomen- ja ruotsinkielisten koulujen oppilaiden eroon oli ruotsinkielisten koulujen oppilaiden heikompi keskimääräinen ratkaisusuus yhteiskuntaopissa. Seuranta-arvioinnissa saavutettuun osaamistasoon suhteutettuna ruotsinkielisten koulujen arvosanakäytännöt noudattivat täsmälleen suomenkielisten koulujen arvosanakäytänteitä.



KUVIO 26. Suomen- ja ruotsinkielisten koulujen oppilaiden keskimääräinen ratkaisuosuus suhteessa yhteiskuntaopin eroon odotusarvoista.

Kotitehtävien tekeminen selitti enemmän oppilaiden välisiä yhteiskuntaopin arvosanaeroja kuin yhteiskuntaopin ratkaisuosuuksien eroja, kuten taulukossa 12 on esitetty. Olisi täysin perusteltua olettaa, että kouluissa, joissa oppilaat tekevät säännöllisesti läksynsä, oppilaiden osaamista palkittaisiin parempina arvosanoina. Kun tarkastellaan koulutasolla säännöllisen läksyjien tekemisen korrelaatiota yhteiskuntaopin keskimääräisen ratkaisuosuuden ($r_{XY} = 0,34$) ja keskimääräisten arvosanojen ($r_{XY} = 0,29$) välillä, näyttääkin siltä että, toisin kuin oppilastasolla, korrelaatio on koulutasolla vahvempi keskimääräiseen ratkaisuosuuteen kuin keskimääräisiin arvosanoihin nähden. Läksyjien säännöllinen tekeminen korreloi kuitenkin positiivisesti ($r_{XY} = 0,25$) koulutason muuttujaan eli eroon yhteiskuntaopin odotusarvosta. Tulos tarkoittaa, että mitä paremmin oppilaat keskimäärin tekevät läksyt, sen heikompia arvosanoja he ovat keskimäärin saaneet suhteessa seuranta-arvioinnissa osoitettuun osaamiseen. Läksyjien säännöllinen tekeminen ei siis selitä koulujen eroja odotusarvoista. Pikemminkin saadun tuloksen valossa on ilmeistä, että kouluissa, joissa tehdään säännöllisesti läksyjä, osataan hyvin ja siksi arvosanan antamisperusteet ovat myös tiukat.

6.4.2 Väitteen, arvosanat eivät ole validi mittari, koetteleminen

Hyvä argumentointitapa edellyttää, että esitetyn väitteen tai tulkinta-argumentin rinnalla tarkastellaan toista vaihtoehtoista väitettä, teoriaa tai argumenttia. Jos nyt oletetaan, että arvosanat eivät ole validi mittari, se tarkoittaisi joko sitä, että

- a) arvosanat on annettu opetussuunnitelman perusteissa määritellyistä kriteereistä ”väärin” tai poikkeavin perustein tai
- b) on olemassa jokin arvosanakäytänteitä tasoittava elementti.

On mahdollista, että yhteiskuntaopin arvosanat olisi annettu opetussuunnitelmassa määritellyistä kriteereistä ikään kuin ”väärin”,⁷⁶ säädöksistä poikkeavin perustein, koska Perusopetuksen opetussuunnitelman perusteissa (2004) on annettu yhteiskuntaopin opetuksen ja arvosanojen antamisperusteiden suhteen paljon väljyyttä ja tulkinnanmahdollisuutta (POPS 2004, 226–227). Mikäli oletus a) eli arvosanat on annettu poikkeavin perustein, pitäisi paikkansa, se tarkoittaisi, että oppilaiden saamat arvosanat ja heidän seuranta-arvioinnissa osoittamansa osaaminen eivät korreloisi toistensa kanssa oppilastasolla ja koulujen sisällä. Luvuissa 6.1.1 ja 6.1.2 on osoitettu, että näin ei kuitenkaan ole. Tämän tutkimuksen empiirisellä aineistolla luvussa 6.1.3 on osoitettu, että yhteiskuntaopin arvosanat ja seuranta-arvioinnissa osoitettu osaaminen olivat myös johdonmukaisesti yhteydessä niihin taustamuuttujiin, joihin niiden tuleekin olla. Yhteiskuntaopin tulosten vertailemiseksi käytettiin vielä kahden muun oppiaineen eli historian ja matematiikan arvosanatiejoja ja seuranta-arviointeja. Oletus a) siitä, että yhteiskuntaopin arvosanat on annettu ”väärin” tai arviointikriteereistä poikkeavin perustein ei voi pitää paikkaansa myöskään siksi, että myös historiassa ja matematiikassa arvosanat ja seuranta-arvioinnissa osoitettu osaaminen käyttäytyivät samalla tavalla kuin yhteiskuntaopissa sekä oppilastasolla, koulujen sisällä että koulujen välillä.

Kuten tämän tutkimuksen luvussa 3.3.4 on esitetty, on mahdollista, että Perusopetuksen opetussuunnitelman perusteet (2004) eivät tue riittävästi arvosanojen antamista. Arvioidessaan oppilaitaan, opettajat saattavat käyttää oman koulunsa oppilaita viiteryhmänä arvosanoja antaessaan, varsinkin kun vanhat säädökset ovat ennen antaneet siihen mahdollisuuden. Myös opettajien itsensä laatimien koulukokeiden arvioinnissa opettajat saattavat suhteuttaa koepisteet ja – arvosanat oppilasryhmään. Näin arvosanakäytänteitä tasoittava elementti muodostuu siitä, kuinka tiedoiltaan ja taidoiltaan hyvästä tai heikosta oppilasryhmästä on kyse.

76 Suomen kielen ilmaisu siitä, että arvosanat olisivat annettu ”väärin perustein” on harhaanjohtava ja siksi tässä yhteydessä käytetään paremmin todellista tilannetta kuvaavaa ilmaisua, joka perustuu siihen, että opettajat ovat mahdollisesti antaneet arvosanat voimassa olevista normeista ja Perusopetuksen päättöarviointikriteereistä (2004) poikkeavin perustein.

Oletusta b) arvosanakäytänteitä tasoittavasta elementistä tukee jo aiemmin tässä tutkimuksessa luvussa 6.2.3 osoitettu tulos. Sen mukaan koulujen sisäiset korrelaatiot yhteiskuntaopissa ovat muutamaa poikkeusta lukuun ottamatta korkeammat kuin koulujen keskimääräisten arvosanojen korrelaatiot suhteessa keskimääräiseen seuranta-arviomenestykseen, mutta koulujen välillä yhteiskuntaopin keskimääräiset ratkaisusuudet korreloivat enää heikosti koulujen keskimääräisiin arvosanoihin nähden: korrelaatio oli yhteiskuntaopissa enää 0,21 ja historiassa 0,38.

Jo aiemmin tässä tutkimuksessa kuvioissa 14 ja 15 esitettiin, että kouluissa (Q4), joissa yhteiskuntaopin tai historian seuranta-arvioinneissa menestytettiin keskimäärin hyvin, annettiin yhteiskuntaopissa ja historiassa samoilla ratkaisusuuksilla alempia arvosanoja kuin niissä kouluissa (Q1), joissa keskimääräiset ratkaisusuudet yhteiskuntaopissa ja historiassa jäivät matalammaksi. Systemaattinen yhteys muodostuu koulujen keskimääräisen osaamistason eli seuranta-arvioinnin keskimääräisten ratkaisusuuksien ja oppilaiden keskimääräisten arvosanojen välille. Kouluissa, joiden osaamistaso on hyvä eli keskimääräinen yhteiskuntaopin tai historian seuranta-arvioinnissa saavutettu ratkaisusuus oli korkea, arvosanoja on annettu tiukemmallalla skaalalla. Vastaavasti kouluissa, joiden yhteiskuntaopin tai historian seuranta-arvioinnissa saavuttama ratkaisusuus on ollut matala, arvosanoja on annettu löyhemmin perustein. Arvosanoja tasoittava elementti eli oppilaan arvosanan muodostuminen suhteuttamalla oppilaan osoittama osaaminen suhteessa luokkaan tai opetusryhmään ei ole riippuvainen oppiaineesta, sillä ilmiö on havaittavissa niin yhteiskuntaopissa kuin historiassakin. Kahden eri oppiaineen samanaikaisella arvioinnilla on voitu ratkaista kahden mittauksen eli arvosanojen ja seuranta-arvioinnin validiusongelma, kuten tämä tutkimus on osoittanut.

Tämän tutkimuksen lähtöoletuksena, kuten luvun 6.4 otsikko antaa ymmärtää, on ollut arvosanojen validiuden asettaminen kyseenalaiseksi. Oletus siitä, että arvosanojen validius mahdollisesti pettää, perustuu tässä tutkimuksessa osoitettuihin tuloksiin. Mikäli kuitenkin oletettaisiin, että arvosanat olisivat valideja, se johtaisi puolestaan mahdottomiin jatko-oletuksiin. Kun lisäksi arvosanojen keskiarvot ovat yleensä taipuneet kohti keskimääräistä, kuten kuvioista 14 ja 15 ilmenee, tutkimusaineiston tuleekin käyttäytyä esitetyllä tavalla eli niin, että arvosanojen validius kannattaa kyseenalaistaa ensimmäiseksi. Väitteet arvosanojen validiudesta tai epävalidiudesta laadittiin Toulminin argumentaatiomallin perusteella siten, että tutkimusaineiston perusteella oli olemassa lähtötieto siitä, että jompikumpi osaamisen mittari, seuranta-arviointi tai oppilaiden saamat arvosanat, ei ole validi. Tämän jälkeen lähtötietoa perusteltiin ja tarkennettiin esimerkiksi tutkimalla sitä, että taustamuuttujat ovat johdonmukaisessa yhteydessä seuranta-arviointiin ja arvosanoihin. Kun argumentaatiomallin perusteella oli muodostettu kaksi kilpailevaa väitettä, niitä koeteltiin ja haastettiin validiusargumentoinnin periaatteiden mukaan.

Tutkimusaineiston esittelyssä, luvuissa 6.1 ja 6.2, ilmenevä ratkaiseva ajatus on ollut, että osaaminen kasautuu, eli kun osataan hyvin yhteiskuntaoppia, osataan myös hyvin historiaa. Oletus koskee myös muita aineita, kuten äidinkieltä ja matematiikkaa, sillä esimerkiksi taulukosta 9 nähdään, että oppilaiden matematiikan, äidinkielen, yhteiskuntaopin ja historian arvosanojen keskinäiset korrelaatiot ovat vahvoja. Näin lähekkäin olevissa aineissa kuten yhteiskuntaopissa ja historiassa seuranta-arviointien tulisi siis antaa samansuuntaiset tulokset. Jos siis toisessa aineessa, esimerkiksi yhteiskuntaopissa, on osattu hyvin tai huonosti arvosanoihin nähden, saman pitäisi toistua myös toisessa aineessa eli historiassa, samalla tavalla ja vielä samoissa kouluissa.

Jos kumpikin seuranta-arvioinnin tehtäväsarjoista olisi omalla laillaan ollut kapealainainen ja siksi yhteiskuntaopin ja historian seuranta-arviointi ei olisi kohdentunut koulujen omiin pedagogisiin painotuksiin, lopputuloksena olisi satunnainen kuvio 22. Nyt kuitenkin yhteiskuntaopin ja historian seuranta-arvioinnille on yhteistä se, että osaaminen kasautuu. Yhteiskuntaopin ja historian arvosanoille on puolestaan yhteistä se, että ne tuntuvat taipuvan kohti keskimääräistä. Yhteistä kummassakin oppineessa on, että kun osataan keskimäärin hyvin, osataan hyvin myös keskimääräisiin arvosanoihin nähden.

Kahden oppiaineen, yhteiskuntaopin ja historian, seuranta-arviointiaineistojen sekä niistä saatujen tulosten perusteella voidaan perustellusti todeta, että oppilaan arviointikäytännöt eivät vaihtele niinkään koulujen sisällä, vaan koulujen välillä koulujen yleisen osaamistason eli koulun keskimääräisen ratkaisuosuuden mukaan, joka toimii arvosanoja tasoittavana elementtinä.

7 KESKEISET TULOKSET

Opetushallituksen julkaisemassa yhteiskunnallisten aineiden oppimistulosten seuranta-arvioinnissa todettiin olevan koulujen välisiä eroja seuranta-arvioinnissa osoitetun osaamisen ja oppilaiden saamien arvosanojen välillä (Ouakrim-Soivio & Kuusela 2012, 110–111). Tässä väitöstutkimuksessa pureuduttiin siihen osaan yhteiskunnallisten aineiden seuranta-arviointiaineistoa, joka jäi Opetushallitukselta aiemmin tutkimatta. Toinen olennainen väitöstutkimuksen ja Opetushallituksen seuranta-arvioinnin välinen ero liittyi oppilasryhmään, joka Opetushallituksen seuranta-arvioinnissa käsitti yleisopetuksen oppilaiden lisäksi myös erityistä tukea saavat oppilaat ja muuta kuin suomea tai ruotsia äidinkielenään puhuvat oppilaat. Tässä tutkimuksessa aineistoon kuuluivat vain yleisopetuksen oppilaat, koska tutkimusongelma kiinnittyi niihin säädöksiin ja normeihin, jotka määrittävät arvosanan ja päättöarvosanan antamista yleisopetuksessa. Siksi tämän tutkimuksen empiirinen aineisto ja siitä laaditut analyysit tehtiin kokonaan uudestaan vain yleisopetuksen oppilailta kerätyllä aineistolla. Kerätty yhteiskuntaopin ja historian tutkimusaineisto mahdollisti kahden erillisen oppiaineen välisen vertailun samojen oppilaiden ja koulujen tuottamalla aineistolla sekä kahdella rinnakkaismittauksella, arvosanoilla ja seuranta-arvioinnissa osoitetulla osaamisella.

Tutkimustulokset vastasivat ensimmäiseen tutkimuskysymykseen siten, että oppimistulosten seuranta-arviointi ja arvosanat vastasivat toisiaan hyvin oppilastasolla sekä yhteiskuntaopissa että historiassa. Näiden oppiaineiden arvosanat ja ratkaisuosuudet korreloivat vahvasti yhteiskuntaopissa ($r_{xy} = 0,55$) ja historiassa ($r_{xy} = 0,61$). Lisäksi oppilaiden keskimääräinen yhteiskuntaopin ja historian seuranta-arvioinnissa osoitettu osaaminen korreloi vahvasti keskenään ($r_{xy} = 0,74$). Vaikka Perusopetuksen opetussuunnitelman perusteissa (2004) annetut päättöarvioinnin kriteerit määrittävät arvosanaa 8, näytti tutkimustulosten mukaan siltä, että arvosanan 8 yhteiskuntaopissa tai päättöarvosanan 8 historiassa saavuttaneiden oppilaiden ratkaisuosuuksien keskijajonta oli yhtä suurta kuin muissakin arvosanoissa tai päättöarvosanoissa saavutetuissa ratkaisuosuuksissa. Lisäksi oppilaiden seuranta-arvioinnissa saamat ratkaisuosuudet ja arvosanat olivat johdonmukaisessa yhteydessä sellaisiin taustamuuttujiin, joihin niiden oletettiin olevan.

Tutkimuksen alussa esitetty kysymys siitä, ilmenikö kahden arvioinnin välillä mahdollisia systemaattisia eroja ja jos ilmeni, niin miten, perustui osittain tutkimuksen taustalla olevaan tietoon siitä, että Perusopetuksen opetussuunnitelman perusteiden (2004) mukaan määritellyt osaamisen mittarit, arvosanat ja seuranta-arviointi, näyttivät antavan koulujen välisessä vertailussa sekä yhteiskuntaopissa että historiassa erilaiset mittaustulokset. Kuvioissa 14 ja 15 esitettiin yleisopetuksen oppilailla, että kyse ei ollut esimerkiksi siitä, että Opetushallituksen otannassa mukana olleet erityistä tukea saavien tai muuta kuin suomea tai ruotsia äidinkielenään puhuvien oppilaiden seuranta-arvioinnissa osoitettu osaaminen olisi vaikuttanut

Opetushallituksen tulokseen, vaan myös yleisopetuksen oppilaat saivat kahdessa oppiaineessa, samalla osaamistasolla, erilaisia arvosanoja. Kuvioiden 14 ja 15 perusteella voitiin päätellä, että koulujen välillä oli eroja. Kun seuranta-arviointiin osallistuneet koulut jaettiin osaamiskvartiileittain neljään ryhmään, havaittiin, että kouluissa (Q4), joissa yhteiskuntaopin tai historian seuranta-arvioinneissa menestytettiin keskimäärin hyvin, annettiin yhteiskuntaopissa ja historiassa samoilla ratkaisuosuuksilla alempia arvosanoja kuin niissä kouluissa (Q1), joissa keskimääräiset ratkaisuosuudet yhteiskuntaopissa ja historiassa jäivät matalammiksi.

Saatujen tutkimustulosten mukaan näytti lisäksi siltä, että yhteiskunnallisten aineiden seuranta-arviointi tuotti enemmän vaihtelua koulujen välille kuin koulujen antamat yhteiskuntaopin arvosanat antoivat olettaa. Sisäkorrelaatiolla laskettuna sekä yhteiskuntaopissa että historiassa koulujen välinen vaihtelu oli suurempaa seuranta-arvioinnissa osoitetussa osaamisessa kuin koulujen keskimääräisissä arvosanoissa. Aineiston perusteella näytti siltä, että arvosanat taipuivat kohti keskimääräistä.

Vaikka arvosanojen ja seuranta-arvioinnissa osoitetun osaamisen välinen korrelaatio oli oppilastasolla ja koulujen sisällä vahva, koulujen väliset yhteiskuntaopin keskimääräiset ratkaisuosuudet korreloivat enää varsin heikosti koulujen keskimääräisiin arvosanoihin nähden: korrelaatio yhteiskuntaopissa oli enää 0,21 ja historiassakin vain 0,38. Tutkittaessa koulujen antamia keskimääräisiä yhteiskuntaopin arvosanoja suhteessa koulujen keskimääräisiin ratkaisuosuuksiin, näytti siltä, että koulujen yhteiskuntaopin seuranta-arvioinnissa osoitetun osaamisen ja annettujen yhteiskuntaopin arvosanojen välillä ei olisi ollut juuri yhteyttä. Yhteistä vaihtelua oli yhteiskuntaopissa vain 4,4 % ja historiassa 14,7 %. Tulokset olivat odottamatomia varsinkin, kun ne olivat samankaltaisia myös historiassa. Edellä esitetyt tutkimustulokset kertoivat siitä, että kyseiset kaksi mittausta, yhteiskuntaopin seuranta-arviointi ja arvosanat näyttivät samaan suuntaan oppilastasolla ja koulujen sisällä, kun konvergenssin mittana pidetään korrelaatiokerrointa, mutta ne eivät konvergoineetkaan enää koulujen tasolla tai ainakin tämä konvergenssi oli huono. Saaduista tutkimustuloksista virisi seuraavat tutkimuskysymykset: voidaanko seuranta-arvioinnin perusteella tehdä johtopäätöksiä arvosanojen validiudesta, vai jääkö kahden mittauksen validiusongelma ratkaisematta sekä voidaanko kahden oppiaineen samanaikaisella arvioinnilla ratkaista kahden mittauksen eli arvosanojen ja seuranta-arvioinnin validiutta koskeva ongelma?

Jotta esitettyihin tutkimuskysymyksiin oli mahdollista vastata, käytettiin Kuuselan (2006) aiemmin kehittämää menetelmää epälineaarisuudesta johtuvan ongelman ratkaisemiseksi. Se edellytti, että kullekin oppilaan saamalle yhteiskuntaopin ja historian arvosanasta lasketulle odotusarvolle laskettiin vielä raja-arvot eli vaihteluvälit. Odotusarvo on abstraktio siitä, mihin tasoon ratkaisuosuuden avulla ilmaistuna oppilaan osaamisen tulisi asettua, jos arvosanat ja seuranta-arviointi vastaisivat todellista tilannetta. Sen sijaan poikkeama odotusarvosta kuvaa todellista tilannetta. Mitä paremmin oli osattu, sitä paremmin oli osattu myös keskimääräisiin

arvosanoihin nähden ja kääntäen. Koulujen välistä, voimakasta systemaattista vaihtelua tutkittiin hyödyntämällä selitettävänä muuttujana uutta koulutason muuttujaa, yhteiskuntaopin eroa odotusarvosta. Selittävänä muuttujana oli yhteiskuntaopin seuranta-arvioinnissa saavutettu ratkaisuosuus. Koulujen välistä yhteistä vaihtelua oli yhteiskuntaopissa 81 % ja historiassa 76 %. Saatujen tulosten mukaan, mitä paremmin oli menestytty seuranta-arvioinnissa, sitä paremmin oli menestytty koulun antamiin keskimääräisiin arvosanoihin nähden. Koulujen väliset erot olivat suuria, äärimmillään ratkaisuprosentteina ilmaistuna 20 prosenttiyksikköä, joka vastasi noin neljää kouluarvosanaa.

Saatu tutkimustulos koulujen välisten ratkaisuosuuksien ja arvosanojen systemaattisista eroista sekä yhteiskuntaopissa että historiassa antoivat aiheita kysymykseen siitä, voidaanko seuranta-arvioinnin perusteella tehdä johtopäätöksiä oppilaiden saamien arvosanojen validiudesta. Haasteena oli, ettei yhteiskuntaopin tai historian osaamiselle ole olemassa mitään absoluuttista mitta, johon voisi tehdä vertailuja eikä oppilaiden arvosanojen ja seuranta-arvioinnin lisäksi ollut olemassa kolmatta osaamisen mittaria, joihin arvosanoja tai seuranta-arviointitulosta olisi voitu verrata. Käytettävissä oli ainoastaan kaksi Perusopetuksen opetussuunnitelman perusteissa (2004) määriteltyyn osaamiseen perustuvaa mittaria, arvosanat ja seuranta-arviointi. Näiden kahden osaamista mittaavan mittarin validiuden kannalta oli olennaista tarkastella sitä, voidaanko yhteiskuntaopin osaamista yleensäkin arvioida yhdellä seuranta-arvioinnilla tai yhdellä arviointikerralla, ja jos voidaan, oliko nyt käytetty seuranta-arviointi sellainen, että se mittasi kohtalaisen virheettömästi yhteiskuntaopin osaamista. Mahdollisia väitteitä oli kolme: oppilaiden saamat yhteiskuntaopin arvosanat olivat validi mittari tai seuranta-arviointi mittasi yhteiskuntaopin osaamista validimmin. Kolmas, periaatteellinen väite oli, että kumpikaan tapa mitata yhteiskuntaopin osaamista ei ollut validi, jolloin arvosanat ja seuranta-arviointi olisivat olleet eri tavalla virheellisiä. Tähän kolmanteen vaihtoehtoon ei ollut mahdollista päästä empiirisesti käsiksi ja siksi sitä ei liioin tarkasteltu tässä tutkimuksessa.

Tutkimuskysymysten, voidaanko seuranta-arvioinnin perusteella tehdä johtopäätöksiä arvosanojen validiudesta vai jääkö kahden mittauksen validiusongelma ratkaisematta sekä voidaanko kahden oppiaineen samanaikaisella arvioinnilla ratkaista kahden mittauksen eli arvosanojen ja seuranta-arvioinnin validiutta koskeva ongelma, ratkaisemiseksi muodostettiin Toulminin argumentaatiomallin mukaan kaksi väitettä: arvosanat ovat validi mittari tai arvosanat eivät ole validi mittari. Ennen väitteiden muodostamisen, tutkittiin ja tarkennettiin taustalla ollutta lähtötietoa siitä, että kummassakin oppiaineessa seuranta-arvioinnissa osoitetun keskimääräisten ratkaisuosuuksien ja keskimääräisten arvosanojen välillä oli koulujen välisiä eroja. Väitteen ja johtopäätöksen muodostamiseksi tutkittiin muun muassa sitä, miten oppilaiden arvosanat ja seuranta-arvioinnissa osoitettu osaaminen korreloivat oppilastasolla, koulujen sisällä ja koulujen välillä sekä sitä, miten erilaiset taustamuuttujat olivat yhteydessä seuranta-arvioinnissa osoitettuun osaamiseen ja arvosanoihin.

Kanen (1990; 2006) mukaan argumentointiin perustuva validiustarkastelu edellyttää, että esitetyn väitteen tai tulkinta-argumentin rinnalla tarkastellaan toista vaihtoehtoista teoriaa, oletusta tai argumenttia. Kuten luvussa 4.4 on todettu, tulkinta-argumenttien tulee olla ymmärrettäviä, järkeviä ja niistä johdettujen oletusten tulee olla uskottavia. Validiusargumentointiin kuuluu, että niistä johdetut tulkinnat ovat tarkkaan määriteltyjä ja oletukset arvioidaan olemassa olevan empiirisen näytön perusteella (Kane, 1990; Kane 2001, 339–340). Tulkinta-argumenteista tehtävistä havainnoista ja syistä tulee myös voida tehdä koko perusjoukkoa koskevia yleistyksiä (Kane 2006, 23). Validiusargumentointia voidaan pitää onnistuneena silloin, kun esitetty tulkinta antaa riittävän taustan ja argumentaatiolla on onnistuttu sulkemaan muut tulkinnat pois (Kane 2006, 29). Kahta edellä esitettyä väitteitä arvosanojen validiudesta tai arvosanojen epävalidiudesta koeteltiin validiusargumentointiin kuuluvan ”jos-niin”-säännön avulla.

Kun tiedettiin, että koulujen keskimääräisiä seuranta-arvioinnissa saavutettuja ratkaisusuosuuksia ja keskimääräisiä arvosanoja verrattaessa toisiinsa toisen mittarin validius pettää, valittiin ensin koeteltavaksi väitteeksi, että arvosanat ovat validi mittari. Jos arvosanat olisivat validi mittari ja seuranta-arviointi ei, se olisi tarkoittanut sitä, että

- a) seuranta-arviointi olisi mitannut väärää asioita tai
- b) seuranta-arviointi olisi ollut liian kapea-alainen.

Tutkimusaineisto osoitti kuitenkin, että seuranta-arvioinnissa saavutetut ratkaisuosuudet ja oppilaiden arvosanat korreloivat vahvasti sekä oppilastasolla että koulujen sisällä. Siksi ei voitu pitää uskottavana, että oletus a) olisi pitänyt paikkansa.

Jos taas oletus b) seuranta-arvioinnin kapea-alaisuudesta olisi pitänyt paikkansa, se olisi tarkoittanut sitä, että seuranta-arvioinnissa oppilaiden osaamista olisi saattanut jäädä piiloon. Mikäli yhteiskuntaopin osaamista olisi jäänyt piiloon, se olisi puolestaan voinut tarkoittaa sitä, että:

1. Opetushallituksen laatimalla yhteiskunnallisten aineiden seuranta-arvioinneilla enempää kuin muidenkaan oppiaineiden oppimistulosten seuranta-arvioinneilla ei ole tavoitettu eikä tavoiteta todellista oppiaineen osaamista tai
2. yhteiskuntaoppi ja historia ovat sellaisia oppiaineita, joissa on poikkeuksellisen vaikea tavoittaa oppilaiden todellista osaamista tai
3. oppilaille olisi sitä enemmän muuta, piiloon jäänyttä osaamista, mitä heikommin he olisivat menestyneet seuranta-arvioinnissa.

Ensimmäinen ja toinen esitetyistä oletuksista ei voinut pitää paikkaansa, koska asiantuntijatriangulaatioperiaatteen mukaan Opetushallitus on pyrkinyt saamaan tehtäväryhmän jäseniksi sekä arvioitavan oppiaineen että oppimistulosten arvioinnin parhaat asiantuntijat, niin myös yhteiskunnallisten aineiden seuranta-arvioinnissa. Ei myöskään ollut perusteltua olettaa, että yhteiskuntaoppi ja historia olisivat niin poikkeuksellisia oppiaineita, että niiden oppimistuloksia ei voitaisi mitata. Esitetyistä oletuksista kolmas oli järkevin ja todennäköisin.

Mikäli oletus 3) olisi pitänyt paikkansa, kyse olisi voinut olla siitä, että seuranta-arviointi olisi osunut vaihtelevasti eri koulujen tai eri opettajien painotuksiin, sillä opettajien pedagoginen vapaus kuuluu olennaisena osana suomalaiseen koulutusjärjestelmään. Tällöin kyse olisi kuitenkin koulujen tasolla tapahtuvasta satunnaisvaihtelusta eli seuranta-arvioinnin reliabiliudesta, ei validiudesta. Mikäli piiloon jäävällä osaamisella taas tarkoitettaisiin sellaista osaamista, joka oli johdettu opetussuunnitelman perusteista, olisi se tarkoittanut sitä, että seuranta-arvioinnin tehtäväsarjat yhteiskuntaopissa ja historiassa olisivat olleet kapea-alaisia, eikä niillä olisi kyetty mittaamaan riittävän monipuolisesti ja laaja-alaisesti erilaisia yhteiskuntaopin tietoja ja taitoja, joita Perusopetuksen opetussuunnitelman perusteiden (2004) mukaan kouluissa opetetaan ja oppilaat osaavat. Kyse olisi tällöin ollut arviointiin liittyvästä epävalidiudesta ja siinä aliedustettavuudesta, joka olisi tarkoittanut sitä, että seuranta-arviointi olisi ollut liian kapea-alainen ja siksi joku käsitevalidiuden ulottuvuus olisi jäänyt ottamatta huomioon (Messick 1989b, 4; 1995, 742–743). Tätäkään oletusta ei voitu pitää empiiristen tutkimustulosten mukaan uskottavana sillä se olisi johtanut siihen, että opetussuunnitelman perusteissa määriteltyä tavoiteltavaa ja oikeaa osaamista olisi jäänyt piiloon sitä enemmän, mitä heikommin oppilaat olivat menestyneet seuranta-arvioinnissa. Oletus kumoutui ristiriitaisena, koska tutkimusaineiston perustella tiedettiin, että oppilaiden seuranta-arviointimenestys ja arvosanat korreloivat vahvasti oppilastasolla, samoin koulujen sisällä. Oletus ”todellisesta piiloon jäävästä osaamisesta” olisi vaatinut jatko-oletuksen, jonka mukaan mitä huonompi seuranta-arviointimenestys ja mitä huonommat arvosanat oppilaalla olisi, sitä enemmän oppilaalla olisi piiloon jäänyttä, ”todellista, opetussuunnitelmaan perustuvaa osaamista”. Kuvion 20 perusteella voitiin kuitenkin olettaa, että vaikka seuranta-arviointi olisi mennyt osittain kouluissa opettajien tietojen ja taitojen ohi ja siksi seuranta-arviointi olisi osunut vaihtelevasti koulujen painotuksiin, oli edelleen todennäköistä, että hyviä yhteiskuntaopin arvosanoja saaneet oppilaat osasivat siitä osasta, joka osui seuranta-arvioinnissa kohdalleen, enemmän kuin ne oppilaat, joiden osaaminen yhteiskuntaopissa oli heikkoa.

Koska yhteiskuntaopin lisäksi yhteiskunnallisten aineiden seuranta-arvioinnissa kerättiin samoilta oppilailta, samoilta kouluilta, samaan aikaan aineisto myös historiassa, saatettiin yhteiskuntaopin tuloksia verrata historian tuloksiin. Mikäli väite arvosanojen validiudesta olisi pitänyt paikkansa, olisi myös historian seuranta-arvioinnissa oppilaiden osaamista jäänyt seuranta-arvioinnin ulkopuolelle. Tällöin myös historian seuranta-arvioinnin ulkopuolelle olisi jäänyt piiloon oppilaiden osaamista

samalla tavalla kuin yhteiskuntaopissakin. Tätä seuranta-arvioinnin ulkopuolelle jäävää osaamista olisi tullut olla samoissa kouluissa, samoilla oppilailla, yhtä paljon kahdessa eri oppiaineessa, joiden tavoitteet, sisällöt ja esimerkiksi opetuksessa käytetyt oppikirjat olivat erilaisia. Tutkimusaineistosta oli mahdollista laskea kahdessa eri oppiaineessa todellinen tilanne, jossa historian koulutason muuttuja eli ero odotusarvosta selitti tilastollisesti 66 % yhteiskuntaopin koulutason muuttujasta tai toisin päin. Tuloksessa ilmeni aito empiirinen yhteys, jossa koulujen odotusarvoista lasketut ylärajojen ylitykset ja alarajojen alitukset olivat samoissa kouluissa, joissa olivat samat oppilaat ja opettajat, mutta nyt tarkastelussa oli kaksi eri oppiainetta ja niiden seuranta-arvioinneista saadut tulokset. Kahden oppiaineen, yhteiskuntaopin ja historian, samanaikainen arviointi, jossa oli samoilta oppilailta, samalla seuranta-arvioinnilla, samaan aikaan ja samoissa olosuhteissa kerätty aineisto, tarjosi ratkaisun kahden mittauksen ongelmaan. Jotta yhteiskuntaopista ja historiasta saatu tulos oli yleistettävissä validiusargumentoinnin periaatteen mukaisesti, seuranta-arvioinnin validiutta tutkittiin vielä ns. ulkoisen ankkurin avulla, joka oli huoltajien koulutustasta. Pääkaupunkiseudun kuntien kaikille kouluille tilaaman matematiikan seuranta-arvioinnin avulla osoitettiin, että sama säännönmukaisuus koulujen yleisen osaamistason yhteydestä oppilaiden osaamiseen päti matematiikassa aivan samalla tavalla kuin yhteiskuntaopissa tai historiassa. Näin siitä huolimatta, että kysymyksessä ovat eri oppiaineet, eri opettajat, eri oppilaat ja eri koulut.

Yhteiskunnallisten aineiden empiirisellä aineistolla tutkittiin vielä lopuksi niitä kahta taustamuuttujaa, koulun opetuskieltä ja annettujen kotitehtävien tekemistä, joilla oli voimakkaampi yhteys joko seuranta-arviointiin tai annettuihin arvosanoihin. Opetuskielillä oli tilastollinen yhteys sekä yhteiskuntaopin että historian ratkaisuosuuksiin, mutta ei arvosanoihin. Kotitehtävien säännöllisellä tekemisellä oli puolestaan voimakkaampi yhteys oppilaiden saamiin arvosanoihin kuin ratkaisuosuuksiin. Ruotsinkielisten koulujen oppilaat saivat parempia yhteiskuntaopin arvosanoja kuin suomenkielisten koulujen oppilaat suhteessa seuranta-arvioinnissa osoitettuun osaamiseen. Kun koulutasolla tarkasteltiin suomenkielisten ja ruotsinkielisten koulujen keskimääräistä ratkaisuosuutta suhteessa yhteiskuntaopin eroon odotusarvosta, ilmeni, että suomenkielisten ja ruotsinkielisten koulujen arvosanakäytännöt eivät eronneet toisistaan. Ainoa syy suomen- ja ruotsinkielisten koulujen oppilaiden eroon oli ruotsinkielisten koulujen oppilaiden heikompi keskimääräinen ratkaisuosuus yhteiskuntaopissa. Seuranta-arvioinnissa saavutettuun osaamistasoon suhteutettuna ruotsinkielisten koulujen arvosanakäytännöt noudattivat täsmälleen suomenkielisten koulujen arvosanakäytännöitä.

Kotitehtävien säännöllinen tekeminen selitti enemmän oppilaiden välisiä yhteiskuntaopin arvosanaeroja kuin yhteiskuntaopin ratkaisuosuuksien eroja. Olisi ollut täysin perusteltua olettaa, että kouluissa, joissa oppilaat tekevät säännöllisesti läksynsä, heitä ”palkittaisiin” paremmilla arvosanoilla. Kun koulutasolla tarkasteltiin säännöllisen läksyn tekemisen korrelaatiota yhteiskuntaopin keskimääräisen ratkaisuosuuden ja keskimääräisten arvosanojen välillä, näytti siltä että, toisin kuin

oppilastasolla, korrelaatio oli koulutasolla vahvempi keskimääräiseen ratkaisuosuuteen kuin keskimääräisiin arvosanoihin. Tulos tarkoitti sitä, että mitä paremmin oppilaat keskimäärin tekivät läksyt, sen heikompi arvosanoja he keskimäärin saivat suhteessa seuranta-arvioinnissa osoitettuun osaamiseen. Läksyjen säännöllinen tekeminen ei siis selittänyt koulujen eroja odotusarvoista. Pikemminkin saadun tuloksen valossa oli ilmeistä, että kouluissa, joissa tehdään säännöllisesti läksyjä, osataan hyvin ja siksi arvosanan antamisperusteet olivat myös tiukat.

Koska hyvä validiusargumentointi edellyttää, että esitetyn väitteen rinnalla tarkastellaan muita kilpailevia vaihtoehtoja, koeteltiin tutkimuksessa vielä väitettä: arvosanat eivät ole validi mittari. Väite sai rinnalleen oletukset:

- a) arvosanat on annettu opetussuunnitelman perusteissa määritellyistä kriteereistä poikkeavin perustein tai
- b) on olemassa jokin koulujen arvosanakäytänteitä tasoittava elementti.

Oletus a) ei voinut pitää paikkaansa, sillä empiiristen tulosten mukaan oppilastasolla ja koulujen sisällä yhteiskuntaopin ja historian arvosanat ja seuranta-arvioinnissa osoittama osaaminen korreloivat vahvasti. Sen sijaan tässä tutkimuksessa esitetyt empiiriset tulokset tukivat oletusta b). Koulujen keskimääräisten ratkaisuosuuksien ja keskimääräisten arvosanojen väliset erot selittyivät arvosanakäytänteitä tasoittavalla elementillä, joka muodostui todennäköisesti siten, että opettaja suhteutti arviointinsa yksittäisestä oppilaasta siihen taustajoukkoon, esimerkiksi luokkaan, opetusryhmään tai koulun saman luokkatason oppilaisiin, joita hän opetti ja arvioi.

Kanen (2006, 29) mukaan validiusargumentointi on onnistunut, kun esitetty tulkinta antaa riittävän taustan ja argumentaatiolla on onnistuttu sulkemaan muut tulkinnat pois. Tämän lisäksi loogisen validiuden eli koetun validiuden periaatteen mukaan se väite, joka antaa oikeamman, perustellumman, loogisemman ja yleisemmän lopputuloksen, tulee valita.

Kahden oppiaineen, yhteiskuntaopin ja historian seuranta-arviointiaineistojen sekä niistä saatujen tulosten perusteella voitiin osoittaa, että vertailtaessa koulujen välisiä eroja seuranta-arvioinnissa saavutettujen keskimääräisten ratkaisuosuuksien ja keskimääräisten arvosanojen suhteen, arvosanat eivät olleet validi mittari. Koulujen arviointikäytänteet vaihtelivat koulujen yleisen osaamistason eli koulun keskimääräisen ratkaisuosuuden mukaan, joka toimi oppilaiden arvosanoja tasoittavana elementtinä. Saatu tutkimustulos yhteiskuntaopissa oli validiusargumentoinnin periaatteiden mukaan yleistettävissä myös historiassa. Lisäksi tulos oli linjassa Kuuselan (2006) ja Mattilan (2010) jo aiemmin vuoden 2002 matematiikan arviointiaineistolla tekemien havaintojen kanssa ja kyseiset tulokset matematiikassa vahvistivat yhteiskuntaopissa ja historiassa saatuja tuloksia.

8 POHDINTA

8.1 Tulosten arviointia

Jotta arvosanojen validiutta koskevan ongelman laajuus konkretisoituisi, tutkimuksessa laskettiin myös se, kuinka suuri oppilasmäärä kävi sellaisissa kouluissa, joissa yhteiskuntaopin arvosanan tai historian päättöarvosanan poikkeamat odotusarvoista olivat suuria. Koulukohtaiset arvosanapoikkeamat odotusarvosta saatiin laskemalla oppilaan saaman tosiasiallisen arvosanan ja seuranta-arviointituloksen perusteella oletetun arvosanan välinen erotus. Mikäli saatu arvo oli negatiivinen, oppilas oli saanut huonomman arvosanan, kun seuranta-arviointituloksella olisi antanut olettaa, jos arvo oli positiivinen, oppilas oli puolestaan saanut paremman arvosanan, kuin olisi voinut olettaa. Mikäli arvo oli nolla, oppilaan saama tosiasiallinen yhteiskuntaopin arvosana tai historian päättöarvosana vastasivat seuranta-arviointituloksen perusteella oletettua arvosanaa. Tämän jälkeen asetettiin kaksi harkinnanvaraista rajaa, joilla pyrittiin kuvaamaan riittävän suurta poikkeamaa. Tiukempi raja asetettiin niin, että koulujen (N = 109) keskimääräiset arvosanat poikkesivat vähintään yhden arvosanan ja matalampi raja asetettiin puolen arvosanan poikkeaman kohdalle joko ylös- tai alaspäin.

Kouluja, joissa yhteiskuntaopin arvosanaerot olivat vähintään yhden arvosanan ylöspäin, oli 17 ja niissä oli oppilaita yhteensä 635. Vastaavasti kouluja, joissa yhteiskuntaopin arvosanaerot olivat vähintään yhden arvosana alaspäin, oli kahdeksan ja oppilaita näissä kouluissa oli yhteensä 299. Näin ollen otoksen oppilaista (N = 4 158) yhteensä 934 eli 23 %, kävi kouluissa (N = 25), joiden keskimääräiset arvosanat poikkesivat vähintään yhdellä arvosanalla jompaankumpaan suuntaan. Kun harkinnanvaraista rajaa laskettiin puoleen arvosanaan, otoskouluista 29:llä yhteiskuntaopin keskimääräiset arvosanat olivat heikompia, kun seuranta-arviointituloksen mukaan olisi voinut olettaa. Näissä kouluissa kävi yhteensä 1 110 (noin 26 %) oppilaista. Vastaavasti niissä 31 koulussa, joissa yhteiskuntaopin keskimääräiset arvosanat olivat parempia, kun olisi voinut olettaa, kävi yhteensä 1 149 (noin 28 %) oppilasta. Näin ollen noin 54 % oppilaista (N = 2 259) oli kouluissa (60 koulua), joissa koulujen keskimääräiset yhteiskuntaopin arvosanaerot olivat enemmän kuin puoli arvosanaa jompaankumpaan suuntaan.

Kun puolestaan tarkastellaan historian arvosanaeroja, kolmessa koulussa kävi yhteensä 114 (noin 3 %) oppilasta, joiden keskimääräiset historian päättöarvosanat olivat vähintään yhden arvosanan heikompia, kun seuranta-arviointituloksen perusteella olisi voinut olettaa ja kymmenessä koulussa, joiden yhteenlaskettu oppilasmäärä oli 296 (lähes 7 %) historian päättöarvosana oli vähintään yhden arvosanan odotusarvoa parempi. Kun rajaa asetettiin puolen arvosanan kohdalle, 17 koulussa kävi yhteensä 604 (noin 15 %) oppilasta, joiden historian päättöarvosana oli odotusarvoa heikempi ja 29 koulussa kävi yhteensä 1 110 oppilasta (noin 26 %),

joiden historian päättöarvosana oli puolestaan odotusarvoa parempi. Tarkasteltaessa historian oppiainetta yhteensä noin 41 % (N = 1 714) oppilaista oli sellaisissa kouluissa (N = 46), joiden keskimääräiset historian päättöarvosanat poikkesivat oletetuista arvosanoista enemmän kuin puoli arvosanaa jompaankumpaan suuntaan.

Edellä kuvattu tulos kertoo siitä, että arvosanojen validiutta koskeva ongelma on laaja ja se antaa aihetta pohtia syvällisemmin sitä, mitä oletuksia tämän tutkimuksen tulosten perusteella tulee tehdä oppilaiden arvosanoista, arvosanojen antamista ohjaavista normeista ja koulujen arvosanojen antamiskäytänteistä.

Huoli koulutuksen tasa-arvosta, erityisesti oppimistulosten koulukohtaisista vaihteluista, johti aikoinaan 1990-luvun lopulla kansallisen oppimistulosten arviointijärjestelmän kehittämiseen (Jakku-Sihvonen 2002a, 11). Järjestelmän tarkoituksena on tuottaa tietoa oppimistulosten laadusta ja perusopetuksen opetussuunnitelmien perusteiden toimivuudesta. Oppimistuloksia käytetään hyödyksi koulutusjärjestelmän ja opetussuunnitelman perusteiden kehittämistyössä, mutta myös käytännön opetustyössä, sillä niiden avulla seurataan koulutuksen tasa-arvon toteutumista. (Koulutuksen tuloksellisuuden arviointimalli 1998, 55.)

Vuoden 1994 Peruskoulun opetussuunnitelman perusteiden ollessa vielä voimassa Opetushallitus julkaisi laajan peruskoulun vuosien 1998–2002 oppimistuloksiin perustuvan raportin, jossa alakoulujen oppilasarvioinnin osalta todettiin, että kouluissa tulisi kiinnittää huomiota oppilasarvioinnin yhdenmukaisuuteen ja oikeudenmukaisuuteen (Jakku-Sihvonen 2002a, 13). Vuosiluokkien 7–9 oppilaiden matematiikan, äidinkielen, luonnontieteiden, A-englannin ja ruotsin kielen seuranta-arviointiaineistojen pohjalta kyseisiltä vuosilta laadittiin meta-aineisto (Jakku-Sihvonen 2002b, 34), joka osoitti, että eri oppiaineiden arvioinneissa ylä- ja alakvartiiliin sijoittuneiden koulujen ratkaisuprosenttien ero oli alimmillaan 13 prosenttiyksikköä ja ylimmillään 21 prosenttiyksikköä (Jakku-Sihvonen 2002b, 38).

Kun nykyisin voimassa olevaa Perusopetuksen opetussuunnitelman perusteita (2004) oltiin kehittämässä, kiinnitettiin päättöarvioinnissa erityistä huomiota oppiainekohtaiseen hyvän osaamisen eli arvosanan 8 kuvaukseen. Hyvän osaamisen kriteereistä ajateltiin muodostuvan mittatikka, johon muut arvosanat suhteutuvat. Tässä tutkimuksessa on kuvioissa 8 ja 9 osoitettu yhteiskuntaopin ja historian arvosanan 8 saaneiden oppilaiden seuranta-arvioinneissa saamat ratkaisuosuudet sekä kuvioissa 10 ja 11 oppilaiden muissa annetuissa yhteiskuntaopin ja historian kouluarvosanoissa saavuttamat ratkaisuosuudet, joista ilmeni, että vaihtelua on paljon eivätkä yhteiskuntaopin arvosana 8 ja historian päättöarvosana 8 näytä muodostavan poikkeusta muihin kyseisissä oppiaineissa annettuihin arvosanoihin nähden. Näiltä osin voi todeta, että hyvän osaamisen kuvaus arvosanalle 8 ei ole onnistunut toimimaan mittatikkuna kummassakaan oppiaineessa.

Kuvioissa 16 ja 17 on yhteiskuntaopin ja historian aineistolla osoitettu, että yhteistä vaihtelua oli yhteiskuntaopissa 4,4 % ja historiassa 14,7 %. Tulosten perusteella näyttää siltä, että koulujen yhteiskuntaopin ja historian seuranta-arvioinnissa osoitetun osaamisen ja annettujen yhteiskuntaopin ja historian arvosanojen välillä ei olisi juuri yhteyttä. Vaikka historiassa yhteistä vaihtelua oli vähän, oli se kuitenkin hieman yli 10 prosenttiyksikköä enemmän kuin yhteiskuntaopissa. Kyseiset erot yhteiskuntaopin ja historian välillä saattavat selittyä osittain sillä, että historian oppiainekohtaisen opetussuunnitelman tavoitteet ja kriteerit on selkeämmin määriteltä, ja ne ovat yhteydessä toisiinsa, kuten luvussa 3.3.4 on osoitettu. Näin voidaan olettaa, että historian oppiaineelle asetetut tavoitteet ja kriteerit tukevat opettajan arviointityötä paremmin kuin yhteiskuntaopin tavoitteet ja kriteerit. Toinen mahdollinen selitys yhteiskuntaopin ja historian yli 10 prosenttiyksikön eroon arvosanojen ja seuranta-arviointien yhteisestä vaihtelusta saattaa olla se, että historia on oppiaineena opettajalle tutumpi, sillä suuri osa yhteiskuntaoppia opettavista opettajista on opiskellut pääaineenaan historiaa, jolloin voidaan olettaa, että opettajat myös tuntevat paremmin historian kuin yhteiskuntaopin didaktiikkaa ja oppilaan arviointiperusteita.

Opettajien oletetaan noudattavan oppilaan arviointia ohjaavia säädöksiä ja normeja. Kuitenkin, yhteiskunnallisten aineiden seuranta-arviointiin osallistuneista opettajista kolmasosa ilmoitti, että he eivät saa historian opetussuunnitelman perusteista riittävästi tukea oppilaan arviointiin, mutta yhteiskuntaopissa noin puolet kyselyyn vastanneista opettajista oli tätä mieltä (Ouakrim-Soivio & Kuusela 2012, 33, 35). Sen sijaan Mattilan tutkimuksessa haastatellut matematiikan ja äidinkielen opettajat kertoivat olevansa perillä Opetushallituksen normeista ja siitä, miten päättöarviointi määräytyy (Mattila 2010, 84).

Tämän tutkimuksen tulokset vahvistavat jo aiemmin Opetushallituksen seuranta-arvioinneissa saatuja tuloksia siitä, että oppilaiden keskimääräiset arvosanat ja seuranta-arvioinneissa osoitettu keskimääräinen osaamistaso korreloivat vahvasti. Myös koulujen sisäiset korrelaatiot yhteiskuntaopin ja historian seuranta-arvioinneissa osoitetun keskimääräisen osaamistason ja kouluissa annettujen yhteiskuntaopin ja historian arvosanojen välillä olivat vahvoja, mutta kun koulujen keskimääräisiä yhteiskuntaopin ja historian ratkaisuosuuksia verrataan oppilaiden keskimääräisiin arvosanoihin, korrelaatiot heikkenevät huomattavasti.

Yhteiskunnallisten aineiden oppimistulosten arvioinnin yhteydessä saadut tulokset sekä tämän tutkimuksen tulokset kertovat siitä, että opettajat tuntevat arviointikriteerit ja arvioivat oppilaita suhteessa oppiaineessa asetettuihin tavoitteisiin, tietoihin ja taitoihin, mutta arvosanojen antaminen tapahtuu todennäköisesti suhteessa joko luokan muihin oppilaisiin tai koulun muihin, saman luokkatason oppilaisiin, eikä suhteessa asetettuihin kriteereihin. On selvää, että käytännössä vain hyvin harvoilla opettajilla on mahdollisuus verrata muiden kuin oman koulunsa oppilaiden osaamistasoa keskenään. Usein ainoana vertailukohteena, johon opettaja oppilaan

suoritusta vertaa, on luokan muut oppilaat tai koulun muut saman vuosiluokan oppilaat. Näin todennäköisesti toimitaan siitä huolimatta, että opetussuunnitelman perusteiden mukaan ”oppilaan edistymistä, työskentelyä ja käyttäytymistä arvioidaan suhteessa opetussuunnitelman tavoitteisiin ja kuvauksiin oppilaan hyvästä osaamisesta” (POPS 2004, 260) eikä suhteessa muuhun oppilasjoukkoon.

Tämä tutkimus on osoittanut yhteiskuntaopin ja historian seuranta-arvioinnissa kerätyllä aineistolla sen systematiikan, joka muodostuu koulujen osaamistason ja oppilasarvioinnin välille: kun koulun osaamistaso on ollut hyvä, se on ollut hyvä myös keskimääräisiin arvosanoihin nähden. Saman huomion ovat tehneet jo aiemmin esimerkiksi Kuusela (2006) sekä Opetushallituksen matematiikan (esim. Mattila 2005) ja äidinkielen (esim. Lappalainen 2004) arvioitsijat. Tämän tutkimuksen tulosten valossa näyttää siis siltä, että oppilaan arvosana annetaan yhä suhteessa muihin oppilaisiin, vaikka virallisesti suhteellinen arviointi poistui normitasolla, kun vuoden 1985 Peruskoulun opetussuunnitelman perusteet astuivat voimaan.

Sytä arviointikulttuurin kerroksellisuuteen ja aiempien arviointiohjeiden käyttämiin oppilaan arvioinnissa on monia. Suomessa opetussuunnitelman perusteiden oppilaan arviointia käsittelevään osuuteen (POPS 2004, 260–265) on omaksuttu aineksia edellisistä opetussuunnitelmien perusteista ilman, että vanhoja ohjeita tai käytänteitä olisi selkeästi purettu. Opetushallitus kuvaa verkkosivuillaan oppilaan arvioinnin kerroksellisuutta 1970-luvun alusta vuoteen 1994 sekä siihen liittyviä ongelmia. Esimerkiksi oppilaan arvioinnin kahta tehtävää, opintojen aikaista arviointia ja päättöarviointia ei erotettu selkeästi toisistaan opetussuunnitelmien perusteissa kolmeenkymmeneen vuoteen. Vaikka suhteellisesta arvioinnista luovuttiin vuonna 1985, sitä ei selkeästi osoitettu perustetekstissä.⁷⁷ Kun suuri joukko arviointityötä tekevistä opettajista on joko ollut itse peruskoulun oppilaina ja tullut arvioiduksi suhteellisen arviointiperiaatteen mukaan tai saanut opettajakoulutuksen suhteellisen arvioinnin ollessa voimassa tai tullut opettajina työelämään 1970- tai 1980-luvulla, ei liene ihme, että suhteellisen arvioinnin käytänteitä on omaksuttu pitkään, vaikka ne oppilaan arviointia ohjaavista säädöksistä on kumottu jo vuosikymmeniä sitten.

Arviointi on aina kontekstisidonnaista. Suomessa oppilaan arvioinnin säädöspohja, ohjeet ja käytänteet poikkeavat monilla tavoin muiden maiden säädöksistä ja käytänteistä, kuten tämän tutkimuksen luvussa kaksi on osoitettu. Kansainvälisistä arvosanan antamista käsittelevistä tutkimustuloksista saattaa löytyä joitain yhteisiä tekijöitä ja selityksiä sille, miksi suomalaisissa opetussuunnitelmien perusteissa jo vuosikymmenten ajan ilmennyt kerroksellisuus (esimerkiksi POPS 1985; POPS 1994) ja opettajien itsenäinen asema arvioijina saattavat johtaa arvosanan luotettavuuden tai arvosanojen valtakunnallisen vertailukelpoisuuden heikkenemiseen,

77 Kyseinen viittaus löytyy oppilaan arviointia käsittelevästä artikkelista Opetushallituksen sivustolta: http://www.oph.fi/saadokset_ja_ohjeet/ohjeita_koulutuksen_jarjestamiseen/perusopetuksen_jarjestaminen/perusopetuksen_oppilaan_arviointi/oppilaan_arviointi_menneina_vuosikymmenina

kuten luvussa 2.3 annetut esimerkit Ruotsista ovat osoittaneet. Toisaalta esimerkiksi Ranskassa arvosanojen luotettavuutta ja arvioinnin objektiivisuutta on pyritty vahvistamaan sekä kansallisesti yhdenmukaisilla ja tarkoilla ohjeilla että oppilaan osaamista arvioivan itsenäisen ryhmän avulla.

Guskeyn ja Baileyn (2001, 11) tutkimus osoitti arvosanan antamista määrittävien ohjeiden ja opettajien arvosanan antamiskäytänteiden välillä olevan eroja. Tutkimuksessaan Guskey ja Bailey löysivät neljä tekijää, jotka vaikuttivat opettajien arvosanan antamiskäytänteisiin, joita olivat 1) opettajien omat kokemukset siitä, miten heitä itseään oli oppilaina arvioitu, 2) se mitä opettajat olivat oppineet opettajakoulutuksen aikana, 3) opettajien henkilökohtainen filosofia tai näkemys opettamisesta sekä 4) se, miten ylempi taho, kuten aluehallinto, koulu tai osasto oli linjannut arvosanan antamista. (Gusekey & Bailey 2001, 16–17.) Brookhart (1994, 284–285; 289; 299) osoitti tekemässään meta-analyysissä, miten opettajien arviointikäytännöt ja arvosanojen antamissuositukset poikkesivat toisistaan esimerkiksi sen suhteen, kuinka opettajat ottivat arvosanaa antaessaan huomioon kirjallisten kokeiden lisäksi oppilaiden muita suorituksia, kuten oppilaiden yritteliäisyyttä suhteessa oppilaiden arvioinnista annettuihin määräyksiin ja ohjeisiin.

Crossin ja Fraryn (1999) tutkimuksen lähtökohtana oli Ebelin ja Frisbien (1986) aiemmin esittämä ajatus siitä, että arvosanan antaminen on aina epäluotettavaa. Heidän mukaansa koulumenestystä ja siinä kehittymistä, erilaisia koulutukseen liittyviä filosofioita ja ristiriitoja on jo teknisestikin vaikea mitata, koska opettajat ovat työssään kaksinaisroolissa opettajana ja arvioijana (Cross & Frary 1999, 53). Brookhartin (1991) lisäksi Cross ja Frary (1999) osoittivat, että 72 % tutkimukseen osallistuneista opettajista ilmoitti nostavansa hieman arvosanoja oppilaille, jotka menestyvät heikommin, tai niille oppilaille, joiden sosiaalinen tausta oli muuta ryhmää heikompi (Cross & Frary 1999, 60). Klapp Lekholm esitti väitöstutkimuksessaan, että Crossin ja Fraryn tutkimuksessa (1999) ilmennyt tulos arvosanojen nostamisesta heikosti menestyvissä ryhmissä tai kouluissa saattaa liittyä esimerkiksi koulun yleiseen arviointikulttuuriin sitä kautta, että opettajat haluavat mahdollisimman monen oppilaan ylittävän hyväksytyt suorituksen rajan (Klapp Lekholm 2008, 87). Mattilan haastattelemat suomalaisopettajat kertoivat ottavansa oppilaiden hankaluudet ja huolet kantaakseen ja opettajat pyrkivät käytössä olevin keinoin saattamaan oppilaansa koulupolulla eteenpäin (Mattila 2010, 170). Crossin ja Fraryn tutkimuksessa opettajat näyttivät palkitsevan oppilaita yrittämisestä, kotitehtävien tekemisestä ja osallistumisesta, vaikka arvosanan antamista ohjaavissa asiakirjoissa kehoitettiin, ettei näin tehtäisi, kyseisten tekijöiden subjektiivisuuden vuoksi. Arvosanoja antaessaan opettajat kertoivat asettavansa oppilaat järjestykseen myös sen mukaan, miten oppilaat olivat menestyneet opettajien itsensä laatimissa kokeissa. (Cross & Frary 1999, 60–69.)

Edellä esitetyt kansainväliset ja kansalliset tutkimustulokset arvosanan antamisesta, seuranta-arviointien tulokset sekä tämän tutkimuksen tulokset palauttavat huomion tämän tutkimuksen varsinaiseen aiheeseen päättöarvioinnin kriteereiden toimivuudesta, arvosanojen validiudesta ja näiden myötä päättöarvosanojen yhdenvertaisuudesta. Cross ja Frary viittasivat tutkimuksessaan (1999), että mittaamisen asiantuntijoilla on laaja yhteysymmärrys siitä, että ainakin lukuaineissa (*academic subjects*) arvosanojen tulisi perustua kulloistenkin saavutusten ja niissä kehittymisen mittaamiseen, eikä arvosanoja annettaessa tulisi huomioida lainkaan esimerkiksi oppilailla olevia erilaisia taitoja, yrittämistä, käyttäytymistä tai muita subjektiiviseen arviointiin perustuvia saavutuksia. (Cross & Frary 1999, 53.) Klapp Lekholmin mukaan niin kauan kun käytössä ei ole kansallisia standardoituja koulukokeita, eivät opetussuunnitelman perusteissa tai paikallisissa opetussuunnitelmissa määritellyt kriteerit arvosanan antamiseksi ole valideja, reliaabeleja, oikeudenmukaisia eivätkä ne ole arvosanoja annettaessa yleistettävissä. Arvosanojen antamiseen ja arvosanoihin on kuitenkin sisäänrakennettu niiden subjektiivinen ulottuvuus. (Klapp Lekholm 2008, 88.) Jotta arvosanat olisivat validi mittari arvioimaan osamista, niiden tulisi olla vertailtavia ja yhdenvertaisin periaattein annettuja. Klapp Lekholmin mukaan Messick (1994) ja Kane (2006) esittivät juuri tästä syystä, että tulkitsevien argumenttien tulisi olla mahdollisimman selviä ja kaikkien oletusten, joita arvosanan antamisprosessista voidaan tehdä, tulisi niin ikään olla läpinäkyviä ja selkeitä (Klapp Lekholm 2008, 90).

Brookhart piti arvosanojen validiuden tarkastelua arvosanojen yleistettävyyden näkökulmasta tärkeänä (Brookhart 1998, 298). Arvioinnin summatiivisuus arvosanoja annettaessa tekee niistä ikään kuin lopullisia, sillä saatua arvosanaa ei voi tehdä tekemättömäksi ja ainoa tehtävissä oleva asia on yrittää korottaa seuraavaa arvosanaa. Mikäli Messickin validiusteoriaa sovelletaan arvosanoihin, edellä esitetty luonnehdinta arvosanojen summatiivisuudesta johtaisi arvosanojen validiuden heikkenemiseen, koska Brookhartin mukaan arvosanojen käyttötarkoituksen ja tulkinnan välinen raja hämärtyy. Ratkaisuna arvosanojen validiuden parantamiseksi Brookhart viittaa Mossin ja muiden tutkimukseen (1992) tulkinnallisten näkökulmien huomioon ottamisesta luokkatyöskentelyn arvioinnissa ja oppilaiden yrittämistä ja edistymistä arvioivien validien ja reliaabelien mittaustapojen kehittämisestä, joka lähtee tavoitteiden asettamisesta. (Brookhart 1994, 298.)

Luvussa 3.3 on esitetty, kuinka yhteiskuntaopin opetuksen painotukset, opetukselle varatut resurssit, oppiaineen tavoitteet sekä oppiaineelle asetetut tiedot ja taidot ovat pysyneet hyvin samanlaisina perusopetuksen alkuvuosista aina 2000-luvulle tultaessa siitä huolimatta, onko yhteiskuntaoppia opetettu yhdessä historian kanssa vai siitä erillisenä oppiaineena. Löfströmin mukaan (2001a, 15) yhteiskuntaopin opetuksen tueksi, toisin kuin esimerkiksi historiassa, ei ole osoitettu mitään käsitteellisiä apuja ja se puolestaan ylläpitää tilannetta, joka Löfströmin, Virran ja van den Bergin (2010, 7) mukaan ilmenee niin, että yhteiskuntaopin opetus on ollut pitkään kuvailtavaa, selostavaa ja toteavaa, mutta ei analyttistä, kriittistä tai problematisoivaa.

Löfström huomauttaa, että suomalainen yhteiskuntaopin opetus on perinteisesti painottunut antamaan tietoja yhteiskunnan instituutioista, ja vasta sen jälkeen oppilaan on ajateltu voivan harjoitella yhteiskunnallista ajattelua, analyysin ja synteessin tekemistä ja argumentointia (Löfström 2012a, 147–148).

Opetussuunnitelmien kerrostuneisuus oppilaan arvioinnissa, arvosanan validius sekä mitattavien tavoitteiden asettamisen tärkeys ilmenevät Virran arvioissa Peruskoulun opetussuunnitelman perusteista (1994) ja niissä määritellyistä tavoitteista, jotka olivat niin yleisiä, että tavoitteiden käyttäminen oppilaan arvioinnin perustana oli ongelmallista. Kun tavoitteiden määrittely ja tulkinta olivat väljiä, ne voitiin eri arviointitilanteissa määritellä hyvin eri tavoin. (Virta 1999, 9.) Vuoden 1999 päättöarvioinnin kriteereissä oli Virran mukaan haasteena se, että niissä vältettiin määrittelemästä vaadittavat sisältöalueet. Sen sijaan esimerkiksi yhteiskuntaopin kriteereissä painottuvat kognitiiviset valmiudet. (Virta 2000, 59.)

Yhteiskuntaopin arvioinnin kerrostuneisuus siirtyi päättöarvioinnin kriteereihin (1999), sillä ne olivat yhteiskuntaopin osalta liian yleisesti määriteltyjä. Virran mukaan kriteerit olivat ylätasoin kuvauksia, jotka eivät kovinkaan selkeästi tarkentaneet, mitä oppilaan olisi todella osattava. Kriteerit jättivät hänen mielestään liian paljon tulkinnan varaa sille, mitä ne merkitsivät ja miten kriteereiden avulla ajateltiin toteuttaa oppilaiden arviointia niin, että oppilaiden oikeusturvaa edistetään samalla kun hänen osaamistaan arvioidaan (Virta 2000, 79–80). Myös Löfström näkee vuoden 1999 päättöarvioinnin kriteereiden ongelmallisuuden ilmenevän siinä, että esimerkiksi yhteiskunnallisen tiedon ymmärtämistä koskeva kriteeri on hyvin yleinen ja jopa ylimalkaisesti määritelty. (Löfström 2001a, 15.) Tässä tutkimuksessa luvussa 3.4.1 on osoitettu, että sama tavoitteiden määrittelyn ylimalkaisuus on siirtynyt Perusopetuksen opetussuunnitelman perusteiden yhteiskuntaopin ainekohtaiseen osaan ja vielä niin, että oppiaineelle asetetut tavoitteet ja päättöarvioinnin kriteerit arvosanalle 8 eivät kiinnity toisiinsa riittävän selkeästi (POPS 2004, 226–227; ks. myös Ouakrim-Soivio 2013, 149–150). Kriteereiden ylimalkaisuudesta seuraa väistämättä, että ne eivät tue oppilaan arviointia riittävästi esimerkiksi päättöarvosanaa annettaessa ja näin ollen päättöarvosanat eivät ole valtakunnallisesti yhteismitallisia, koska koulujen osaamistaso määrittää sen, kuinka tiukkaa tai väljää arviointikulttuuria koulussa noudatetaan, kuten tässä tutkimuksessa on osoitettu. Kyse on paitsi päättöarvosanojen valtakunnallisesta vertailukelpoisuudesta myös oppilaan päättöarvioinnin ja jatko-opintokelpoisuuden yhdenvertaisuudesta ja sitä kautta oppilaan oikeusturvasta.

Luvussa 6.1 on osoitettu, kahden eri oppiaineen, historian ja yhteiskuntaopin identtisesti kerätyllä seuranta-arviointiaineistolla, että historiassa, jossa oppilaat olivat saaneet päättöarviointikriteereiden mukaan päättöarvosanan ja yhteiskuntaopissa, jossa oppilaat olivat saaneet arvosanan opintojen aikaiseen arviointiin perustuen, yhteiskuntaopin ja historian arvosanat käyttäytyivät suhteessa ratkaisuosuuksiin ja taustamuuttujiin hyvin samalla tavalla. Empiirisellä aineistolla saatu tulos antaa aiheutta olettaa, että opettajat eivät juuri tee eroa sen suhteen, arvioivatko he oppilasta

päättöarvioinnin kriteereiden vai opintojen aikaisten hyvän osaamisen (arvosanan 8) mukaiseen kuvauksen perustuen. Kyseinen tulos myös vahvistaa käsitystä siitä, että päättöarvioinnin kriteerit eivät ohjaa riittävän hyvin opettajaa arviointityössä.

8.2 Tutkimuksen arviointia ja jatkotutkimusaiheita

Tämän tutkimuksen tarkoituksena on ollut vastata siihen, miten oppimistulosten seuranta-arviointi ja arvosanat vastaavat toisiaan, esiintyykö kahden arvioinnin välillä mahdollisia systemaattisia eroja ja jos esiintyy, miten ne ilmenevät, sekä kysymyksiin siitä, voidaanko seuranta-arvioinnin perusteella tehdä johtopäätöksiä arvosanojen validiudesta, vai jääkö kahden mittauksen validiusongelma ratkaisematta ja voidaanko kahden oppiaineen samanaikaisella arvioinnilla ratkaista kahden mittauksen eli arvosanojen ja seuranta-arvioinnin validiutta koskeva ongelma.

Tutkimusaineistona on hyödynnetty yhteiskunnallisten aineiden seuranta-arvioinnin tuloksia yhteiskuntaopissa ja tulosten vertailemiseksi myös historiassa. Lisäksi tässä tutkimusaineistossa on käytetty oppilaan oman ilmoituksen mukaista yhteiskuntaopin viimeisintä arvosanaa ja oppilaiden keväällä 2010 saamaa historian päättöarvosanaa.

Tämän väitöstutkimuksen luotettavuutta olisi voinut heikentää se, että väitöskirjan tekijä toimi Opetushallituksessa yhteiskunnallisten aineiden projektipäällikkönä, jonka työtehtäviin kuului osallistua asiantuntijana niin seuranta-arvioinnissa käytettyjen tehtäväsarjojen sekä pisteitys- ja arviointiohjeiden laatimiseen kuin oppilaiden tehtävävihkojen tarkistamiseen, seuranta-arviointiaineiston tulkitsemiseen aineistosta tehtyjen analyysien perusteella kuin seuranta-arviointiraportin laatimiseen kerätyn aineiston perusteella. Opetushallituksella on kuitenkin vakiintuneet menetelmät ja prosessit, joiden mukaan oppimistulosten seuranta-arviointeja on jo vuosien aikana muissa oppiaineissa laadittu, aivan kuten luvussa 1.2 on esitetty. Seuranta-arvioinnin tehtäväsarjoja laatineeseen ryhmään kuului projektipäällikön lisäksi viisi muuta henkilöä ja tehtäväsarjat laadittiin työryhmässä Opetushallituksen ohjeiden mukaan vastaamaan niitä tiedollisia ja taidollisia tavoitteita, jotka on määritelty Perusopetuksen opetussuunnitelman perusteissa (2004) yhteiskuntaopille ja historialle. Seuranta-arvioinnin tehtäväsarja koostettiin esitestauksen perusteella laaditun osio-analyysin jälkeen. Osio-analyysin laativat Helsingin yliopiston Koulutuksen arviointikeskuksen ja Opetushallituksen metodikot. Näin ollen yksittäisen tehtäväryhmän jäsenellä ei ole ollut mahdollisuutta vaikuttaa tehtävien valintaan enempää kuin projektipäälliköllä vaikuttaa Opetushallituksen vakiintuneeseen oppimistulosten arviointiprosessiin. Sen sijaan se, että oppimistulosten seuranta-arviointiprosessi, siinä hyödynnetyt analyysit ja menetelmät ovat vuosien aikana vakiintuneet sekä se, että seuranta-arviointiprosessin läpiviemiseen osallistuu lähes kymmenen asiantuntijaa, lisäävät myös tämän väitöstutkimusaineiston, tässä tutkimuksessa hyödynnetyjen valmiiden mittareiden ja oppimistulosten arvioinnissa vakiintuneiden analyysimenetelmien luotettavuutta.

Tutkimuksen alussa esitettyihin tutkimusongelmiin pystyttiin vastaamaan, joskin yhteiskunnallisten aineiden seuranta-arvioinnin tulosten muutosta ei ole ollut mahdollista vertailla, koska yhteiskuntaopissa ja historiassa ei ole aiemmin tehty oppimistulosten seuranta-arviointia. Tämän tutkimuksen vahvuutena on kuitenkin se, että samoilta oppilailta, samoista kouluista ja samaan aikaan kerättiin yhteiskuntaopin lisäksi toisen oppiaineen eli historian seuranta-arviointitulokset, joita voitiin käyttää yhteiskuntaopin tulosten ristiinvalidoimisessa.

Sekä oppimistulosten arviointeihin yleensä että yhteiskunnallisten aineiden seuranta-arviointiin liittyy Korkeakosken (2005) esittämä arvio, jonka mukaan seuranta-arviointeja laadittaessa tulisi kiinnittää enemmän huomiota siihen, että arviointiin laaditut tehtävät edustavat samanlaista oppimiskäsitystä kuin opetussuunnitelman perusteet. Korkeakoski pitää tärkeänä myös sitä, mitkä oppiaineen sisällöt ja tavoitteet on valittu seuranta-arvioinnin tehtäväsarjoihin, mitkä laadituista tehtävistä ovat valikoituneet seuranta-arvioinnin osaamistasoa mittaaviksi mittareiksi ja mihin tehtävien lopullinen valinta perustuu. (Korkeakoski 2005, 179.) Tämä väitöstutkimus on tarjonnut mahdollisuuden tarkastella tutkimuskysymystä arvosanojen ja seuranta-arvioinnin validiudesta kahteen eri tieteenalaan perustuvan oppiaineen, kahden eri oppiaineen opetussuunnitelmiin perustuvien erilaisten tavoitteiden, sisältöjen ja kriteereiden pohjalta laadituista seuranta-arviointikysymysten lähtökohdista käsin.

Sekä yhteiskuntaopissa että historiassa on ollut välttämätöntä kiinnittää näiden oppiaineiden osaaminen Perusopetuksen opetussuunnitelman perusteiden (2004) mukaan määrittelyiksi tiedoiksi ja taidoiksi, joiden mittaaminen arvosanoilla tai seuranta-arvioinnilla perustuu samoihin säädöksiin ja ohjausasiakirjoihin. Määrittely ei välttämättä tee oikeutta laajemmalle näkemykselle oppilaiden yhteiskunnallisesta tai historiallisesta ajattelusta, jotka ovat aina yksilöllisiä ja tulkinnallisia ja jotka jokainen konstruoi omista lähtökohdistaan käsin. Osaamista enempää kuin asioiden arvoa ei voida määrittää koskaan täysin objektiivisesti, sillä ne perustuvat subjektiivisiin arvostuksiin, ennalta sovittuihin kriteereihin tai eri tavalla todettujen arvojen tulkintoihin (Korkeakoski 2005, 180). Lisäksi arviointitieto on aina jossain määrin suhteellista.

Samalla tavalla kuin arvosanoilla tai seuranta-arvioinnilla mitattava yhteiskuntaopin tai historian osaamistaso on ollut välttämätöntä määrittää suhteessa Perusopetuksen opetussuunnitelman perusteisiin (2004), tutkimuksessa esiintyvä validius on ennen kaikkea käsitteellinen tai teoreettinen mittauksen ominaisuus (Metsämuuronen 2006, 110) ja se on kytketty tässä tutkimuksessa validiusteoreetikkojen näkemyksiin käsitevalidiudesta, sillä nykykäsitys testin, kokeen tai tutkimusasetelman validiudesta perustuu pitkälti käsitevalidiuden periaatteisiin.

Tässä tutkimuksessa on käytetty Kuuselan (2006) kehittämää menetelmää epälineaarisuudesta johtuvan ongelman ratkaisemiseksi, kun haluttiin tutkia koulujen välisiä arvosana- ja osaamiseroja yhteiskuntaopissa ja historiassa. Uuden koulutason muuttujan avulla oli havaittavissa systemaattinen yhteys, joka muodostui koulujen osaamistason ja oppilasarvioinnin välille: kun osaamistaso oli hyvä, se oli hyvä myös keskimääräisiin arvosanoihin nähden. Sama systematiikka havaittiin yhteiskuntaopin lisäksi historiassa, ja tässä tutkimuksessa saatu tutkimustulos on linjassa Kuuselan (2006, 96) aiemman tutkimustuloksen kanssa, jossa hän osoitti matematiikan seuranta-arviointiaineistolla, että arvosanan 8 voi saada koulusta riippuen hyvin eritasoisen osaamisen perusteella.

Tämä väitöstutkimus on lisännyt tietoa koulujen välisistä eroista, jotka muodostuvat arvosanojen ja seuranta-arvioinnissa osoitetun osaamisen suhteen, arvosanojen ja seuranta-arvioinnin validiudesta koulujen osaamista mittaavina mittareina sekä siitä, miten yhteiskuntaopin oppiaineelle määritellyt hyvän osaamisen kriteerit arvosanalle 8 toimivat.

Tämän tutkimuksen tehtävänä ei ole kuitenkaan ollut selvittää sellaisten taustamuuttujien, kuten opettajien subjektiivisten ominaisuuksien tai opetuksessa hyödynnettävien eri oppimateriaalien vaikutusta koulujen välisiin eroihin yhteiskuntaopin ja historian keskimääräisissä arvosanoissa ja seuranta-arviointituloksissa. Ennen kuin yksilöllisiin efekteihin, kuten opettajaefektiin, päästään käsiksi, on tutkimusaineistosta perattava koulutason efekti, joka tässä tutkimuksessa on tehty. Opettajaefektin tutkiminen olisi sekä kiinnostavaa että hyödyllistä, mutta se edellyttäisi kokonaan uutta tutkimusta, eikä sitä ole edes yritetty sisällyttää tähän tutkimukseen.

Edellä esitetty ehdotus opettajien subjektiivisten ominaisuuksien tutkimisesta nostaa esiin myös muita arvosanan antamiseen liittyviä uusia tutkimusaiheita. Koska oppilaan arviointi on yhteydessä muun muassa oppimiskäsitykseen ja koulutusjärjestelmään, olisi hyvä tutkia sitä, miten ja millä perusteella eri oppiaineiden opettajat antavat päättöarvosanoja. Tutkimuksessa olisi hyvä tarkastella sitä, miten opettajat hyödyntävät Perusopetuksen opetussuunnitelman perusteita (2004) ja siinä eri oppiaineiden hyvän osaamisen kriteereitä antaessaan päättöarvosanoja, missä määrin opettajat painottavat kirjallisia tuotoksia, luokkatyöskentelyä, oppilaiden aktiivisuutta tai kotitehtävien tekemistä. Olisi tärkeää saada tutkimustietoa siitä, missä määrin opettajat mahdollisesti keskustelevat antamistaan arvosanoista kollegoidensa kanssa sekä siitä, miten päättöarvosanan antamisperusteita käydään läpi oppilaiden tai heidän huoltajiensa kanssa.

Vaikka tässä tutkimuksessa ei ole painotettu arvioinnin ja arvosanojen antamisen etiikkaa, olisi tarvetta tutkimukselle, joka pureutuisi opettajien arviointikäytäntöiden eettiseen puoleen. Kyseinen tutkimus täydentäisi koulujen arviointikäytäntöiden ja arviointikulttuuriin liittyvää tutkimustietoa myös siinä, mitkä ovat ne tilanteet tai asiat, jotka mahdollisesti saavat opettajan joko lieventämään tai kiristämään oppilaidensa arviointia ja korottamaan tai laskemaan oppilaiden arvosanoja.

Yhteiskuntaopin ja historian arvosanojen antamiskäytänteissä olisi kiinnostavaa ja hyödyllistä tutkia niin sanottua haloilmiön vaikutusta annettuihin arvosanoihin. Haloilmiöllä tarkoitetaan tässä yhteydessä sitä, että opettaja nostaa oppilaan arvosanaa siksi, että oppilas on tunnetusti hyvä muissa oppiaineissa tai esimerkiksi opettaja antaa heikon arvosanan oppilaalle yhteiskuntaopissa, koska oppilas on saanut heikon arvosanan jo aiemmin historiassa. Ajatus tutkimusaiheesta virisi tämän tutkimuksen empiriaosan taulukoista 9 ja 10. Taulukon 9 avulla on havaittavissa, että oppilaat, jotka olivat saaneet hyviä arvosanoja esimerkiksi yhteiskuntaopissa, olivat saaneet niitä myös historiassa, äidinkielessä ja matematiikassakin. Taulukko 10 sen sijaan kuvaa tämän tutkimuksen aineistolla tilannetta keväällä 2011, jossa yli 50 % oppilaista oli saanut saman arvosanan yhteiskuntaopissa kuin historiassakin. Tulos koski kaikkia annettuja yhteiskuntaopin arvosanoja 5:stä 10:een.

Opetushallituksen seuranta-arviointulokset esimerkiksi äidinkielessä ja matematiikassa ovat osoittaneet, että myös niissä näkyy samalla tavalla kuin yhteiskuntaopissa ja historiassa oppilaiden saamien arvosanojen ja kyseisten oppiaineiden seuranta-arvioinnissa osoitetun osaamisen välillä koulujen välisiä eroja. Mikäli käytettävissä olisi useamman eri arviointikerran arvosana- ja seuranta-arviointitietoja, olisi hyödyllistä tarkastella myös arvosanojen pysyvyyttä ja sitä kautta arvosanojen yhdenvertaisuutta ja vertailtavuutta suhteessa eri vuosina annettuihin arvosanoihin. Käytännössä vuosittain jatko-opintoihin ja työelämään pyrkii tuhansia nuoria, jotka ovat saaneet perusopetuksen päättöarvosanansa eri vuosina, mutta jotka valitaan olettaen, että annetut arvosanat ovat vuodesta toiseen yhteismitallisia.

8.3 Tutkimustulosten yhteiskunnallinen merkitys

Tällä tutkimuksella on tieteellisen merkityksen lisäksi koulutuspoliittista ja yhteiskunnallista merkitystä, jotka kiinnittyvät oppilaan arvioinnin ja siinä erityisesti päättöarvioinnin yhdenvertaisuuteen ja arvosanojen vertailukelpoisuuteen. Virta (1999, 30, 39, 94) ja Jakku-Sihvonen (2001, 85) ovat todenneet, että päättöarvosanan käyttäminen jatkokoulutukseen valikoitumisen perusteena asettaa päättöarvioinnille oppilaiden oikeusturvaan ja arvosanojen vertailukelpoisuuteen liittyviä haasteita.

Arvioinnin yhdenvertaisuus ei yhteiskunnallisena ja koulutuspoliittisena kysymyksenä ole uusi. Siihen kiinnitettiin suurta huomiota esimerkiksi 1960-luvun lopussa, kun peruskoulun periaatteita linjattiin Peruskoulun komiteanmietinnöissä (1970: 4A, 5A). Päättöarvioinnin hyvän osaamisen kriteereillä halutaan nykyisin turvata arvosanojen yhdenvertaiset antamisperusteet, koska niillä pyritään ja valikoidutaan jatko-opintoihin (Lankinen & Lahtinen 2009, 236, 240). Samalla tavalla jo 1970-luvun alussa korostettiin, että oppilaan arvosanan tuli perustua selkeisiin tavoitteisiin, joiden avulla oppilaan tietoja ja taitoja voitiin arvioida mahdollisimman yhteismitallisesti, koska arvosanoilla oli niin suuri merkitys oppilaan jatko-opintojen kannalta (Komiteanmietintö 1970: A4, 158).

Suomalainen koulutusjärjestelmä, jonka lähtökohtiin on koko peruskoulun voimassaoloajan kuulunut koulutuksellinen tasa-arvo ja yhdenvertaisuus, joutuu koeltavaksi, kun oppiaineesta annettujen arvosanojen ja seuranta-arvioinnissa osoitetun osaamisen välillä on suuria koulujen välisiä eroja. Kahden eri oppiaineen, yhteiskuntaopin ja historian, aineistolla tehty tutkimus osoittaa, että koulujen keskimääräisellä osaamistasolla on selkeä yhteys siihen, kuinka vaativaa tai lievää arvosanan antamiskulttuuria koulussa noudatetaan. Tämän tutkimuksen tulos ei tue ainakaan yhteiskuntaopin eikä myöskään historian oppiaineiden osalta käsitystä siitä, että Perusopetuksen opetussuunnitelman perusteissa (2004) määritellyt päättöarvioinnin kriteerit toimisivat sellaisena mittatikkuna, jolla päättöarvosanojen yhdenvertaiset antamisperusteet voitaisiin skaalata. Tutkimustuloksen todennäköisenä selityksenä on, että opetussuunnitelman perusteet eivät tarjoa riittävästi tukea yhteiskuntaopin ja historian absoluuttiseen ankkurointiin suhteessa hyvän osaamisen kriteereihin ja siksi opettajat käyttävät joko tietoisesti tai tiedostamattaan oman opetusryhmän tai oman koulun oppilaita ainakin osittain viiteryhmänä, johon yksittäisen oppilaan päättöarvosana suhteutetaan.

Päättöarvosanan valtakunnallista vertailukelpoisuutta, oppilaiden yhdenvertaisuutta heidän pyrkimään päättöarvosanoilla toisen asteen opintoihin ja tätä kautta oppilaiden oikeusturvaa tulisi lisätä seuraavin koulutuspoliittisin toimenpitein: 1) perusopetuksen opetussuunnitelmauudistuksen yhteydessä tulisi koko oppilaan arviointia koskeva osuus rakentaa uudelleen, 2) perusopetuksessa opetettaville oppiaineille tulisi rakentaa koepankki, jonka avulla koulutuksenjärjestäjillä, kouluilla ja opettajilla olisi mahdollisuus arvioida oman koulunsa osaamistasoa suhteessa valtakunnalliseen osaamistasoon, 3) eri oppiaineiden seuranta-arviointien yhteydessä tulisi nykyistä systemaattisemmin kerätä arviointitietoa sekä koulujen, koulutuksen järjestäjien että poliittisten päättäjien käyttöön siitä, onko oppilaiden saamien arvosanojen ja kyseisten oppiaineiden seuranta-arvioinnissa osoitetun osaamisen välillä sellaisia koulujen välisiä eroja, jotka vaarantavat oppilaan oikeusturvaa ja päättöarvioinnin yhdenvertaisuutta.

Seuraavan perusopetuksen opetussuunnitelmauudistuksen yhteydessä tulisi oppilaan arviointia tarkastella kokonaisuutena ja rakentaa se niin, että oppilaan arviointia ja arvosanan antamista ohjaavat normit tukisivat sekä opintojen aikaista arviointia että päättöarviointia. Opettajan arviointityötä ja arvosanan antamista tulisi opetussuunnitelman perusteissa ohjata selkeämmin niin, että asiakirjaan olisi määriteltä opintojen aikaisen arvioinnin ja siinä arvosanan antamisen sekä päättöarvioinnin ja siinä päättöarvosanan antamisen periaatteet, arviointiohjeet ja arviointikriteerit. Tämän lisäksi eri oppiaineiden kriteerit tulisi määritellä niiden luonteesta käsin, mutta samalla ottaen huomioon, että kriteerit olisivat enemmän yhteismittallisia. Jo voimassaolevia historian ja yhteiskuntaopin päättöarvioinnin kriteereitä vertailemalla on nähtävissä, että esimerkiksi yhteiskuntaopin päättöarvosanan kriteerit ovat historian päättöarviointikriteerejä ylimalkaisemmat ja ne eivät kiinnity riittävästi yhteiskuntaopissa asetettuihin tavoitteisiin, vaikka arvioinnin tulisi aina

lähteä asetetuista tavoitteista (Atjonen 2005, 144). Yhteiskuntaopin kriteereitä ei myöskään ole johdettu samalla lailla tiedon luonteesta kuin historian oppiaineen kriteereitä, koska yhteiskuntaopilla ei ole yhtä tiedonalataustaa ja siihen perustuvaa tiedon luonnetta (Löfström 2012a, 148).

Opetussuunnitelman arviointiosuutta uudistettaessa olisi hyvä pohtia perusteellisesti, tulisiko myös muille kouluarvosanoille kuin arvosanalle 8 laatia oppiainekohdattaiset kuvaukset, sillä yhteiskuntaopin ja historian osalta tutkimustulokset osoittavat, että arvosana 8 enempää kuin arvosana 5 niukkoine kuvauksineen ei toimi toivotun mukaisesti mittatikkuna muita arvosanoja annettaessa.

1990-luvun lopulla peruskoulun oppilasarvioinnin yhtenäistämiseksi oli kehitteillä koepankki, josta kouluilla suunniteltiin olevan mahdollisuus tilata eri oppiaineiden tehtäväsarjoja oppilaidensa osaamisen tason määrittämiseksi (Koulutuksen tuloksellisuuden arviointimalli 1998, 47). Uuden arviointiviraston aloittaessa toimintansa vuonna 2014 olisi hyödyllistä sen tehtäviä linjattaessa pohtia myös sitä, tulisiko sen tai jonkun toisen organisaation tehtäviin kuulua eri oppiaineiden tehtäväsarjojen ja arviointiohjeiden laatiminen. Tehtäväsarjojen avulla opettajat, koulut ja koulutuksenjärjestäjät voisivat arvioida omien oppilaidensa osaamistasoa suhteessa valtakunnalliseen osaamistasoon, jolloin esimerkiksi oppiaineen päättöarvosanaa annettaessa oman koulun oppilaat eivät muodostaisi ainoaa käytettävissä olevaa viiteryhmää, vaan opettajalla olisi mahdollisuus hyödyntää tietoa valtakunnallisesta osaamistasosta ja suhteuttaa omien oppilaidensa osaamistaso siihen päättöarvosanaa antaessaan. Koepankista saatavat eri oppiaineiden tehtävät antaisivat opettajille, kouluille ja koulutuksenjärjestäjille mahdollisuuden skaalata oppilaiden osaamistasoa ilman, että se asettaisi suomalaiselle koulutusjärjestelmälle paineita järjestää joissain oppiaineissa kansallisia päättökokeita, kuten esimerkiksi Ranskassa. Eri oppiaineiden tehtäväsarjojen avulla saatavaa tietoa opettajat, koulut ja koulutuksen järjestäjät voisivat hyödyntää sekä opetuksen kehittämisessä että oppilaan arvioinnin yhdenmukaistamisessa. Mahdollinen koepankki voisi toimia vapaaehtoisuuteen perustuen. Koepankin tehtäväsarjojen tulisi olla kouluille tai koulutuksenjärjestäjille vapaaehtoisia ja maksuttomia, jotta tehtävistä perittävät mahdolliset maksut eivät asettaisi kouluja eriarvoiseen asemaan.

Tutkimuksesta nousee esiin myös suositus tulevia kansallisia arviointeja suunniteltaessa. Eri oppiaineiden seuranta-arviointeihin tulisi liittää koulutuksellisen tasarvon kuten sukupuolten, eri alueiden ja kieliryhmien välistä tasa-arvoa mittaavien indikaattoreiden lisäksi oppilaiden osaamistasoa ja arvosanojen yhdenvertaisuutta mittaava indikaattori, joka tarkastelisi ennen kaikkea koulujen välisiä osaamistasoeroja suhteessa annettuihin arvosanoihin.

- Abrams, L. M. 2007. Implications of High Stakes testing for the Use of Formative Classroom Assessment. Teoksessa McMillan, J. Formative Classroom Achievement. New York: Teachers College Press, 79–98.
- Ahonen, S. 1996. Historiatietoisuus oppimistuloksena. Teoksessa Jakku-Sihvonen, R., Lindström, A., Lipsanen, S. (toim.). Toteuttaako peruskoulu tasa-arvoa? Opetushallituksen julkaisuja 1/96. Helsinki: Opetushallitus, 375–390.
- Ahonen, S. 1998. Historiaton sukupolvi. Historian vastaanotto ja historiallisen identiteetin rakentuminen 1990-luvun nuorison keskuudessa. Historiallisia tutkimuksia 202. Helsinki: Suomen Historiallinen Seura.
- American Psychological Association. 1966. Standards of educational and psychological tests. Washington DC: Author.
- Ansley, T. 1997. The Role of Standardized Achievement Tests in Grades K-12. Teoksessa Phye G. D. (toim.). Handbook of Classroom Assessment. Learning, Adjustment and Achievement. San Diego: Academic Press, 265–285.
- Apajalahti, M. 2004. Vuosiluokittaiset tavoitteet ja sisällöt. Teoksessa Vitikka, E. & Saloranta-Eriksson, O. (toim.). Uudistuva perusopetus. Näkökulmia opetuksen ja opetussuunnitelman kehittämiseen. Helsinki: Opetushallitus, 126–127.
- Arola, P. 2001. Yhteiskuntaopista 2000-luvun kansalaiskasvatukseen. Teoksessa Löfström J. (toim.). Yhteiskuntatiedon tiedetausta. Ainedidaktisia lähtökohtia oppiaineen kehittämiseen. Helsinki, 1–8.
- Arola, P. 2005. Aktiiviseksi kansalaiseksi. Kansalaisvaikuttaminen Viikin normaalikoulussa 2004–2005. Teoksessa Ojanperä, T. (toim.). Näkökulmia aktiiviseen kansalaisuuteen. Viikin Normaalikoulun julkaisuja 2, 31–41. Noudettu 30.8.2012 osoitteesta http://www.vink.helsinki.fi/files/aktiivinen_kansalainen.pdf
- Atjonen, P. 1993. Kunnan opetussuunnitelma koulun hallinnollisen ja pedagogisen kehittämisen kohteena ja välineenä. Acta Univ. Oul. E 11. Oulun yliopisto. Kajaanin opettajankoulutuslaitos.
- Atjonen, P. 2005. Arviointi opetuksen ja oppimisen tueksi. Teoksessa Lyytinen, H. K. & Räsänen, A. (toim.). Kehittämissuuntaa arvioinnista. Koulutuksen arviointineuvoston julkaisuja 6. Jyväskylä: Koulutuksen arviointineuvosto, 143–152.
- Atjonen, P. 2007. Hyvä, paha arviointi. Helsinki: Kustannusosakeyhtiö Tammi.
- Autio, T. 2002. Teaching Under Siege. Beyond the Traditional Curriculum Studies and/or Didaktik Split. Acta Universitatis Tamperensis 904. Tampere: University of Tampere.
- Autio, T. 2007. Curriculum-tutkimus ja opettajankoulutus. Aikuiskasvatus. Aikuiskasvatus-tieteellinen aikakauslehti 27 (1), 43–47.
- Bachman, L. F. 2005. Building and supporting a case for test use. Language Assessment Quarterly, 2(1), 1–34.
- Baxter, G. P. & Glaser, R. 1998. The Cognitive Complexity of Science Performance Assessment in the Science Classroom, Educational Measurement: Issues and Practice, 17 (3), 37–45.
- van den Berg, M. 2010. Nuorten näkemyksiä Suomen yhteiskunnallisesta kehityksestä. Helsinki: Historiallis-yhteiskuntatiedollisen kasvatuksen tutkimus- ja kehittämiskeskus.

- van den Berg, M. & Löfström, J. 2011. Katsaus peruskoulun yhteiskuntatiedolliseen opetukseen 2000-luvun Suomessa: puitteet, tarpeet, kehitysnäkymät. *Kasvatus ja aika* 5 (1) 2011, 79–86.
- Bracey, G. W. 1991. Why can't they be like we were? *Phi Delta Kappan*, 73, 104–117.
- Brennan, R. L. 2006. Perspectives on the Evolution and Future of Educational Measurement. Teoksessa Brennan R. L. *Educational Measurement* (4th edition). Westport, CT: American Council on Education and Praeger Publishers, 1–16.
- Broadfoot, P. 1995. Performance assessment in perspective: international trends and current English practice. Teoksessa Torrance, H. (toim.). *Evaluating authentic assessment. Problems and possibilities in new approaches to assessment*. Buckingham: Open University Press, 9–4.
- Broadfoot, P. 1996. *Education, assessment and society*. Buckingham: Open University Press.
- Brookhart, S. M. 1991. Grading practices and validity. *Educational Measurement: Issues and Practice*, 10(1), 35–36.
- Brookhart, S. M. 1994. *Teacher's Grading: Theory and Practice*. *Applied measurement in Education* 7(4), 279–301.
- Brookhart, S. M. 2003. Developing measurement theory for classroom assessment purposes and uses. *Educational Measurement: Issues and Practice*, 22(4), 5–12.
- Brookhart, S. 2007. Expanding Views About Formative Classroom Assessment: A Review of the Literature. Teoksessa McMillan, J. *Formative Classroom Achievement*. New York: Teachers College Press, 43–62.
- Brunell, V. 2002. CIVED Suomen näkökulmasta. Teoksessa Brunell, V. & Törmäkangas, K. (toim.). *Tulevaisuuden yhteiskunnan rakentajat. Yhteiskunnallisen opetuksen kansainvälinen tutkimus Suomen näkökulmasta*. Jyväskylä: Jyväskylän yliopisto. Koulutuksen tutkimuslaitos, 5–11.
- Bulterman-Bos, J., Terwel, J., Verloop, N., & Wardekker, W. 2002. Observation in teaching: Toward a practice of objectivity. *Teachers College record*, 104, 1069–1100.
- Calfee, R.C. 1987. The School as a context for assessment of literacy. *The Reading Teacher*, 40, 738–743.
- Calfee, R.C. 1995. Implications of cognitive psychology for authentic assessment and instruction. Teoksessa Oakland, T. & Hambleton, R. (toim.). *International perspective on academic assessment*. Boston: Kluwer Academic Publishers, 254–258.
- Calfee, R. C. & Masuda, W. V. 1997. Classroom Assessment as Inquiry. Teoksessa Phye, G. D. (toim.). *Handbook of Classroom Assessment. Learning, Adjustment and Achievement*. San Diego: Academic Press, 69–102.
- Campbell, D. T. & Fiske, D. W. 1959. Convergent and discriminant validity by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81–105.
- Cizek, G. J. 1996. Setting passing scores. *Educational Measurement: Issues and Practice*, 15 (2), 20–31.
- Cizek, G. J. 1997. Learning, Achievement and Assessment: Constructs at a crossroads. Teoksessa Phye G. D. (toim.). *Handbook of Classroom Assessment. Learning, Adjustment and Achievement*. San Diego: Academic Press, 2–29.
- Cizek, G. C., Fitzgerald, S. M. & Rachor, R. E. 1995/1996. Teachers' assessment practices: Preparation, isolation, and the kitchen sink. *Educational Assessment* 3(2), 159–179.

- Coleman, J. S., Campbell, E.Q., Hobson, C. J., McPartland, J., Mood, A.M., Weinfield, D. F., York, R. L. 1966. Equality of Educational Opportunity. Washington: U.S. Department of Health, Education, and Welfare.
- Cook, T. D. & Campbell, D. T. 1979. Quasi-experimentation: Design and analysis issues for field settings. New York: Basic Books.
- Cresswell, M. 2000. The Role of Public Examinations in Defining and Monitoring Standards. Teoksessa Goldstein, H. & Heath, A. Educational Standards. Proceedings of the British Academy 102. Oxford University Press, 69–120.
- Cronbach, L. J. 1971. Test validation. Teoksessa Thorndike, R.L. (toim.). Educational measurement (2nd Ed.) Washington DC: American Council on Education and Praeger Publishers, 443–507.
- Cronbach, L. J. 1980. Selection theory for a political world. Public Personnel Management, 9 (1), 37–50.
- Cronbach, L.J. 1988. Five perspectives on validation argument. Teoksessa Wainer, H. & Braun, H. (toim.). Test validity. Hillsdale, New Jersey: Erlbaum, 34–35.
- Cronbach, L.J. 1989. Construct validation after thirty years. Teoksessa Linn, R.E. (toim.). Intelligence: Measurement, theory, and public policy. Urbana: University of Illinois Press, 147-171.
- Cronbach, L. J. & Meehl, P. E. 1955. Construct validity in psychological tests. Psychological Bulletin, 52, 281–302.
- Cronbach, L. J. & Snow, R. E. 1977. Aptitudes and instructional methods; A handbook for research on interaction. New York: Irvington Publishers.
- Cross, L. H. & Frary, R. B. 1999. Hodgebodge Grading: Endorsed by Students and Teachers alike. Applied measurement in Education, 12(1), 53–72.
- Cureton, E. E. 1951. Validity. Teoksessa Lindquist, E. F. (toim.). Educational measurement. 3rd.ed. Washington DC: American Council of Education, 621–694.
- Dillon, S. 2006. Schools Cut Back Subjects to Push Reading and Math. The New York Times. Noudettu 29.8.2012 osoitteesta http://www.nytimes.com/2006/03/26/education/26child.html?_r=2
- Dunnette, M. D. & Borman, W. C. 1979. Personnel selection and classification systems. Teoksessa Rosenzweig, M. R. & Porter, L. W. (toim.). Annual Review of Psychology (Vol. 30). Palo Alto, California: Annual Reviews, 1979.
- Ebel, R. 1961. Must all tests be valid? American Psychologists, 16, 640–647.
- ECTS (European Credit Transfer and Accumulation System). 2009. Users' Guide. Belgium. Noudettu 31.8.2012 osoitteesta http://ec.europa.eu/education/lifelong-learning-policy/doc/ects/guide_en.pdf
- Eduscol. Portail national des professionnels de l'éducation. Noudettu 31.8.2012 osoitteesta <http://eduscol.education.fr/cid46835/organisation-et-modalites-d-attribution.html>
- Embertson, S. E. 2007. Construct Validity: A Universal Validity System or Just Another Test Evaluation Procedure? Educational Research, 36, 449–455.
- Eränpalo, T. 2012. Onko kuva suomalaisnuorten yhteiskunnallisesta passivisuudesta harhaa? Kasvatus & Aika 6 (1) 2012, 23–38.
- European Commission. European Credit Transfer and Accumulation System (ECTS). Noudettu 31.8.2012 osoitteesta http://ec.europa.eu/education/lifelong-learning-policy/ects_en.htm

- Eurypedia. European Encyclopedia on National Education Systems. Noudettu 31.8.2012 osoitteesta https://webgate.ec.europa.eu/fpfis/mwikis/eurydice/index.php/France_fr:Aper%C3%A7u_des_principaux_%C3%A9l%C3%A9ments
- Evans, R. W. 2004. *Social Studies Wars. What Should We Teach the Children?* New York: Teachers College Press.
- Ferrara, S. & DeMauro, G. 2006. *Standardized Assessment of Individual Achievement in K-12.* Teoksessa Brennan R. L. *Educational Measurement* (4th edition). Westport, CT: American Council on Education and Praeger Publishers, 579–621.
- Glaser, R. & Klaus, D. J. 1962. *Proficiency measurement: Assessing human performance.* Teoksessa Gagné, R. (toim.). *Psychological principles in system development.* New York: Holt, Rinehart & Winston, 421–427.
- Glaser, R., Lesgold, A., & Lajoie, S. 1989. *Toward a cognitive theory for the measurement of achievement.* Teoksessa Ronning, R., Glover, Conoley, J. C., & Witt, J. (toim.). *The influence of cognitive psychology on testing and measurement*, 96–131. Hillsdale, NJ: Erlbaum.
- Goals 2000: Educate America Act. Noudettu 31.8.2012 osoitteesta <http://www2.ed.gov/legislation/GOALS2000/TheAct/index.html>
- Gordon, M. E. & Fay C. H. 2010. *The Effects of Grading and Teaching Practices on Students' Perceptions of Grading Fairness.* *College Teaching*, 58, 93–98.
- Gordon, T., Lahelma, E., Lyytinen, H., Niemivirta, M., Scheinin, P., Siimes. M. 2004. *Koulu – sukupuoli – oppimistulokset.* Helsinki: Opetushallitus.
- Gorin, J. S. 2006. *Test Design with Cognition in Mind.* *Educational measurement: Issues and Practice*, 25, 351–373.
- Gorin, J. S. 2007. *Reconsidering Issues in Validity Theory.* *Educational Research*, 36, 456–462.
- Grant, S. G. 2010. *High-Stakes Testing: How Are Social Studies Teachers Responding?* Teoksessa Parker, W.C. (toim.). *Social Studies Today.* New York: Routledge, 43–52.
- Grant, S. G. & Salinas, C. 2008. *Assessment and Accountability in the Social Studies.* Teoksessa Levstick, L. S. & Tyson, C. A. (toim.). *Handbook of Research in Social Studies Education.* New York: Routledge Falmer Taylor & Francis Group. S., 219–267.
- Green, K. H. & Emerson, A. 2007. *A New framework for grading.* *Assessment and Evaluation in Higher Education* 32, 496–511.
- Greer, S. 1969. *The Logic of Social Inquiry.* Chicago: Aldine.
- Guion, R. M. 1978. *Scoring of content domain samples: The problem of fairness,* *Journal of Applied Psychology*, 63, 499–506.
- Guion, R.M. 1980. *On trinitarian conception of validity.* *Professional Psychology*, 11, 358–398.
- Gullickson, A. R. 1985. *Student evaluation techniques and their relationship to grade and curriculum.* *Journal of Educational Research*, 72 (2), 96–100.
- Guskey, T. R. 2011. *Stability and Change in High School Grades.* *National Association of Secondary School Principals bulletin*, 95 (2), 85–98.
- Guskey, T. R. & Bailey, J. M. 2001. *Developing Grading and Reporting Systems for Student Learning.* California: Corwin Press.

- Hambleton, R. K. & Pitoniak, M. J. 2006. Setting Performance Standards. Teoksessa Brennan R. L. 2006. Educational Measurement (4th edition). Westport, CT: American Council on Education and Praeger Publishers, 433–470.
- Hamilton, L. S., Nussbaum, E. M., & Snow, R. E. 1997. Interview Procedures for Validating Science Assessments. *Applied Measurement in Education*, 10, 181–200.
- Hannén, K. 2001. Utvärdering av inlärningsresultat i modersmål i åk 6 i den grundläggande utbildningen våren 2000. Utvärdering av inlärningsresultat 4/2001. Helsingfors: Utbildningsstyrelsen.
- Hautamäki, J., Arinen, P., Hautamäki, A., Ikonen-Varila, M., Kupiainen, S., Lindblom, B., Niemivirta, M., Rantanen, P., Ruuth, M. & Scheinin, P. 2000. Oppimaan oppiminen yläasteella. Oppimistulosten arviointi 7/2000. Helsinki: Opetushallitus.
- Hautamäki, J. & Kuusela, J. 2005. Mitataan oppilaita, mutta päätellään koulusta. Teoksessa Lyytinen, H. K. & Räisänen, A. (toim.). 2005. Kehittämissuuntaa arvioinnista. Jyväskylä: Koulutuksen arviointineuvosto, 229–241.
- Hautamäki, J. 2010. Minne vaihtelu menee? Oppilaiden, luokkien ja koulujen väliset erot oppimistuloksissa. Teoksessa Rimpelä, M. & Bernelius, V. (toim.). 2010. Peruskoulujen oppimistulokset ja oppilaiden hyvinvointi eriytyvällä Helsingin seudulla. MetrOP-tutkimus 2010–2013. Mitä tiedettiin tutkimuksen käynnistyessä keväällä 2010? Geotieteiden ja maantieteen laitoksen julkaisuja B1. Helsinki: University of Helsinki, 49–54. Noudettu 29.11.2012 osoitteesta https://helda.helsinki.fi/bitstream/handle/10138/17076/MetrOP-raportti_1_verkkoversio.pdf?sequence=1
- Heubert, J. P. (toim.) & Hauser, R. M. 1999. National Research Council. High Stakes: Testing for Tracking, Promotion and Graduation. Washington D.C: National Academic Press. Noudettu 29.8.2012 osoitteesta http://www.nap.edu/openbook.php?record_id=6336&page=1
- Heinonen, J.-P. 2005. Opetussuunnitelmat vai oppimateriaalit. Peruskoulun opettajien käsitteitä opetussuunnitelmien ja oppimateriaalien merkityksestä opetuksessa. Soveltavan kasvatustieteen laitos. Tutkimuksia 257. Helsinki: Helsingin yliopisto.
- Heinonen, S. 2001a. Arvioinnin teoreettisia lähtökohtia. Teoksessa Jakku-Sihvonen, R. & Heinonen S. 2001. Johdatus koulutuksen uudistuvaan arviointikulttuuriin. Arviointi 2/2001. Helsinki: Opetushallitus, 21–46.
- Heinonen, S. 2001b. Arviointi koulutuksen eri tasoilla. Teoksessa Jakku-Sihvonen, R. & Heinonen S. 2001. Johdatus koulutuksen uudistuvaan arviointikulttuuriin. Arviointi 2/2001. Helsinki: Opetushallitus, 48–63.
- Heinonen, S. & Jakku-Sihvonen, R. 2001. Koulutuksen arvioinnin etiikka. Teoksessa Jakku-Sihvonen, R. & Heinonen S. Johdatus koulutuksen uudistuvaan arviointikulttuuriin. Arviointi 2/2001. Helsinki: Opetushallitus, 95–110.
- Helakorpi, S. & Olkinuora, H. 1997. Asiantuntijuutta oppimassa. Porvoo: WSOY.
- Hellgren, J. 2011. Modersmål och litteratur i årskurs 9. En utvärdering av inlärningsresultat i modersmål och litteratur i årskurs 9 våren 2010. Uppföljningsrapporter 1011: 1. Helsingfors: Utbildningstyrelsen.
- Hildén, R. 2008. Validiusargumentin rakentuminen suullisen kielitaidon arviointiprojektissa. Teoksessa Kallioniemi, A. (toim.). Uudistuva ja kehittyvä ainedidaktiikka. Ainedidaktinen symposiumi 8.2.2008 Helsingissä. Osa 2., Soveltavan kasvatustieteen laitos. Tutkimuksia, 299. Helsinki: Helsingin yliopisto, 751–762.

- Hills, J. R. 1991. Apathy concerning grading and testing. *Phi Delta Kappan*, 72, 540–545.
- Hirsjärvi, S., Remes, P. & Sajavaara, P. 2009. *Tutki ja kirjoita*. Helsinki: Kustannusosakeyhtiö Tammi.
- Hirvonen, K. 2012. Onko laskutaito laskussa? *Matematiikan oppimistulokset perusopetuksen päättövaiheessa 2012*. Koulutuksen seurantaraportit 2012: 4. Helsinki: Opetushallitus.
- Huisman, T. 2004. *Liikunnan arviointi peruskoulussa 2003. Yhdeksäsluokkalaisten kunto, liikunta-aktiivisuus ja koululiikuntaan asennoituminen*. Oppimistulosten arviointi 1/2004. Helsinki: Opetushallitus.
- Impara, J.C., Plake, B.S. & Fager, J. J. 1993. Educational administrators' and teachers' knowledge of classroom assessment. *Journal of School Leadership*, 3, 510–521.
- IEA-Civics. Nuori kansalainen. Yhteiskunnallisen opetuksen kansainvälinen tutkimus. Koulutuksen tutkimuslaitos, Jyväskylän yliopisto. Noudettu 1.9.2012 osoitteesta <http://ktl.jyu.fi/arkisto/civics/civics.htm>
- Jakku-Sihvonen, R. 1996. Opetuksen saavutettavuus ja koulutuksellinen tasa-arvo arvioinnin kohteena. Teoksessa Jakku-Sihvonen, R., Lindström, A. & Lipsanen, S. (toim.). *Toteuttaako peruskoulu tasa-arvoa? Oppimistulosten arviointeja 1/96*. Helsinki: Opetushallitus, 2–6.
- Jakku-Sihvonen, R. 2001. Arviointitiedon luotettavuuden osoittaminen. Teoksessa Jakku-Sihvonen, R. & Heinonen, S. *Johdatus koulutuksen uudistuvaan arviointikulttuuriin*. Arviointi 2/2001. Helsinki: Opetushallitus, 111–135.
- Jakku-Sihvonen, R. 2002a. Perusopetusta vuosina 1995–2002 koskeneiden arviointien tuloksia. Teoksessa Jakku-Sihvonen, R. & Kuusela, J. 2002. *Mahdollisuuksien koulutuspolitiikan tasa-arvo*. Arviointi/2001. Helsinki: Opetushallitus, 9–26.
- Jakku-Sihvonen, R. 2002b. Perusopetuksen oppimistuloksissa ja opiskeluun asennoitumisessa koulukohtaista vaihtelua. Teoksessa Jakku-Sihvonen, R. & Kuusela, J. 2002. *Mahdollisuuksien koulutuspolitiikan tasa-arvo*. Arviointi/2001. Helsinki: Opetushallitus, 27–66.
- Jakku-Sihvonen, R. 2010. Oppimistulosten arviointitiedon puntarointia. *Hallinnon Tutkimus* 29 (4), 317–324.
- Jakku-Sihvonen, R. 2011. Oppimistuloksia arvioidaan koulutuksen kehittämisen perustaksi. Teoksessa Laitinen, S. & Hilmola, A. (toim.). *Taito- ja taideaineiden oppimistulokset asiantuntijoiden arviointia*. Raportit ja selvitykset 2011:11. Helsinki: Opetushallitus.
- Jakku-Sihvonen, R. 2012. Peruskoulusta perusopetuksiksi. Teoksessa Jakku-Sihvonen, R. & Kuusela, J. 2012. *Perusopetuksen aika*. Selvitys koulujen toimintaympäristöä kuvaavista indikaattoreista. Opetus- ja kulttuuriministeriön työryhmämuistioita ja selvityksiä 2012:13. Noudettu 30.7.2012 osoitteesta <http://www.minedu.fi/julkaisut>
- Jakku-Sihvonen, R. 2013. Oppimistulosten arviointijärjestelmistä ja niiden kehittämishaasteista. Teoksessa Räisänen, A. (toim.). *Oppimisen arviointin kontekstit ja käytännöt*. Raportit ja selvitykset 2013:3. Helsinki: Opetushallitus, 13–36. Noudettu 31.7.2013 osoitteesta http://www.oph.fi/download/149650_Oppimisen_arvioinnin_kontekstit_ja_kaytannot_2.pdf
- Jakku-Sihvonen, R. & Heinonen, S. 2001. Opiskelija-arviointi. Teoksessa Jakku-Sihvonen, R. & Heinonen, S. *Johdatus koulutuksen uudistuvaan arviointikulttuuriin*. Arviointi 2/2001. Helsinki: Opetushallitus, 78–94.

- Jakku-Sihvonen & Kuusela 2002. Mahdollisuuksien koulutuspolitiikan tasa-arvo. Arviointi 7/2002. Helsinki: Opetushallitus.
- Jakku-Sihvonen, R. & Kuusela, J. 2012a. Perusopetuksen aika. Selvitys koulujen toimintaympäristöä kuvaavista indikaattoreista. Opetus- ja kulttuuriministeriön työryhmämuistioita ja selvityksiä 2012:13. Noudettu 30.7.2012 osoitteesta <http://www.minedu.fi/julkaisut>
- Jakku-Sihvonen, R. & Kuusela, J. 2012b. Ehdotukset eriytymisen estotoimenpiteiksi. Teoksessa Jakku-Sihvonen, R ja Kuusela, J. 2012. Perusopetuksen aika. Selvitys koulujen toimintaympäristöä kuvaavista indikaattoreista. Opetus- ja kulttuuriministeriön työryhmämuistioita ja selvityksiä 2012:13, 55–87. Noudettu 30.7.2012 osoitteesta <http://www.minedu.fi/julkaisut>
- Jakku-Sihvonen, R. & Komulainen, E. 2004. Perusopetuksen oppimistulosten meta-arviointi. Arviointi 1/2004. Helsinki: Opetushallitus.
- Jakku-Sihvonen, R., Lindström, A. ja Lipsanen, S. (toim.). 1996. Toteuttaako peruskoulu tasa-arvoa? Opetushallituksen julkaisuja 1/96. Helsinki: Opetushallitus.
- Janson, T. 2011. Vad kommer på provet? Gymnasielärares provpraxis i samhällskunskap. Karlstad University Studies 2011: 2. Karlstad: Karlstad University Press.
- Kane, M. T. 1994. Validating the performance standards associated with passing scores. *Review of Educational Research*, 64, 425–461.
- Kane, M. T. 1998. Choosing between examinee-centered and test-centered standard-setting methods. *Educational Assessment*, 5, 129–145.
- Kane, M. T. 2001. Current Concerns in Validity Theory. *Journal of Educational measurement*, 38, 319–342.
- Kane, M. T. 2004. Certification testing as an illustration of argument based Validation. *Measurement*, 2(3), 135–170.
- Kane, M. T. 2006. Validation. Teoksessa Brennan, R. L. (toim.). *Educational Measurement* (4th edition). Westport, CT: American Council on Education and Praeger Publishers, 17–64.
- Kansallinen oppimistulosten arviointimalli. 1998. Arviointi 4/98. Helsinki: Opetushallitus.
- Kansanen P. 2004. Opetuksen käsitemaailma. Jyväskylä: PS-kustannus.
- Kansanen, P. 1998. Opetusprosessin arviointi. Teoksessa Jakku-Sihvonen, R. (toim.). *Onnistuuko oppiminen – oppimistuloksien ja opetuksen laadun arviointiperusteita peruskoulussa ja lukiossa*. Arviointi 3/1997. Helsinki: Opetushallitus, 429–448.
- Kaplan, A. 1964. *The conduct of inquiry: Methodology for behavioral science*. San Francisco: Chandler.
- Kartovaara, E. 2004. Perusopetuksen yleiset valtakunnalliset tavoitteet. Teoksessa Vitikka, E. & Saloranta-Eriksson, O. (toim.). *Uudistuva perusopetus. Näkökulmia opetuksen ja opetussuunnitelman kehittämiseen*. Helsinki: Opetushallitus, 66–68.
- Ketokivi, M. 2009. Tilastollinen päättely ja tieteellinen argumentointi. Helsinki: Gaudeamus. Helsinki University Press.
- Klapp Lehkholm, A. 2008. Grades and grade assignment: effects of student and school characteristics. Göteborg: Göteborgs Universitetet. Noudettu 29.11.2012 osoitteesta: <https://gupea.ub.gu.se/handle/2077/18673>

- Kolen, M. J. 2006. Scoring and Norming. Teoksessa Brennan R. L. 2006. Educational Measurement (4th edition). Westport, CT: American Council on Education and Praeger Publishers, 155–186.
- Komiteanmietintö 1970: A4. Peruskoulun opetussuunnitelmakomitean mietintö I. Opetussuunnitelman perusteet. Helsinki: Kouluhallitus.
- Komiteanmietintö 1970: A5. Peruskoulun opetussuunnitelmakomitean mietintö II. Oppiaineiden opetussuunnitelmat. Helsinki: Kouluhallitus.
- Konttinen, E., Suortti, J. & Väyrynen, K. 1980. Hegeliläinen perinne ja opetussuunnitelma. Jyväskylän yliopisto. Kasvatustieteiden tutkimuslaitos. Selosteita ja tiedotteita 141/1980.
- Korhonen, H. 1999. Peruskoulun matematiikan oppimistulosten kansallinen arviointi 1998. Oppimistulosten arviointi 1/1999. Helsinki: Opetushallitus.
- Korhonen, H. 2001. Perusopetuksen päättövaiheen matematiikan oppimistulosten kansallinen arviointi 2000. Oppimistulosten arviointi 3/2001. Helsinki: Opetushallitus.
- Koppinen, M-L., Korpinen, E. & Pollari, J. 1994. Mitä arviointi on? Juva: WSOY.
- Koretz, D. M. & Hamilton, L. S. 2006. Testing for Accountability in K-12. Teoksessa Brennan R.L. 2006. Educational Measurement (4th edition). Westport, CT: American Council on Education and Praeger Publishers, 531–578.
- Korkeakoski, E. 1997. Opettajan tavoitetietoisuus opetuksessa ja oppimisessa. Opetustapah-tuman eräiden piirteiden sekä oppimisen ja opiskelun mielekkyyden tarkastelua opetta-jan tavoitetietoisuuden näkökulmasta peruskoulun ala-asteella. Kasvatustieteen väitös-kirja. Acta Universitatis Tamperensis A. 533. Tampere: Tampereen yliopisto.
- Korkeakoski, E. 2001. Perusopetuksen äidinkielen oppimistulosten kansallinen arviointi 6. vuosiluokalla keväällä 2000. Oppimistulosten arviointi 1/2001. Helsinki: Opetushalli-tus.
- Korkeakoski, E. 2005. Perusopetuksen oppimistulokset ohjausjärjestelmän osaksi. Teoksessa Lyytinen, H. K. & Räisänen, A. (toim.). 2005. Kehittämissuuntaa arvioinnista. Jyväskylä: Koulutuksen arviointineuvosto, 175–192.
- Korpinen, E. 1998. Peruskoulun oppilasarvostelua kehittämässä. Teoksessa Törmä, E. (toim.). Tutkiva opettaja. Journal of Teacher researcher 3/1998. Jyväskylä: TUOPE, 8–33.
- Kosunen, T. 2002. Opettaja opetussuunnitelman ja oman työnsä kehittäjänä. Teoksessa Huusko, J. & Pietarinen, J. (toim.). Yhä parempi paikka kasvaa ja oppia ☒ punnittua puhetta koulun kehittämisestä. Kasvatustieteiden tiedekunnan selosteita 83. Joensuun yliopisto.
- Kosunen T. & Huusko J. 2002. Opetussuunnitelma opettajan työn ja kouluyhteisön kehit-tämisen välineenä. Teoksessa Julkunen, M-L. (toim.) 2002. Opetus, oppiminen, vuoro-vaikutus. Porvoo: WSOY, 202–226.
- Koulutuksen tuloksellisuuden arviointimalli 1998. Arviointi 7/98. Helsinki: Opetushallitus.
- Kuusela J. 2002. Oppimistulosten yhteydet demografisiin tekijöihin. Teoksessa Jakku-Sihvon-nen, R. & Kuusela, J. Mahdollisuuksien koulutuspolitiikan tasa-arvo. Arviointi 7/2002. Helsinki: Opetushallitus, 67–87.
- Kuusela, J. 2003. Koulujen paremmuusjärjestyksestä. Helsinki: Opetushallitus. Noudettu 30.8.2012 osoitteesta http://www.oph.fi/download/48971_koulujen_paremmuusjar-jestyksesta.pdf

- Kuusela, J. 2006. Temaattisia näkökulmia perusopetuksen tasa-arvoon. Oppimistulosten arviointi 6/2006. Helsinki: Opetushallitus. Noudettu 29.11.2012 osoitteesta: http://www.oph.fi/download/47156_Temaattinen_tasa_arvo.pdf
- Kuusela, J. 2010. Oppilaiden sosioekonomisen taustan yhteys koulumenestykseen koulutasolla. Teoksessa Rimpelä, M. & Bernelius, V. (toim.). Peruskoulujen oppimistulokset ja oppilaiden hyvinvointi eriytyvällä Helsingin seudulla. MetrOP-tutkimus 2010–2013. Mitä tiedettiin tutkimuksen käynnistyessä keväällä 2010? Geotieteiden ja maantieteen laitoksen julkaisuja B1. Helsinki: University of Helsinki, 49–54. Noudettu 29.11.2012 osoitteesta https://helda.helsinki.fi/bitstream/handle/10138/17076/MetrOP-raportti_1_verkkoversio.pdf?sequence=1
- Kuusela, J. 2012. Eriytymiskehitys oppimistulosten valossa. Teoksessa Jakku-Sihvonen, R. & Kuusela, J. 2012. Perusopetuksen aika. Selvitys koulujen toimintaympäristöä kuvaavista indikaattoreista. Opetus- ja kulttuuriministeriön työryhmämuistioita ja selvityksiä 2012:13. Noudettu 30.7.2012 osoitteesta <http://www.minedu.fi/julkaisut>
- Kuusela, P. 2001. Julkisen sektorin modernisaatio, tuloksellisuus ja arviointi. Kuopion yliopiston selvityksiä E. Yhteiskuntatieteet 30.
- Kuusela, J., Etelälahti, A., Hagman, Å., Hievanen, R., Karppinen, K., Nissilä, L., Rönneberg, U. & Siniharju, M. 2008. Maahanmuuttajaoppilaat ja koulutus – tutkimus oppimistuloksista, koulutusvalinnoista ja työllistymisestä. Helsinki: Opetushallitus.
- Kuusinen, J. 1986. Koulumenestys, lahjakkuus ja sosiaalinen tausta. Kasvatus, 17, 192–197.
- Kvale, S. 1995. Evaluation as construction of knowledge. Teoksessa Jussila, J. & Rajala, R. (toim.). Rajanylityksiä: Monipuolistuva kasvatustutkimus. XVI kasvatustieteen päivät Rovaniemellä 24.–26.11.1994. Lapin yliopiston kasvatustieteellisiä julkaisuja C 10.
- Kärnä, P., Hakonen, R. & Kuusela, J. 2012. Luonnontieteellinen osaaminen perusopetuksen 9.luokalla 2011. Koulutuksen seurantaraportit 2012:2. Helsinki: Opetushallitus.
- Lahdes, E. 1997. Peruskoulun uusi didaktiikka. Keuruu: Otava.
- Laitinen, S., Hilmola, A. & Juntunen, M-L. 2011. Perusopetuksen musiikin, kuvataiteen ja käsityön oppimistulosten arviointi 9.vuosiluokalla. Koulutuksen seurantaraportit 2011:1. Helsinki: Opetushallitus.
- Lankinen, T. 2007. Opetushallituksen asema, rooli ja tehtävät sekä koulutustoimialan ohjaus muuttuvassa toimintaympäristössä. Opetusministeriön työryhmämuistioita ja selvityksiä 2007: 46. Noudettu 28.8.2012 osoitteesta <http://www.minedu.fi/export/sites/default/OPM/Julkaisut/2007/liitteet/tr46.pdf?lang=fi>
- Lankinen, T. & Lahtinen, M. 2009. Koulutuksen lainsäädäntö käytännössä. Porvoo: WS Bookwell.
- Lappalainen, H.-P. 2000. Peruskoulun äidinkielen oppimistulosten kansallinen arviointi 9. vuosiluokalla 1999. Oppimistulosten arviointi 1/2000. Helsinki: Opetushallitus.
- Lappalainen, H.-P. 2001. Peruskoulun äidinkielen ja kirjallisuuden oppimistulosten kansallinen arviointi 9. vuosiluokalla 2001. Oppimistulosten arviointi 6/2001. Helsinki: Opetushallitus.
- Lappalainen, H.-P. 2004. Kerroin kaiken tietämäni. Perusopetuksen äidinkielen ja kirjallisuuden oppimistulosten arviointi 9. vuosiluokalla 2003. Oppimistulosten arviointi 2/2004. Helsinki: Opetushallitus.

- Lappalainen, H.-P. 2006. Ei taito taakkana ole. Perusopetuksen äidinkielen ja kirjallisuuden oppimistulosten arviointi 9. vuosiluokalla 2005. Oppimistulosten arviointi 1/2006. Helsinki: Opetushallitus.
- Lappalainen, H.-P. 2011. Sen edestään löytää. Äidinkielen ja kirjallisuuden oppimistulokset perusopetuksen päättövaiheessa 2010. Opetushallitus. Koulutuksen seurantaraportti 2011:2. Helsinki: Opetushallitus.
- Lawton, D. 1975. *Class, Culture and the Curriculum*. London: Routledge & Kegan Paul.
- Lehtisalo, L. & Raivola, R. 1999. *Koulutus ja koulutuspolitiikka 2000-luvulle*. Juva: WSOY.
- Lilja, K. 2002. Matematiikan oppimistuloksiin yhteydessä olevat tekijät peruskoulussa. Helsinki: Opetushallitus.
- Linnakylä, P. 2002. Kansainvälisten ja kansallisten oppimistulosten arviointien välisestä suhteesta. Teoksessa Olkinuora E., Jakku-Sihvonen, R. & Mattila E. (toim.). *Koulutuksen arviointi – lähtökohtia, malleja ja tilannekatsauksia*. Turun yliopiston kasvatustieteiden tiedekunnan julkaisuja B:70. Turku: Turun yliopiston kasvatustieteiden tiedekunta, 31–53. Noudettu 29.8.2012 osoitteesta <http://www.edu.utu.fi/erillisyksikot/koulutuksenarviointi/julkaisut/arviointi.pdf>
- Linnakylä, P. & Välijärvi, J. 2005. Arvon mekin ansaitsemme. Kansainvälinen arviointi suomalaisen koulun kehittämiseksi. Jyväskylä: PS-kustannus.
- Lissitz, R. W. & Samuelsen, K. 2007a. A Suggested Change in Terminology and Emphasis Regarding Validity and Education. *Educational Research*, 36, 437–448.
- Lissitz, R. W. & Samuelsen, K. 2007b. Further Clarification Regarding Validity and Education. *Educational Research*, 36, 482–484.
- Llosa, L. 2008. Building and Supporting Validity Argument for a Standards-Based Classroom Assessment of English Proficiency Based on Teacher Judgments. *Educational Measurement: Issues and Practice*, Fall 2008, 32–42.
- Loyd, B. H. & Loyd, D.E. 1997. Kindergarten through Grade 12 Standards: A Philosophy of Grading. Teoksessa Phye G.D. *Handbook of Classroom Assessment*. Learning, Adjustment and Achievement. San Diego: Academic Press, 481–489.
- Löfström, J. 2001a. Yhteiskuntatiedon ongelma. Teoksessa Löfström J. (toim.). *Yhteiskuntatiedon tiedetausta. Ainedidaktisia lähtökohtia oppiaineen kehittämiseen*. Helsinki: Helsingin yliopisto, 9–19.
- Löfström, J. 2001b. Yhteiskuntatiedon kehittämissuunta: tiedetaustan selkiyttäminen ja teoreettisuuden lisääminen. Teoksessa Löfström J. (toim.). *Yhteiskuntatiedon tiedetausta. Ainedidaktisia lähtökohtia oppiaineen kehittämiseen*. Helsinki: Helsingin yliopisto, 20–25.
- Löfström, J. 2002. Yhteiskuntatieto ja yhteiskunnallinen kasvatustieteiden periaatekysymyksiä ja suuntaviivoja. Teoksessa Löfström, J. (toim.). *Kohti tulevaa menneisyyttä. Historiallis-yhteiskunnallinen kasvatustieteiden uudella vuosituohannella*. Jyväskylä: PS-kustannus, 89–131.
- Löfström, J. 2008. Uudistuva yhteiskuntaopin didaktiikka – tavoitteet, sisällöt ja toteutus. Teoksessa Kallioniemi, A. (toim.). *Uudistuva ja kehittyvä ainedidaktiikka. Osa 1. Ainedidaktinen symposiumi 8.2.2008 Helsingissä*. Helsingin yliopiston tutkimuksia 298. Helsinki: Helsingin yliopisto, 278–289.

- Löfström, J. 2012a. Miten taitoja opetetaan yhteiskuntaopissa? Teoksessa Ouakrim-Soivio, N. & Kuusela, J. 2012. *Historian ja yhteiskuntaopin oppimistulokset perusopetuksen päättövaiheessa 2011*. Koulutuksen seurantaraportit 2012:3. Helsinki: Opetushallitus, 147–152.
- Löfström, J. 2012b. Yhteiskunnallisten aineiden didaktiikan tutkimus Suomessa. Teoksessa Kallioniemi, A., & Virta, A. (toim.), *Ainedidaktiikka tutkimuskohteena ja tiedonalana*. Turku: Suomen kasvatustieteellinen seura, 172–194.
- Löfström, J., Virta, A. & van den Berg M. 2010. Who Actually Sets the Criteria for Social Studies Literacy? The National Core Curricula and the Matriculation Examination as Guidelines for Social Studies teaching in Finland in the 2000's. *Journal of Social Sciences Education*. 9 (4), 6–14.
- Malin, A. 2005. Monitasomallit ja tutkimusaineiston rakenne. Koulutuksen tutkimuslaitos: Jyväskylän yliopisto. Noudettu 13.10.2012 osoitteesta <https://www.jyu.fi/erillis/agora-center/tutkimus/.../Monitasomallit.ppt>
- Malinen P. 1987. Opetussuunnitelmatutkimuksen jäsentelyä yhteistyöryhmässä. Teoksessa Malinen, P. & Kansanen, P. (toim.). *Opetussuunnitelman tutkimukselliset kehyykset. Opetussuunnitelmatutkijoiden yhteistyöryhmän ja Suomen kasvatustieteellisen seuran opetussuunnitelma seminaari 25.–26.9.1986*. Helsingin yliopisto. Opettajankoulutuslaitos. Tutkimuksia 48, 7–15.
- Mathinson, S. & Freeman, M. 2003. Constraining elementary teacher's work: Dilemma and paradoxes created by state mandated testing. *Education Policy Analysis Archive* 11 (34), 1–23.
- Mathinson, S., Ross, E. W. & Vinson K. D. 2006. *Defining the Social Studies Curriculum: Influence of Resistance to Curriculum Standards and Testing in Social Studies*. Teoksessa Ross, E.W. (toim.). *Social Studies Curriculum Purposes, Problems and Possibilities*. Third Edition. Albany: State University of New York Press, 99–114.
- Mattila, L. 2002. Perusopetuksen matematiikan oppimistulosten kansallinen arviointi 9. vuosiluokalla 2002. Oppimistulosten arviointi 8/2002. Helsinki: Opetushallitus.
- Mattila, L. 2005. Perusopetuksen matematiikan kansalliset oppimistulokset 9. vuosiluokalla 2004. Oppimistulosten arviointi 2/2005. Helsinki: Opetushallitus.
- Mattila, L. 2010. Perusopetuksen äidinkielen ja matematiikan päättöarvosanat. *Lisensiaatin-tutkimus*. Helsingin yliopisto. Käyttäytymistieteellinen tiedekunta.
- McKernan, J. 2008. *Curriculum and Imagination. Process Theory, Pedagogy an Action Research*. London: Routledge.
- McMillan, J., Myran, S., & Workman, D. 2002. Elementary Teachers' Classroom Assessment and Grading Practices. *The Journal of Educational Research*, 95(4), 203–213.
- McNeill, L. M. 2000. *Contradictions of school reform: Education costs of standardized testing*. New York: Routledge.
- Mertler, C.A. 2000. Teacher-centered fallacies of classroom assessment validity and reliability. *Mid-Western Educational Researcher*, 13(4), 29–35.
- Messick, S. 1975. The Standard problem: Meaning and values in in measurement and evaluation. *American Psychologist*, 30, 294–304.
- Messick, S. 1980. Test Validity and the Ethics of Assessment. *American Psychologist*, 35(11), 1012–1027.

- Messick, S. 1988. The once and future issues of validity. Assessing the meaning and consequences of measurement. Teoksessa Wainer, H. & Braun, H. (toim). Test validity. Hillsdale, NJ: Lawrence Erlbaum, 33-45.
- Messick, S. 1989a. Validity. Teoksessa Linn, R. L. (toim). Educational measurement (3rd Ed.). New York: American Council of Education and Macmillan, 13-103.
- Messick, S. 1989b. Meaning and values in test validation: the science and ethics of assessment. Educational Researcher 18 (2), 5-11.
- Messick, S. 1994. The interplay of evidence and consequences in the validation of performance assessment. Educational Researcher, 23(2), 13-23.
- Messick, S. 1995. Validation of Inferences from Persons' Responses and Performances as Scientific Inquiry into Score Meaning. American Psychologist, 50 (9), 741-749.
- Metsämuuronen, J. 2006. Tutkimuksen tekemisen perusteet ihmistieteissä 2. International Methelp Ky. Vaajakoski: Gummerus.
- Metsämuuronen, J. 2009. Metodit arvioinnin apuna. Perusopetuksen oppimistulosarvioinnin ja -seurantojen menetelmäratkaisut Opetushallituksessa. Oppimistulosten arviointi 1/2009. Helsinki: Opetushallitus.
- Ministère de l'Éducation Nationale. Noudettu 31.8.2012 osoitteesta <http://www.education.gouv.fr/>
- Mislevy, R. J., Yamamoto, K. & Anacker, S. 1991. Toward a test theory for assessing student understanding (Tutkimusraportti No. RR-91-32-ONR). Princeton, NJ: Educational Testing Service.
- McMillan, J. 2007. Formative Classroom Assessment: The Key to Improving Student Achievement. Teoksessa McMillan, J. 2007. Formative Classroom Achievement. New York: Teachers College Press, 1-7.
- Moss, P. A. 2003. Reconceptualizing validity for classroom assessment. Educational Measurement: Issues and Practice, 22 (4), 13-25.
- Moss, P. A. 2007. Reconstructing Validity. Educational Research, 36, 470-476.
- Moss, P. A., Beck, J.S., Ebbs, C., Matson, B. Muchmore, J. & Steele, D., Taylor, C. & Herter, R. 1992. Portfolios, accountability, and an interpretive approach to validity. Educational Measurement: Issues and Practice, 11 (2), 12-21.
- NCCS (National Council for the Social Studies). Noudettu 31.8.2012 osoitteesta <http://www.socialstudies.org/>
- Nichols, P. D. 1994. A framework for developing cognitively diagnostic assessments. Review of Educational Research, 64, 575-603.
- Niemi, H. 2009. Why Finland on the top? Reflections on the reasons for the PISA success. Keynote paper presented at the 3rd Redesigning Pedagogy International Conference June 2009, Singapore. Noudettu 29.11.2012 osoitteesta http://conference.nie.edu.sg/2009/downloads/hannele_niemi_2009_rp_keynote_fullpaper.pdf
- No Child Left Behind Act of 2001. Noudettu 7.11.2012 osoitteesta <http://www2.ed.gov/policy/elsec/leg/esea02/107-110.pdf>
- Nousiainen, L. & Piekkari, U. 2007. Osallistuva oppilas, yhteisöllinen koulu. Toimivan oppilaskunnan opas. Opetusministeriön julkaisuja 2007:32.
- Nummenmaa T., Konttinen R., Kuusinen J. & Leskinen E. 1997. Tutkimusaineiston analyysi. Porvoo: WSOY.

- Numminen, U. & Ouakrim-Soivio, N. 2009. Joustava perusopetus. JOPO-toiminnan aloittaminen ja vakiinnuttaminen. Opetusministeriön julkaisuja 2009:26.
- Näveri, L. 2009. Aritmetiikasta algebraan: Muutoksia osaamisessa peruskoulun päättöloukalla 20 vuoden aikana. Helsingin yliopisto, käyttäytymistieteellinen tiedekunta, soveltavan kasvatustieteen laitos, tutkimuksia 309. Noudettu 31.8.2012 osoitteesta <https://helda.helsinki.fi/bitstream/handle/10138/20064/aritmeti.pdf?sequence=1>
- Odenstad, C. 2010. Prov och bedömning i samhällskunskap. En analys av gymnasielärares skriftliga prov. Karlstad University Studies 2010:41.
- OECD 2007. Education at Glance 2007.
- OECD 2010. Education at Glance 2010.
- OECD 2011. Reviews of Evaluation and Assessment in Education. Sweden. Noudettu 30.8.2012 osoitteesta <http://www.oecd.org/dataoecd/38/42/47169533.pdf>
- Opetushallitus 1996. Peruskoulun oppilasarvioinnin vertailtavuuden parantaminen.
- Opetushallitus 1997. Muistio 25.10.1997.
- Opetushallitus. Edu.fi-portaali. Noudettu 31.8.2012 osoitteesta www.edu.fi
- Opetushallitus. Perusopetuksen oppilaan arviointi. Noudettu 31.8.2012 osoitteesta http://www.oph.fi/saadokset_ja_ohjeet/ohjeita_koulutuksen_jarjestamiseen/perusopetuksen_jarjestaminen/perusopetuksen_oppilaan_arviointi
- Opetushallitus 2012. Demokratiaselvitys. Raportit ja selvitykset 2012:27. Helsinki: Opetushallitus. Noudettu 31.10.2012 osoitteesta http://www.oph.fi/download/139654_Demokratiakasvatusselvitys.pdf
- Opetusministeriö 2001. Perusopetuksen uudistamistyöryhmän muistio 11:2001. Noudettu 30.8.2012 osoitteesta http://www.minedu.fi/export/sites/default/OPM/Julkaisut/2001/liitteet/opm_523_perusopetuksen_uudistamistyoryhma.pdf?lang=fi
- Opetusministeriö 2008. Koulutuksen arviointisuunnitelma vuosille 2009–2011. Opetusministeriön julkaisuja 2008: 38. Noudettu 30.8.2012 osoitteesta <http://www.minedu.fi/export/sites/default/OPM/Julkaisut/2008/liitteet/opm38.pdf?lang=fi>
- Opetusministeriö 2010. Perusopetuksen laatukriteerit. Opetusministeriön julkaisuja 2010:6. Noudettu 30.8.2012 osoitteesta <http://www.minedu.fi/export/sites/default/OPM/Julkaisut/2010/liitteet/opm06.pdf>
- Opetus- ja kulttuuriministeriö 2012. Tulevaisuuden perusopetus – valtakunnalliset tavoitteet ja tuntijako. Opetus- ja kulttuuriministeriön työryhmämuistioita ja selvityksiä 2012:6.
- Ouakrim-Soivio, N. 2013. ”Sitä osataan, mitä oppiaineiden tavoitteissa määritellään ja päätötarvioinnin kriteereillä arvioidaan”. Teoksessa Räsänen, A. (toim.) 2013. Oppimisen arvioinnin kontekstit ja käytännöt. Raportit ja selvitykset 2013:3. Helsinki: Opetushallitus, 141–158. Noudettu 18.8.2013 osoitteesta http://www.oph.fi/download/149650_Oppimisen_arvioinnin_kontekstit_ja_kaytannot_2.pdf
- Ouakrim-Soivio, N. & Kuusela, J. 2012. Historian ja yhteiskuntaopin oppimistulokset perusopetuksen päättövaiheessa 2011. Koulutuksen seurantaraportit 2012:3. Helsinki: Opetushallitus. Noudettu 29.11.2012 osoitteesta http://www.oph.fi/download/139945_Historian_ja_yhteiskuntaopin_oppimistulokset_perusopetuksen_paattovaiheessa_2011.pdf
- Oscarsson, V. 2005. Elevers syn på globala förhållanden och framtiden. Göteborgs universitet Institutionen för pedagogik och didaktik 2005:05.

- Oscarsson, V. & Svingby G. 2005. Sätt lärande I centrum I samhällsorienterande ämnen. I Grundskolans ämnen i ljuset av Nationella utvärderingen 2003. Nuläge och framåtblickar. Stockholm: Skolverket, 122–134. Noudettu 22.1.2013 osoitteesta <http://www.mai.liu.se/~chber/kurser/pdf1497.pdf>
- Palomäki, S. & Heikinaro-Johansson, P. 2011. Liikunnan oppimistulosten seuranta-arviointi perusopetuksessa 2010. Koulutuksen seurantaraportit 2001:4. Helsinki: Opetushallitus.
- Pellegrino, J. W., Baxter, G. P., & Glaser, R. 1999. Addressing the 'Two Disciplines' Problem: Linking Theories of Cognition and Learning with Assessment and Instructional Practice, *Review of Research in Education*, 24, 307–353.
- Pickford, R. & Brown, S. 2006. *Assessing skills and practice*. London: Routledge.
- POPS 1985. Peruskoulun opetussuunnitelman perusteet. Helsinki: Kouluhallitus.
- POPS 1994. Peruskoulun opetussuunnitelman perusteet. Helsinki: Opetushallitus.
- POPS 2004. Perusopetuksen opetussuunnitelman perusteet. Helsinki: Opetushallitus.
- Perusopetuksen oppilaan arvioinnin perusteet. 1999. Helsinki: Opetushallitus.
- Perusopetuksen päättöarviointin kriteerit. Arvosanan hyvä (8) kriteerit yhteisissä oppiaineissa. 1999. Helsinki: Opetushallitus.
- Peruskouluasetus (718/1984)
- Perusopetuslaki (628/1998)
- Perusopetusasetus (825/1998)
- Phye, G. D. 1997. *Classroom Assessment: A Multidimensional Perspective*. Teoksessa Phye G. D. *Handbook of Classroom Assessment. Learning, Adjustment and Achievement*. San Diego: Academic Press 33–51.
- Pinar, E.F. 2004. *What is Curriculum Theory?* Mahwah: Lawrence Erlbaum Associated Publishers.
- Popham, W. J. 1997. Consequential validity: Right concern-wrong concept. *Educational Measurement: Issues and Practice*, 16(2), 9–13.
- Puhakka, E. 2002. Civic-tutkimuksen otannasta. Teoksessa Brunell, V & Törmäkangas, K. (toim.). 2002. *Tulevaisuuden yhteiskunnan rakentajat. Yhteiskunnallisen opetuksen kansainvälinen tutkimus Suomen näkökulmasta*. Jyväskylä: Jyväskylän yliopisto, 5–11.
- Quellmalz, E. & Hoskyn, J. 1997. *Classroom Assessment of Reasoning Strategies*. Teoksessa Phye G. D. *Handbook of Classroom Assessment. Learning, Adjustment and Achievement*. San Diego: Academic Press, 103–130.
- Randal, J. & Engelhard, G. 2009. Differences Between Teachers' Grading Practices in Elementary and Middle Schools. *The Journal of Educational Research*, 102(3), 175–185.
- Rantala, J. 1999. Historian kouluopetus muuttuu. *Historiallinen aikakauskirja* (3), 258–263.
- Rantanen, P. 2003. Enemmän vähemmällä. Monivalintatehtävien mittaustarkkuuden nostaminen: Kasvatusalan tutkimuksia. Turku: Suomen kasvatustieteellinen seura.
- Rauste-von Wright M. & von Wright J. 1994. *Oppiminen ja koulutus*. Porvoo: Werner Söderström Osakeyhtiö.
- Ravitch, D. 2010. *The Death and Life of the Great American School System. How Testing and Choice Are Undermining Education*. New York: Basic Books
- Regressiomallilla estimoitujen residuaalien analysointi. Helsingin yliopisto. Noudettu 11.11.2012 osoitteesta <http://www.helsinki.fi/atk/tilasto/tao/regresiduals.html>

- Rimpelä, M. & Bernelius, V. (toim.). 2010. Peruskoulujen oppimistulokset ja oppilaiden hyvinvointi eriytyvällä Helsingin seudulla. MetrOP-tutkimus 2010–2013. Mitä tiedettiin tutkimuksen käynnistytessä keväällä 2010? Geotieteiden ja maantieteen laitoksen julkaisuja B1. Helsinki: University of Helsinki, 49- 54. Noudettu 29.11.2012 osoitteesta https://helda.helsinki.fi/bitstream/handle/10138/17076/MetrOP-raportti_1_verkkoversio.pdf?sequence=1
- Rinne, R. 1984. Suomen oppivelvollisuuskoulun opetussuunnitelman muutokset vuosina 1916–1970. Julkaisuja c 44. Turun yliopisto.
- Robson, C. 2001. Käytännön arvioinnin perusteet. Opas evaluaation tekijöille ja tilaajille. Helsinki: Tammi.
- Rokka, P. 2011. Peruskoulun ja perusopetuksen vuosien 1985, 1994 ja 2004 opetussuunnitelmien perusteet poliittisen opetussuunnitelman teksteinä. Kasvatustieteiden yksikkö. Tampereen yliopisto. Noudettu 24.1.2013 osoitteesta <http://tampub.uta.fi/handle/10024/66741>
- Rusama, J. 2002. Uskonto, elämäntietä ja tapakasvatus. Oppimistulosten arviointi perusopetuksen päättövaiheessa. Oppimistulosten arviointi 5/2002. Helsinki: Opetushallitus.
- Salmio, Kaija. 2004. Esimerkkejä peruskoulun valtakunnallisista arviointihankkeista kestävä kehityksen didaktiikan näkökulmasta. Vuosien 1993–1995 valtakunnalliset kokeet ja vuoden 1998 luonnontieteiden oppimistulosten arviointi. Joensuun yliopiston kasvatustieteellisiä julkaisuja 98.
- Sanders, C. E. 1997. A Preschool Assessment. Teoksessa Phye G.D. Handbook of Classroom Assessment. Learning, Adjustment and Achievement. San Diego: Academic Press, 227–264.
- Sarjala, J. 2002. Koulutuspolitiikan uudet ohjausmenettelyt. Teoksessa Olkinuora E., Jakkusihvonen, R., & Mattila E. (toim.). 2002. Koulutuksen arviointi – lähtökohtia, malleja ja tilannekatsauksia. Turun yliopiston kasvatustieteiden tiedekunnan julkaisuja B:70. Turku: Turun yliopiston kasvatustieteiden tiedekunta, 9-14. Noudettu 29.8.2012 osoitteesta <http://www.edu.utu.fi/erillisyksikot/koulutuksenarviointi/julkaisut/arviointi.pdf>
- Scheurich, J.J. 1997. The Masks of Validity. A Deconstructive Investigation. Teoksessa Research Method in Postmodern. Qualitative Studies Series, n:o 3. London: The Falmer Press.
- Shepard, L.A. 2006. Classroom Assessment. Teoksessa Brennan R.L. 2006. Educational Measurement (4th edition). Westport, CT: American Council on Education and Praeger Publishers, 623–646.
- Silverström, C. 2002. Utvärdering av inlärningsresultat i modersmål och litteratur i åk 9 våren 2001. Utvärdering av inlärningsresultat 4/2002. Helsingfors: Utbildningsstyrelsen.
- Silverström, C. 2003. Modersmål och litteratur i sex år. En utvärdering av inlärningsresultat i modersmål och litteratur hos elever som slutfört årskurs 6 i den grundläggande utbildningen år 2002. Utvärdering av inlärningsresultat 5/2003. Helsingfors: Utbildningsstyrelsen.
- Silverström, C. 2004. ”Ett viktigt men utmanande ämne”. En utvärdering av inlärningsresultat i modersmål och litteratur i årskurs 9 våren 2003. Utvärdering av inlärningsresultat 3/2004. Helsingfors: Utbildningsstyrelsen.

- Silverström, C. 2006. Modersmål och litteratur i nio år. En utvärdering av inlärningsresultat i modersmål och litteratur i årskurs 9 våren 2005. Utvärdering av inlärningsresultat 2/2006. Helsingfors: Utbildningsstyrelsen.
- Silverström, C. 2008. Modersmålskunskaper, attityder och betyg. En utvärdering av inlärningsresultat i modersmål och litteratur efter årskurs 6 år 2007. Utvärdering av inlärningsresultat 4/2008. Helsingfors: Utbildningsstyrelsen.
- SIRIS (Skolverkets Internetbaserade Resultat- och kvalitets Informations System). Noudettu 21.9.2012 osoitteesta <http://sir.is.skolverket.se/apex/f?p=SIRIS:1:0::NO::>
- Skolverket 2005. Nationella utvärderingen av grundskolan 2003. Samhällsorienterande ämnen. Ämnesrapport till rapport 252. Stockholm: Skolverket.
- Skolverket 2007a. Provbetyg – Slutbetyg – Likvärdig bedömning? En statistisk analys av sambandet mellan nationella prov och slutbetyg i grundskolan 1998-2006. Sammanfattning av rapport 300. Stockholm: Skolverket.
- Skolverket 2007b. Skolverkets lägesbedömning 2007 Förskoleverksamhet, skolbarnsomsorg, skola och vuxenutbildning. Rapport 303: 2007. Stockholm: Skolverket.
- Skolverket 2009. Likvärdig betygssättning i gymnasieskolan? En analys av sambandet mellan nationella prov och kursbetyg. Rapport 338:2009. Stockholm: Skolverket.
- Skolverket 2011a. Läroplan för grundskolan, förskoleklassen och fritidshemmet 2011. Stockholm: Edita. Noudettu 7.9.2013 osoitteesta http://www.skolverket.se/om-skolverket/visa-enskildpublikation?_xurl_=http%3A%2F%2Fwww5.skolverket.se%2Fwtpub%2Fws%2Fskolbok%2Fwpubext%2Ftrycksak%2FRecord%3Fk%3D2575
- Skolverket 2011b. Planering och genomförande av undervisningen – för grundskolan, grundsärskolan, specialskolan och sameskolan. Skolverket: Stockholm. Noudettu 7.9.2013 osoitteesta http://www.skolverket.se/om-skolverket/visa-enskild-publikation?_xurl_=http%3A%2F%2Fwww5.skolverket.se%2Fwtpub%2Fws%2Fskolbok%2Fwpubext%2Ftrycksak%2FRecord%3Fk%3D2698
- Skolverket 2012. Kommentarmaterial till kunskapskraven i samhällskunskap. Skolverket: Stockholm. Noudettu 7.9.2013 osoitteesta http://www.skolverket.se/om-skolverket/visa-enskild-publikation?_xurl_=http%3A%2F%2Fwww5.skolverket.se%2Fwtpub%2Fws%2Fskolbok%2Fwpubext%2Ftrycksak%2FRecord%3Fk%3D2826
- Skolverket. Läroplaner, ämnesplaner och kursplaner. Noudettu 7.9.2013 osoitteesta <http://www.skolverket.se/laroplaner-amnen-och-kurser>
- Skolverket föreskrift. 2011:19. Den nya skollagen – för kunskap, valfrihet och trygghet. Lättläst version av sammanfattningen av regeringens proposition 2009/2010:165. Noudettu 30.8.2012 osoitteesta <http://www.lerum.se/Documents/Dokumenthanteraren/Utbildning%20och%20barnomsorg/L%3A%44ttl%3A%44st/skollagen%20ll.pdf?epslanguage=sv>
- Skolverket föreskrift 2011: 157. Föreskrift: Skolverkets föreskrifter om utformningen av slutbetyg i grundskolan, grundsärskolan och specialskolan. Noudettu 31.8.2012 osoitteesta http://www.skolverket.se/lagar-och-regler/2.3134/2.5007?_xurl_=http%3A%2F%2Fwww4.skolverket.se%3A8080%2Fwtpub%2Fws%2Fskolfs%2Fwpubext%2Ffs%2FRecord%3Fk%3D2331

- Skolverket föreskrift 2011: 196. Statens skolverks författningssamling. Skolverkets föreskrifter om vilka kurser enligt kursplaner som motsvaras av kurser enligt ämnesplaner och får ingå i en gymnasieexamen. Noudettu 30.8.2012 osoitteesta http://www.umu.se/digitalAssets/89/89234_versttningsnyckel-gy2011.pdf
- Smith, J. K. 2003. Reconsidering reliability in classroom assessment and grading. *Educational Measurement: Issues and Practice*, 22(4), 26–33.
- Stiggins, R. J., Frisbie, D. A., & Griswold, P. A. 1989. Inside high school grading practices: Building a research agenda. *Educational Measurement: Issues and Practice*, 8(2), 5–14.
- Suoninen, A., Kupari, P., Törmäkangas, K., 2010. ICCS. Kansainvälisen ICCS 2009-tutkimuksen päätulokset. Jyväskylä: Koulutuksen tutkimuslaitos. Noudettu 29.11. 2012 osoitteesta <https://jyx.jyu.fi/dspace/bitstream/handle/123456789/37474/978-951-39-4107-9.pdf?sequence=1>
- Suutarinen, S. 2000. Yhteiskunnallisen opetuksen asema Suomessa. Teoksessa Suutarinen S. (toim.). Nuoresta pätevä kansalainen. Yhteiskunnallinen opetus Suomen peruskoulussa. IEA Civics – Nuori Kansalainen tutkimuksen julkaisuja I. Jyväskylä: Koulutuksen tutkimuslaitos, 33–54.
- Suutarinen, S. 2002. Tietävillä nuorilla syrjäytyneiden asenteet. Suomalaisten nuorten yhteiskunnallinen ajattelu ja toiminta kansainvälisessä vertailussa. Teoksessa Brunell, V. & Törmäkangas, K. (toim.). Tulevaisuuden yhteiskunnan rakentajat. Yhteiskunnallisen opetuksen kansainvälinen tutkimus Suomen näkökulmasta. Jyväskylä: Koulutuksen tutkimuslaitos, 11–60. Noudettu 29.8.2012 osoitteesta http://ktl.jyu.fi/img/portal/17757/Jatkoanalyysi_D057.pdf?cs=1276080719
- Suutarinen, S. (toim.). 2006 a. Turvallinen keskustelu vähäistä Suomen peruskouluissa – oppilaitokset tulemaan keskustelevaa kansalaisuutta. Teoksessa Suutarinen, S. (toim.). Aktiiviseksi kansalaiseksi. Kansalaisvaikuttamisen haaste. Juva: PS-kustannus, 63–97.
- Suutarinen, S. (toim.). 2006 b. Tietopainotteisen kansalaiskasvatuksen aika ohi Suomessa – miten kansalaiskasvatusta uudistetaan? Teoksessa Suutarinen, S. (toim.). Aktiiviseksi kansalaiseksi. Kansalaisvaikuttamisen haaste. Juva: PS-kustannus, 99–123.
- Suutarinen, S., Brunell, V., Poutiainen, A., Puhakka, E., Saari, H. & Törmäkangas, K., 2001. Suomen peruskoulun päättövaiheessa olevien nuorten yhteiskunnalliset tiedot, taidot, käsitykset, asenteet ja toiminta 28 maan vertailussa. Jyväskylä: Koulutuksen tutkimuslaitos. Noudettu 29.11.2012 osoitteesta http://ktl.jyu.fi/img/portal/17757/Yhteen veto_1999.pdf?cs=1276080659
- Torney-Purta, J., Lehman, R., Oswald, H. & Schulz, W. 2001. Citizenship and Education in Twenty-Eight countries: Civic Knowledge and Engagement at Age Fourteen. Amsterdam: International Association for the Evaluation of Educational Achievement. Noudettu 29.8.2012 osoitteesta http://www.iea.nl/fileadmin/user_upload/Publications/Electronic_versions/CIVED_Phase2_Age_Fourteen.pdf
- Torney-Purta, J., Schwille, J. & Amadeo, J. 1999. Civic education across countries: twenty four national case studies from the IEA Civic Education Project. Amsterdam: International Association for the Evaluation of Educational Achievement.
- Toropainen, O. 2002. Nationell utvärdering av inlärningsresultat i finska i åk 9 våren 2001. Utvärdering av inlärningsresultat 1/2002. Helsingfors: Utbildningsstyrelsen.

- Toulmin, S. E. 2003 [1958]. *The Uses of Argument*. Updated edition first published 2003, 8th printing 2008. Cambridge: Cambridge University Press.
- Tukeyn b (Post hoc-testaus). Noudettu 16.1.2013 osoitteesta: <http://www.uku.fi/~mauranen/spss/jatko/anova18.html>
- Tuokko, E. 2002. Perusopetuksen päättövaiheen ruotsin kielen oppimistulosten kansallinen arviointi 2001. Oppimistulosten arviointi 3/2002. Helsinki: Opetushallitus.
- Tuokko, E. 2003. Perusopetuksen päättövaiheen englannin kielen oppimistulosten kansainvälinen arviointi 2002 (Espanja – Hollanti – Norja – Ranska – Ruotsi – Saksa – Suomi – Tanska). Suomen tulokset. Oppimistulosten arviointi 3/2003. Helsinki: Opetushallitus.
- Tuokko, E. 2009. Miten ruotsia osataan peruskoulussa? Perusopetuksen päättövaiheen ruotsin kielen B-oppimäärän oppimistulosten kansallinen arviointi 2008. Oppimistulosten arviointi 2/2009. Helsinki: Opetushallitus.
- Törmä, E. 1998. Peruskoulun oppilasarvostelua kehittämässä. Teoksessa Törmä, E. (toim.). Aina ajankohtainen arviointi. Opettaja avain arvioinnin kehittämässä. *Journal of Teacher researchers* 3/1998. Jyväskylä: TUOPE, 8–25.
- Uusikylä, K. & Atjonen, P. 2000. *Didaktiikan perusteet*. Helsinki: Werner Söderström Osakeyhtiö.
- Valtioneuvoston päätös peruskoulun tuntijaosta (839/1993).
- Valtioneuvoston asetus (423/2012) Perusopetusasetuksen 3 §:n muuttamisesta.
- Valtioneuvoston asetus (805/2008) Opetushallituksesta.
- Valtioneuvoston asetus (1435/2001) perusopetuslaissa tarkoitettujen opetuksen valtakunnallisista tavoitteista ja perusopetuksen tuntijaosta.
- Wickström, C. 2006. *Education and Assessment in Sweden, Assessment in Education: Principles, Policy and Practice*, Vol. 13, No. 1, 113–128.
- Wiggins, G. 1996/1997. Practicing what we preach in designing authentic assessments. *Educational Leadership* 54 (4), 18–25.
- Virta, A. 1993. Oppilasarviointi yhteiskuntatiedon oppimisprosessin osana. Teoksessa Arola, P., Elio, K., Huuhtanen, T. & Pilli, A. *Yhteiskuntatieto koulussa*. Helsinki: Yliopistopaino, 135–158.
- Virta, A. 1995. Abiturientin historian ja yhteiskuntaopin tieto. Reaalikoevastaukset oppilaiden tiedonrakenteiden ja tiedon laadun ilmentäjinä. *Turun yliopiston julkaisuja C:112*. Turku.
- Virta, A. 1999. Uudistuva oppimisen arviointi. Mahdollisuuksia ja varauksia. *Turun yliopiston kasvatustieteiden tiedekunta. Julkaisusarja B:65*. Turku.
- Virta, A. 2000. Perusopetuksen opetussuunnitelma ja demokratiaan kasvaminen. Teoksessa Suutarinen, S. (toim.). 2000. Nuoresta pätevä kansalainen. Yhteiskunnallinen opetus Suomen peruskoulussa. IEA Civics – Nuori Kansalainen tutkimuksen julkaisuja I. Jyväskylä: Koulutuksen tutkimuslaitos, 55–86.
- Virta, A. 2002. Kohti aktiivista ja tutkivaa historian ja yhteiskuntatiedon oppimista. Lähtökohdat puntarissa. Teoksessa Löfström, J. (toim.). *Kohti tulevaa menneisyyttä. Historiallis-yhteiskunnallinen kasvatus uudella vuosituhannella*. Jyväskylä: PS-kustannus, 35–65.

- Virta, A. 2006. Kansalaisen osallistumiskanavat yhteiskuntaopin kirjojen kuvaamina. Teoksessa Rantala, J. & Salminen, J. (toim.). 2006. Kansalaisvaikuttamisen edistäminen koulussa ja opettajakoulutuksessa. Helsinki: Historiallis-yhteiskuntatiedollisen kasvatuksen tutkimus- ja kehittämiskeskus, 96–108. Noudettu 24.1.2013 osoitteesta <http://www.enorssi.fi/hankkeet/kansalaisvaikuttaminen-1/julkaisut-1/tutkimuksia5.pdf>
- Vitikka, E. 2009. Opetussuunnitelman mallin jäsenys. Sisältö ja pedagogiikka kokonaisuuden rakentajina. Kasvatusalan tutkimuksia 44. Jyväskylä: Suomen kasvatustieteellinen seura.
- Vitikka, E. & Hurmerinta, E. 2011. Kansainväliset opetussuunnitelmavirtaukset. Raportit ja selvitykset 11:4. Helsinki: Opetushallitus. Noudettu 29.11.2012 osoitteesta http://www.oph.fi/download/131521_Kansainvaliset_opetussuunnitelmasuuntaukset.pdf
- Wolf, A. 2000. A Comparative Perspective on Educational Standards. Teoksessa Goldstein H., & Heath, A. (toim.). 2000. Educational Standards. Oxford: Oxford University Press, 9–37.
- Väisänen, T. 2003. Perusopetuksen 9. luokan A-kielenä opetettavan ranskan, saksan ja venäjän oppimistulosten kansallinen arviointi. Oppimistulosten arviointi 1/2003. Helsinki: Opetushallitus. Noudettu 29.11. 20012 osoitteesta http://www.oph.fi/download/115538_perusopetuksen_9_luokan_a_kielena_opetettavan_ranskan_saksan_ja_venajan_oppimistulosten_kansallinen_arviointi.pdf
- Vänttinen, M. 2011. Oikeasti hyvä numero. Oppilaiden arvioinnin totuudet ja totuustuotanto rinnakkaiskoulusta yhtenäiskouluun. Publications of the University of Eastern Finland Dissertations in Education, Humanities, and Theology 17. Joensuu: University of Eastern Finland. Noudettu 29.11.2012 osoitteesta http://epublications.uef.fi/pub/urn_isbn_978-952-61-0514-7/urn_isbn_978-952-61-0514-7.pdf
- Zeller, R. A. 1988. Validity. Teoksessa Keevs, J.P. (toim.). Educational Research, Methodology, and Measurement: An International Handbook. Oxford. Pergamon Press.

LIITETAULUKKO 1. Tyttöjen ja poikien keskimääräiset arvosanat yhteiskuntaopissa, historiassa, äidinkiessä ja matematiikassa.

Sukupuoli		Yhteiskuntaopin arvosana	Historian arvosana	Äidinkielen arvosana	Matematiikan arvosana
poika	Keskiarvo	7,63	7,75	7,47	7,51
	N	2 006	2 010	2 013	2 016
	Keskihajonta	1,21	1,26	1,09	1,41
tyttö	Keskiarvo	8,06	7,89	8,26	7,73
	N	2 097	2 099	2 105	2 103
	Keskihajonta	1,14	1,24	0,98	1,35
Kaikki	Keskiarvo	7,85	7,82	7,87	7,62
	N	4 103	4 109	4 118	4 119
	Keskihajonta	1,19	1,25	1,11	1,38

LIITETAULUKKO 2. Yhteiskuntaopin, historian, matematiikan ja äidinkielen arvosanojen korrelaatiot oppilaiden itsensä ilmoittamien arvosanatietojen mukaan laskettuna.

		Yhteiskuntaopin arvosana	Historian arvosana	Äidinkielen arvosana	Matematiikan arvosana
Yhteiskuntaopin arvosana	r_{xy}	1	0,76	0,65	0,61
	P (kaksisuuntainen)		0,00	0,00	0,00
	N	4 103	4 088	4 095	4 097
Historian arvosana	r_{xy}	0,76	1,00	0,61	0,59
	P (kaksisuuntainen)	0,00		0,00	0,00
	N	4 088	4 109	4 104	4 102
Äidinkielen arvosana	r_{xy}	0,65	0,61	1,00	0,62
	P (kaksisuuntainen)	0,00	0,00		0,00
	N	4 095	4 104	4 118	4 115
Matematiikan arvosana	r_{xy}	0,61	0,59	0,62	1,00
	P (kaksisuuntainen)	0,00	0,00	0,00	
	N	4 097	4 102	4 115	4 119

LIITETAULUKKO 3a. Oppilaiden saama keskimääräinen ratkaisuosuus yhteiskuntaopin seuranta-arvioinnissa.

	N	Minimi	Maksimi	Keskiarvo	Keskihajonta
Yhteiskuntaopin ratkaisuosuus (%)	4 158		96,97	65,15	12,74

LIITETAULUKKO 3b. Oppilaiden saama keskimääräinen ratkaisuosuus historian seuranta-arvioinnissa.

	N	Minimi	Maksimi	Keskiarvo (%)	Keskihajonta
Historian ratkaisuosuus (%)	4 158	5,48	94,52	51,13	13,83

LIITETAULUKKO 3c. Yhteiskuntaopin keskimääräinen ratkaisuosuus arvosanoittain.

Yhteiskuntaopin arvosana	Yhteiskuntaopin keskimääräinen ratkaisuosuus (%)	N	Keskihajonta
4	46,36	5	10,20
5	50,15	140	11,10
6	54,61	393	10,37
7	59,59	972	11,09
8	66,19	1 252	10,67
9	72,26	1 104	10,21
10	77,55	237	10,08
Yhteensä	65,23	4 103	12,69
Yhteiskuntaopin arvosanojen selitysosuus arviointimenestyksestä (eetan neliö).		Eeta	Eetan neliö
Yhteiskuntaopin ratkaisuosuus (%) *Yhteiskuntaopin arvosana		0,55	0,30

LIITETAULUKKO 3d. Historian keskimääräinen ratkaisuosuus päättöarvosanoittain.

Historian päättöarvosana	Historian keskimääräinen ratkaisuosuus (%)	N	Keskihajonta
5	34,11	152	9,41
6	39,84	481	9,85
7	45,33	943	10,62
8	51,88	1 229	10,91
9	59,11	974	11,74
10	67,14	330	12,36
Yhteensä	51,25	4 109	13,79
Historian arvosanojen selitysosuus arviointimenestyksestä (eetan neliö).		Eeta	Eetan neliö
Historian ratkaisuosuus (%) * Historian arvosana		0,60	0,36

LIITETAULUKKO 4. Oppilaiden yhteiskuntaopin ja historian ratkaisuosuuksien, arvosanojen ja arvosanoista laskettujen odotusarvojen ylitysten ja alitusten korrelaatiot.

		Yhteis- kuntaopin ratkaisuo- suus (%)	Yhteis- kuntaopin arvosana	Historian ratkaisuo- suus (%)	Historian arvosana	Ero odotus- arvosta yhteis- kunta- opissa	Ero odotus- arvosta histori- assa
Yhteis- kuntaopin ratkaisuosuus (%)	r_{xy}	1,00	0,55	0,74	0,55	0,80	0,46
	P (kaksi- suuntainen)		0,00	0,00	0,00	0,00	0,00
	N	4 158	4 103	4 158	4 109	4 103	4 109
Yhteis- kuntaopin arvosana	r_{xy}	0,55	1,00	0,53	0,76	- 0,03	0,02
	P (kaksi- suuntainen)	0,00		0,00	0,00	0,10	0,15
	N	4 103	4 103	4 103	4 088	4 103	4 088
Historian ratkaisuosuus (%)	r_{xy}	0,74	0,53	1,00	0,60	0,51	0,71
	P (kaksi- suuntainen)	0,00	0,00		0,00	0,00	0,00
	N	4 158	4 103	4 158	4 109	4 103	4 109
Historian arvo- sana	r_{xy}	0,55	0,76	0,60	1,00	0,13	- 0,09
	P (kaksi- suuntainen)	0,00	0,00	0,00		0,00	0,00
	N	4 109	4 088	4 109	4 109	4 088	4 109
Ero odotus- arvosta yhteis- kuntaopissa	r_{xy}	0,80	- 0,03	0,51	0,13	1,00	0,53
	P (kaksi- suuntainen)	0,00	0,10	0,00	0,00		0,00
	N	4 103	4 103	4 103	4 088	4 103	4 088
Ero odotus- arvosta histo- riassa	r_{xy}	0,46	0,02	0,71	- 0,09	0,53	1,00
	P (kaksi- suuntainen)	0,00	0,15	0,00	0,00	0,00	
	N	4 109	4 088	4 109	4 109	4 088	4 109

LIITETAULUKKO 5a. Yhteiskuntaopin ja historian ratkaisuosuudet ja arvosanat sukupuolen mukaan luokiteltuna.

Sukupuoli		Yhteiskuntaopin ratkaisuosuus (%)	Yhteiskuntaopin arvosana	Historian ratkaisuosuus (%)	Historian arvosana
poika	Keskiarvo	63,73	7,63	51,10	7,75
	N	2 041	2 006	2 041	2 010
	Keskihajonta	13,40	1,21	14,41	1,26
tyttö	Keskiarvo	66,52	8,06	51,17	7,89
	N	2 117	2 097	2 117	2 099
	Keskihajonta	11,91	1,14	13,25	1,24
Sukupuolen selitysosuus (eetan neliö) yhteiskuntaopin ja historian ratkaisuosuuksissa ja arvosanoissa.					
Sukupuoli			Eeta	Eetan neliö	
Yhteiskuntaopin ratkaisuosuus (%)			0,11	0,01	
Yhteiskuntaopin arvosana			0,18	0,03	
Historian ratkaisuosuus (%)			0,00	0,00	
Historian arvosana			0,05	0,00	

LIITETAULUKKO 5b. Yhteiskuntaopin ja historian ratkaisuosuudet ja arvosanat koulun opetuskielen mukaan luokiteltuina.

Koulun opetuskieli		Yhteiskuntaopin ratkaisuosuus (%)	Yhteiskuntaopin arvosana	Historian ratkaisuosuus (%)	Historian arvosana
suomi	Keskiarvo	66,05	7,86	51,79	7,82
	N	3 613	3 562	3 613	3 568
	Keskihajonta	12,59	1,18	13,78	1,25
ruotsi	Keskiarvo	59,21	7,81	46,77	7,83
	N	545	541	545	541
	Keskihajonta	12,09	1,26	13,35	1,26
Koulun opetuskielen selitysosuus (eetan neliö) yhteiskuntaopin ja historian ratkaisuosuuksissa ja arvosanoissa.					
Koulun opetuskieli			Eeta	Eetan neliö	
Yhteiskuntaopin ratkaisuosuus (%)			0,18	0,03	
Yhteiskuntaopin arvosana			0,01	0,00	
Historian ratkaisuosuus (%)			0,12	0,02	
Historian arvosana			0,00	0,00	

LIITETAULUKKO 5c. Yhteiskuntaopin ja historian ratkaisuosuudet ja arvosanat vanhempien koulutustaustan mukaan luokiteltuina.

Onko äitisi/isäsi ylioppilas?		Yhteiskuntaopin ratkaisuosuus (%)	Yhteiskuntaopin arvosana	Historian ratkaisuosuus (%)	Historian arvosana
ei kumpikaan	Keskiarvo	62,52	7,50	47,31	7,40
	N	1 623	1 597	1 623	1 602
	Keskihajonta	12,43	1,21	12,72	1,23
toinen	Keskiarvo	65,51	7,94	51,49	7,93
	N	1 302	1 291	1 302	1 289
	Keskihajonta	12,72	1,15	13,31	1,17
molemmat	Keskiarvo	68,72	8,26	56,36	8,33
	N	1 136	1 131	1 136	1 134
	Keskihajonta	12,35	1,07	14,19	1,13
Vanhempien koulutustaustan selitysosuus (eetan neliö) yhteiskuntaopin ja historian ratkaisuosuuksissa ja arvosanoissa					
Vanhempien koulutustausta			Eeta	Eetan neliö	
Yhteiskuntaopin ratkaisuosuus (%)			0,20	0,04	
Yhteiskuntaopin arvosana			0,26	0,07	
Historian ratkaisuosuus (%)			0,27	0,07	
Historian arvosana			0,31	0,20	

LIITETAULUKKO 5d. Yhteiskuntaopin ja historian ratkaisuosuudet ja arvosanat ensisijaisen jatko-opintopaikan mukaan luokiteltuina.

Yhteishaussa pyrin		Yhteiskuntaopin ratkaisuosuus (%)	Yhteiskuntaopin arvosana	Historian ratkaisuosuus (%)	Historian arvosana
lukioon	Keskiarvo	69,70	8,36	56,41	8,38
	N	2 313	2 297	2 313	2 300
	Keskihajonta	11,28	0,97	12,97	1,03
ammattili- seen koulu- tukseen	Keskiarvo	59,62	7,20	44,59	7,10
	N	1 715	1 681	1 715	1 683
	Keskihajonta	12,12	1,11	11,81	1,12
hankin lisä- valmiuksia jatko-opin- toihin	Keskiarvo	53,10	6,64	39,25	6,68
	N	46	44	46	44
	Keskihajonta	13,39	1,46	14,17	1,29
aion mennä töihin tai pitää vapaa- vuoden	Keskiarvo	53,87	6,56	42,77	6,78
	N	9	9	9	9
	Keskihajonta	9,97	1,74	14,43	1,30
Jatko-opintosuuntautumisvaihtoehdon selitysosuus (eetan neliö) yhteiskuntaopin ja historian ratkaisuosuuksissa ja arvosanoissa.					
Esisijainen jatko-opintosuuntautumis- vaihtoehto			Eeta	Eetan neliö	
Yhteiskuntaopin ratkaisuosuus (%)			0,40	0,16	
Yhteiskuntaopin arvosana			0,49	0,24	
Historian ratkaisuosuus (%)			0,43	0,18	
Historian arvosana			0,51	0,26	

LIITETAULUKKO 5e. Yhteiskuntaopin ja historian ratkaisuosuudet ja arvosanat annettujen kotitehtävien tekemisen mukaan luokiteltuina.

Teen annetut kotitehtävät sovittulla tavalla		Yhteiskuntaopin ratkaisuosuus (%)	Yhteiskuntaopin arvosana	Historian ratkaisuosuus (%)	Historian arvosana
aina	Keskiarvo	68,81	8,35	55,37	8,31
	N	1 234	1 226	1 234	1 225
	Keskihajonta	11,92	1,05	13,44	1,14
usein	Keskiarvo	65,63	7,83	51,34	7,80
	N	1 915	1 886	1 915	1 888
	Keskihajonta	12,16	1,12	13,40	1,19
silloin tällöin	Keskiarvo	59,95	7,24	45,63	7,28
	N	758	745	758	748
	Keskihajonta	12,81	1,20	13,13	1,24
en juuri koskaan	Keskiarvo	59,16	7,27	45,44	7,23
	N	245	240	245	242
	Keskihajonta	13,60	1,27	13,45	1,30
Annettujen kotitehtävien tekemisen selitysosuus (eetan neliö) yhteiskuntaopin ja historian ratkaisuosuuksissa ja arvosanoissa.					
Annettujen kotitehtävien tekeminen			Eeta	Eetan neliö	
Yhteiskuntaopin ratkaisuosuus (%)			0,26	0,07	
Yhteiskuntaopin arvosana			0,34	0,11	
Historian ratkaisuosuus (%)			0,26	0,07	
Historian arvosana			0,31	0,09	

LIITETAULUKKO 5f. Yhteiskuntaopin ja historian ratkaisuosuudet ja arvosanat sen mukaan luokiteltuina, kuinka huoltajat seuraavat annettujen kotitehtävien tekemistä.

Huoltajat seuraavat kotitehtäviä		Yhteiskuntaopin ratkaisuosuus (%)	Yhteiskuntaopin arvosana	Historian ratkaisuosuus (%)	Historian arvosana
aina	Keskiarvo	64,42	8,21	51,17	8,26
	N	54	53	54	53
	Keskihajonta	13,75	1,26	16,01	1,21
usein	Keskiarvo	62,77	7,78	50,06	7,81
	N	316	311	316	312
	Keskihajonta	13,66	1,21	14,14	1,23
silloin tällöin	Keskiarvo	63,90	7,72	49,71	7,72
	N	1 371	1 355	1 371	1 354
	Keskihajonta	12,65	1,19	13,51	1,24
eivät juuri lainkaan	Keskiarvo	66,22	7,92	52,12	7,88
	N	2 408	2 375	2 408	2 381
	Keskihajonta	12,54	1,18	13,83	1,25
Huoltajien annettujen kotitehtävien seuraamisen selitysosuus (eetan neliö) yhteiskuntaopin ja historian ratkaisuosuuksissa ja arvosanoissa.					
Huoltajat seuraavat kotitehtävien tekemistä			Eeta	Eetan neliö	
Yhteiskuntaopin ratkaisuosuus (%)			0,10	0,01	
Yhteiskuntaopin arvosana			0,09	0,01	

LIITETAULUKKO 5g. Yhteiskuntaopin ja historian ratkaisuosuudet ja arvosanat sen mukaan luokiteltuina, kuinka kauan oppilas on harrastanut historiaan ja yhteiskuntaoppiin liittyviä asioita.

Olen harrastanut historiaan ja yhteiskuntaoppiin liittyviä asioita		Yhteiskuntaopin ratkaisuosuus (%)	Yhteiskuntaopin arvosana	Historian ratkaisuosuus (%)	Historian arvosana
en yhtiään	Keskiarvo	63,76	7,67	47,74	7,53
	N	1745	1725	1745	1729
	Keskihajonta	12,72	1,21	12,98	1,22
alle vuoden	Keskiarvo	63,29	7,74	48,50	7,66
	N	613	599	613	601
	Keskihajonta	12,27	1,17	12,65	1,20
1-2 vuotta	Keskiarvo	66,66	8,03	53,20	7,99
	N	640	636	640	635
	Keskihajonta	12,27	1,14	12,86	1,19
3-4 vuotta	Keskiarvo	67,31	8,12	55,45	8,23
	N	548	543	548	543
	Keskihajonta	13,01	1,15	14,25	1,21
yli 5 vuotta	Keskiarvo	67,82	8,05	57,88	8,30
	N	595	585	595	586
	Keskihajonta	12,68	1,18	14,17	1,18
Historian ja yhteiskuntaopin harrastuneisuus ja sen selitysosuus (eetan neliö) yhteiskuntaopin ja historian ratkaisuosuuksissa ja arvosanoissa.					
Harrastuneisuus			Eeta	Eetan neliö	
Yhteiskuntaopin ratkaisuosuus (%)			0,14	0,02	
Yhteiskuntaopin arvosana			0,16	0,03	
Historian ratkaisuosuus (%)			0,29	0,08	
Historian arvosana			0,25	0,06	

LIITETAULUKKO 5h. Yhteiskuntaopin ja historian ratkaisuosuudet ja arvosanat entisen lääninjaon mukaan luokiteltuina.

Lääni koodi		Yhteiskuntaopin ratkaisuosuus (%)	Yhteiskuntaopin arvosana	Historian ratkaisuosuus (%)	Historian arvosana
Etelä-Suomen lääni	Keskiarvo	65,64	7,88	52,09	7,88
	N	1 649	1 625	1 649	1 627
	Keskihajonta	12,82	1,23	14,02	1,26
Länsi-Suomen lääni	Keskiarvo	64,53	7,85	50,76	7,82
	N	1 469	1 459	1 469	1 459
	Keskihajonta	12,73	1,19	13,76	1,24
Itä-Suomen lääni	Keskiarvo	64,46	7,79	49,39	7,69
	N	561	551	561	556
	Keskihajonta	12,86	1,15	13,95	1,26
Oulun lääni	Keskiarvo	66,70	7,80	50,96	7,78
	N	348	340	348	339
	Keskihajonta	12,67	1,13	13,31	1,21
Lapin lääni	Keskiarvo	64,91	7,84	51,26	7,82
	N	131	128	131	128
	Keskihajonta	10,69	1,11	12,22	1,27
Entisen lääninjaon selitysosuus (eetan neliö) yhteiskuntaopin ja historian ratkaisuosuuksissa ja arvosanoissa.					
			Eeta	Eetan neliö	
Yhteiskuntaopin ratkaisuosuus (%)			0,06	0,00	
Yhteiskuntaopin arvosana			0,03	0,00	
Historian ratkaisuosuus (%)			0,07	0,00	
Historian arvosana			0,05	0,00	

LIITETAULUKKO 5i. Yhteiskuntaopin ja historian ratkaisuosuudet ja arvosanat tilastollisen kuntatyyppin mukaan luokiteltuina.

Kuntaryhmäkoodi		Yhteiskuntaopin ratkaisuosuus (%)	Yhteiskuntaopin arvosana	Historian ratkaisuosuus (%)	Historian arvosana
kaupun- kimaiset kunnat	Keskiarvo	65,22	7,90	51,45	7,88
	N	2 668	2 637	2 668	2 641
	Keskihajonta	12,66	1,19	13,66	1,24
taajaan asu- tut kunnat	Keskiarvo	63,15	7,93	49,53	7,86
	N	540	535	540	533
	Keskihajonta	13,06	1,25	14,08	1,26
maaseu- tumaiset kunnat	Keskiarvo	66,10	7,66	51,17	7,65
	N	950	931	950	935
	Keskihajonta	12,65	1,17	14,10	1,25
Tilastollisen kuntatyyppin selitysosuus (eetan neliö) yhteiskuntaopin ja historian ratkaisuosuuksissa ja arvosanoissa.					
			Eeta	Eetan neliö	
Yhteiskuntaopin ratkaisuosuus (%)			0,07	0,00	
Yhteiskuntaopin arvosana			0,09	0,01	
Historian ratkaisuosuus (%)			0,05	0,00	
Historian arvosana			0,07	0,01	

Post Hoc-testaus yhteiskuntaoppi.

LIITETAULUKKO 6a. Vanhempien koulutustausta ja yhteiskuntaopin ratkaisuosuus.

	Ryhmä	N			
Onko äitisi/isäsi ylioppilas?	ei kumpikaan	1 623			
	toinen	1 302			
	molemmat	1 136			
Riippuva muuttuja: Yhteiskuntaopin ratkaisuosuus (%)					
	III-tyyppin neliösumma	df	Keskineliö	F	P
Korjattu kokonais- varianssi	25852,29	2	12926,1	82,73	0,00
Vakio	17094581,53	1	17094581,53	109409,88	0,00
Vanhempien koulutus	25852,29	2	12926,15	82,73	0,00
Virhe	634036,08	4 058	156,24		
Yhteensä	17930089,53	4 061			
Yhteisvarianssi	659888,38	4 060			
Tukeyn testi (Ba, b,c)		N	Osajoukko		
Onko äitisi/isäsi ylioppilas?			1	2	3
ei kumpikaan		1 623	62,52		
toinen		1 302		65,51	
molemmat		1 136			68,72

LIITETAULUKKO 6b. Vanhempien koulutustausta ja yhteiskuntaopin arvosana.

	Ryhmä	N			
Onko äitisi/isäsi ylioppilas?	ei kumpikaan	1 597			
	toinen	1 291			
	molemmat	1 131			
Riippuva muuttuja: Yhteiskuntaopin arvosana					
	III-tyyppin neliösumma	df	Keskineliö	F	P
Korjattu kokonaisvarianssi	400,17	2	200,08	150,75	0,00
Vakio	245773,35	1	245773,35	185169,49	0,00
Vanhempien koulutus	400,16	2	200,08	150,75	0,00
Virhe	5330,39	4 016	1,33		
Yhteensä	253656,00	4 019			
Yhteisvarianssi	5730,56	4 018			
Tukeyn testi (Ba, b,c)		N	Osajoukko		
Onko äitisi/isäsi ylioppilas?			1	2	3
ei kumpikaan		1 597	7,50		
toinen		1 291		7,94	
molemmat		1 131			8,26

LIITETAULUKKO 6c. Ensisijainen jatko-opintosuuntautumismuutostie ja yhteiskuntaopin ratkaisuosuus.

	Ryhmä	N			
Yhteishaussa pyrin	lukioon	2 313			
	ammattilliseen koulutukseen	1 715			
	hankin lisävalmiuksia jatko-opintoihin	46			
	aion mennä töihin tai pitää vapaavuoden	9			
Riippuva muuttuja: Yhteiskuntaopin ratkaisuosuus (%)					
	III-tyyppin neliösumma	df	Keskineliö	F	P
Korjattu kokonaisvarianssi	108081,52	3	36027,17	264,77	0,00
Vakio	417099,56	1	417099,56	3065,32	0,00
Ensisijainen jatko-opintopaikka yhteishaussa	108081,52	3	36027,17	264,77	0,00
Virhe	555030,73	4 079	136,07		
Yhteensä	18045277,78	4 083			
Yhteisvarianssi	663112,25	4 082			
Tukeyn testi (Ba, b,c)			N	Osajoukko	
Yhteishaussa pyrin				1	2
hankin lisävalmiuksia jatko-opintoihin			46	53,20	
aion mennä töihin tai pitää vapaavuoden			9	53,87	
ammattilliseen koulutukseen			1 715	59,63	
lukioon			2 313		69,70

LIITETAULUKKO 6d. Ensisijainen jatko-opintovaihtoehto ja yhteiskuntaopin arvosana.

	Ryhmä	N			
Yhteishaussa pyrin	lukioon	2 297			
	ammattilliseen koulutukseen	1 681			
	hankin lisävalmiuksia jatko-opintoihin	44			
	aion mennä töihin tai pitää vapaavuoden	9			
Riippuva muuttuja: Yhteiskuntaopin arvosana					
	III-tyyppin neliösumma	df	Keskineliö	F	P
Korjattu kokonaisvarianssi	1395,71	3	465,24	431,20	0,00
Vakio	6127,91	1	6127,91	5690,07	0,00
Ensisijainen jatko-opintopaikka yhteishaussa	1395,71	3	465,24	431,20	0,00
Virhe	4336,87	4 027	1,08		
Yhteensä	254253,00	4 031			
Yhteisvarianssi	5732,58	4 030			
Tukeyn testi (Ba, b, c)		N		Osajoukko	
Yhteishaussa pyrin				1	2
aion mennä töihin tai pitää vapaavuoden		9		6,56	
hankin lisävalmiuksia jatko-opintoihin		44		6,64	
ammattilliseen koulutukseen		1 681		7,20	
lukioon		2 297			8,36

LIITETAULUKKO 6e. Annettujen kotitehtävien tekeminen ja yhteiskuntaopin ratkaisuosuus.

	Ryhmä		N		
Teen annetut kotitehtävät sovitulla tavalla	aina		1 234		
	usein		1 915		
	silloin tällöin		758		
	en juuri koskaan		245		
Riippuva muuttuja: Yhteiskuntaopin ratkaisuosuus (%)					
	III-tyyppin neliösumma	df	Keskineliö	F	P
Korjattu kokonaisvarianssi	46227,87	3	15409,29	101,89	0,00
Vakio	9547149,41	1	9547149,41	63129,92	0,00
Läksyt	46227,87	3	15409,29	101,89	0,00
Virhe	627302,84	4 148	151,23		
Yhteensä	18299582,19	4 152			
Yhteisvarianssi	673530,71	4 151			
Yhteiskuntaopin ratkaisuosuus (%)					
Tukeyn testi (Ba, b, c)					
		N	Osajoukko		
Teen annetut kotitehtävät sovitulla tavalla			1	2	3
en juuri koskaan		245	59,16		
silloin tällöin		758	59,95		
usein		1 915		65,63	
aina		1 234			68,81

LIITETAULUKKO 6f. Annettujen kotitehtävien tekeminen ja yhteiskuntaopin arvosana.

	Ryhmä	N			
Teen annetut kotitehtävät sovitulla tavalla	aina	1 226			
	usein	1 886			
	silloin tällöin	745			
	en juuri koskaan	240			
Riippuva muuttuja: Yhteiskuntaopin arvosana					
	III-tyyppin neliösumma	df	Keskineliö	F	P
Korjattu kokonaisvariassi	668,10	3	222,70	176,33	0,00
Vakio	137471,81	1	137471,81	108847,14	0,00
Annettujen kotitehtävien tekeminen	668,10	3	222,70	176,33	0,00
Virhe	5169,38	4 093	1,26		
Yhteensä	258235,00	4 097			
Yhteisvariassi	5837,48	4 096			
Tukeyn testi (Ba, b, c)					
	N	Osajoukko			
Teen annetut kotitehtävät sovitulla tavalla		1	2	3	
silloin tällöin	745	7,24			
en juuri koskaan	240	7,27			
usein	1 886		7,83		
aina	1 226			8,35	

LIITETAULUKOT 6g. Huoltajat seuraavat annettuja kotitehtäviä ja yhteiskuntaopin ratkaisuosuus.

	Ryhmä	N			
Huoltajat seuraavat kotitehtävien tekemistä	aina	54			
	usein	316			
	silloin tällöin	1 371			
	eivät juuri lainkaan	2 408			
Riippuva muuttuja: Yhteiskuntaopin ratkaisuosuus (%)					
	III-typin neliösumma	df	Keskineliö	F	P
Korjattu kokonaisvarianssi	6714,42	3	2238,14	13,92	0,00
Vakio	2900583,88	1	2900583,88	18042,92	0,00
Huoltajat seuraavat kotitehtävien tekemistä	6714,42	3	2238,14	13,92	0,00
Virhe	666351,23	4 145	160,76		
Yhteensä	18294281,45	4 149			
Yhteisvarianssi	673065,65	4 148			
Tukeyn testi (Ba, b, c)					
	N	Osajoukko			
Huoltajat seuraavat kotitehtävien tekemistä		1			
usein	316	62,77			
silloin tällöin	1 371	63,90			
aina	54	64,42			
eivät juuri lainkaan	2 408	66,22			

LIITETAULUKOT 6h. Huoltajat seuraavat annettuja kotitehtäviä ja yhteiskuntaopin arvosana.

	Ryhmä		N			
Huoltajat seuraavat kotitehtävien tekemistä	aina		53			
	usein		311			
	silloin tällöin		1 355			
	eivät juuri lainkaan		2 375			
Riippuva muuttuja: Yhteiskuntaopin arvosana						
	III-tyyppin neliösumma	df	Keskineliö	F	P	
Korjattu kokonaisvariassi	42,66	3	14,22	10,06	0,00	
Vakio	43068,20	1	43068,20	30462,14	0,00	
Huoltajat seuraavat kotitehtävien tekemistä	42,66	3	14,22	10,06	0,00	
Virhe	5782,55	4 090	1,41			
Yhteensä	258125,00	4 094				
Yhteisvariassi	5825,22	4 093				
Tukeyn testi (Ba, b, c)			N		Osajoukko	
Huoltajat seuraavat kotitehtäviä					1	2
siltoin tällöin			1 355		7,72	
usein			311		7,78	
eivät juuri lainkaan			2 375		7,92	7,92
aina			53			8,21

LIITETAULUKKO 6i. Historian ja yhteiskuntaopin aihealueiden harrastaminen ja yhteiskuntaopin ratkaisuosuus.

	Ryhmä	N			
Olen harrastanut historiaan ja yhteiskunta-oppiin liittyviä asioita	en yhtään	1 745			
	alle vuoden	613			
	1-2 vuotta	640			
	3-4 vuotta	548			
	yli 5 vuotta	595			
Riippuva muuttuja: Yhteiskuntaopin ratkaisuosuus (%)					
	III-tyyppin neliösumma	df	Keskineliö	F	P
Korjattu kokonaisvarianssi	13767,02	4	3441,75	21,62	0,00
Vakio	14869284,26	1	14869284,26	93411,11	0,00
Harrastuneisuus	13767,02	4	3441,75	21,62	0,00
Virhe	658373,10	4 136	159,18		
Yhteensä	18270656,57	4 141			
Yhteisvarianssi	672140,12	4 140			
Tukeyn testi (Ba, b, c)			Osajoukko		
Olen harrastanut historiaan ja yhteiskuntaoppiin liittyviä asioita	N		1	2	
alle vuoden	613		63,29		
en yhtään	1 745		63,76		
1-2 vuotta	640			66,66	
3-4 vuotta	548			67,31	
yli 5 vuotta	595			67,82	

LIITETAULUKKO 6j. Historian ja yhteiskuntaopin aihealueiden harrastaminen ja yhteiskuntaopin arvosana.

		Ryhmä	N		
Olen harrastanut historiaan ja yhteiskuntaoppiin liittyviä asioita		en yhtään	1 725		
		alle vuoden	599		
		1-2 vuotta	636		
		3-4 vuotta	543		
		yli 5 vuotta	585		
Riippuva muuttuja: Yhteiskuntaopin arvosana					
	III-typin neliösumma	df	Keskineliö	F	P
Korjattu kokonaisvariassi	146,81	4	36,70	26,40	0,00
Vakio	212892,40	1	212892,40	153141,43	0,00
Harrastuneisuus	146,81	4	36,70	26,40	0,00
Virhe	5676,06	4 083	1,39		
Yhteensä	257833,00	4 088			
Yhteisvariassi	5822,87	4 087			
Tukeyn testi (Ba, b, c)		N		Osajoukko	
Olen harrastanut historiaan ja yhteiskuntaoppiin liittyviä asioita				1	2
en yhtään		1 725		7,67	
alle vuoden		599		7,74	
1-2 vuotta		636			8,03
3-4 vuotta		543			8,12
yli 5 vuotta		585			8,05

Post Hoc-testaus historia.

LIITETAULUKKO 7a. Vanhempien koulutustausta ja historian ratkaisuosuus.

	Ryhmä	N			
Onko äitisi/isäsi ylioppilas?	ei kumpikaan	1 623			
	toinen	1 302			
	molemmat	1 136			
Riippuva muuttuja: Historian ratkaisuosuus (%)					
	III-tyyppin neliösumma	df	Keskineliö	F	P
Korjattu kokonaisvarianssi	54917,16	2	27458,58	154,47	0,00
Vakio	10632579,29	1	10632579,29	59814,85	0,00
Vanhempien koulutustausta	54917,16	2	27458,58	154,47	0,00
Virhe	721342,73	4 058	177,76		
Yhteensä	11415485,08	4 061			
Kokonaisvarianssi	776259,89	4 060			
Tukey (Ba, b, c)					
Onko äitisi/isäsi ylioppilas?		Osajoukko			
	N	1	2	3	
ei kumpikaan	1 623	47,31			
toinen	1302		51,49		
molemmat	1136			56,36	

LIITETAULUKKO 7b. Vanhempien koulutustausta ja historian arvosana.

	Ryhmä	N			
Onko äitisi/isäsi ylioppilas?	ei kumpikaan	1 602			
	toinen	1 289			
	molemmat	1 134			
Historian arvosana					
	III-typin neliösumma	df	Keskineliö	F	P
Korjattu kokonaisvarianssi	590,258	2	295,13	210,11	0,00
Vakio	245167,47	1	245167,47	174540,52	0,00
Vanhempien koulutustausta	590,26	2	295,13	210,11	0,00
Virhe	5649,48	4 022	1,41		
Yhteensä	252965,00	4 025			
Yhteisvarianssi	6239,74	4 024			
Tukeyn testi (Ba, b, c)					
Onko äitisi/isäsi ylioppilas?	Osajoukko				
Ryhmä	N	1	2	3	
ei kumpikaan	1 602	7,40			
toinen	1 289		7,93		
molemmat	1 134			8,33	

LIITETAULUKKO 7c. Ensisijainen jatko-opintosuuntautumismuutokset ja historian ratkaisu-
osuus.

	Ryhmä	N			
Yhteishaussa pyrin	lukioon	2 313			
	ammattilliseen koulutukseen	1 715			
	hankin lisävalmiuksia jatko-opintoihin	46			
	aion mennä töihin tai pitää vapaavuoden	9			
Riippuva muuttuja: Historian ratkaisuosuus (%)					
	III-tyyppin neliösumma	df	Keskineliö	F	P
Korjattu kokonaisvarianssi	145034,95	3	48344,99	308,62	0,00
Vakio	250233,42	1	250233,42	1597,39	0,00
Ensisijainen jatko-opintopaikka	145034,95	3	48344,99	308,62	0,00
Virhe	638981,77	4 079	156,65		
Yhteensä	11497429,16	4 083			
Yhteisvarianssi	784016,72	4 082			
Tukeyn testi (Ba, b, c)		N	Osajoukko		
Yhteishaussa pyrin		1	2		
hankin lisävalmiuksia jatko-opintoihin		46	39,25		
aion mennä töihin tai pitää vapaavuoden		9	42,77		
ammattilliseen koulutukseen		1 715	44,59		
lukioon		2 313	56,41		

LIITETAULUKKO 7d. Ensisijainen jatko-opintosuuntautumisvaihtoehto ja historian arvosana.

	Ryhmä	N			
Yhteishaussa pyrin	lukioon	2 300			
	ammattilliseen koulutukseen	1 683			
	hankin lisävalmiuksia jatko-opintoihin	44			
	aion mennä töihin tai pitää vapaavuoden	9			
Riippuva muuttuja: Historian arvosana					
	III-tyyppin neliösumma	df	Keskineliö	F	P
Korjattu kokonaisvarianssi	1668,34	3	556,11	483,57	0,00
Vakio	6211,48	1	6211,48	5401,17	0,00
Ensisijainen jatko-opintopaikka	1668,34	3	556,11	483,57	0,00
Virhe	4636,90	4 032	1,15		
Yhteensä	253515,00	4 036			
Yhteisvarianssi	6305,25	4 035			
Tukeyn testi (Ba, b, c)		N	Osajoukko		
Yhteishaussa pyrin			1	2	
hankin lisävalmiuksia jatko-opintoihin		44	6,68		
aion mennä töihin tai pitää vapaavuoden		9	6,78		
ammattilliseen koulutukseen		1 683	7,10		
lukioon		2 300		8,38	

LIITETAULUKKO 7e. Annettujen kotitehtävien tekeminen ja historian ratkaisuosuus.

	Ryhmä	N			
Teen annetut kotitehtävät sovitulla tavalla	aina	1 234			
	usein	1 915			
	silloin tällöin	758			
	en juuri koskaan	245			
Riippuva muuttuja: Historian ratkaisuosuus (%)					
	III-tyyppin neliösumma	df	Keskineliö	F	P
Korjattu kokonaisvarianssi	53119,88	3	17706,63	99,11	0,00
Vakio	5809603,72	1	5809603,72	32519,01	0,00
Läksyt	53119,88	3	17706,63	99,11	0,00
Virhe	741050,81	4 148	178,65		
Yhteensä	11656496,53	4 152			
Yhteisvarianssi	794170,69	4 151			
Tukeyn testi (Ba, b, c)			Osajoukko		
Teen annetut kotitehtävät sovitulla tavalla	N	1	2	3	
en juuri koskaan	245	45,44			
silloin tällöin	758	45,63			
usein	1 915		51,34		
aina	1234			55,37	

LIITETAULUKKO 7f. Annettujen kotitehtävien tekeminen ja historian arvosana.

	Ryhmä	N			
Teen annetut kotitehtävät sovitulla tavalla	aina	1 225			
	usein	1 888			
	silloin tällöin	748			
	en juuri koskaan	242			
Riippuva muuttuja: Historian arvosana					
	III-typin neliösumma	df	Keskineliö	F	P
Korjattu kokonaisvariassi	602,86	3	200,95	142,16	0,00
Vakio	137549,90	1	137549,90	97309,91	0,00
Läksyt	602,86	3	200,95	142,16	0,00
Virhe	5794,04	4 099	1,41		
Yhteensä	257517,00	4 103			
Yhteisvariassi	6396,89	4 102			
Tukeyn testi (Ba, b, c)			Osajoukko		
Teen annetut kotitehtävät sovitulla tavalla	N	1	2	3	
en juuri koskaan	242	7,23			
silloin tällöin	748	7,28			
usein	1 888		7,80		
aina	1 225			8,31	

LIITETAULUKKO 7g. Huoltajat seuraavat annettujen kotitehtävien tekemistä ja historian ratkaisuosuus.

	Ryhmä	N			
Huoltajat seuraavat kotitehtäviä	aina	54			
	usein	316			
	silloin tällöin	1 371			
	eivät juuri lainkaan	2 408			
Riippuva muuttuja: Historian ratkaisuosuus (%)					
	III-typin neliösumma	df	Keskineliö	F	P
Korjattu kokonaisvariassi	5491,20	3	1830,40	9,64	0,00
Vakio	1806237,86	1	1806237,86	9509,90	0,00
Huoltajat seuraavat läksyjien tekemistä	5491,20	3	1830,40	9,64	0,00
Virhe	787269,54	4 145	189,93		
Yhteensä	11650463,50	4 149			
Yhteisvariassi	792760,74	4 148			
Tukeyn testi (Ba, b, c)				Osajoukko	
Huoltajat seuraavat kotitehtäviä			N	1	
silloin tällöin			1 3071	49,71	
usein			316	50,06	
aina			54	51,17	
eivät juuri lainkaan			2 408	52,12	

LIITETAULUKKO 7h. Huoltajat seuraavat annettujen kotitehtävien tekemistä ja historian arvosana.

	Ryhmä	N			
Huoltajat seuraavat kotitehtävien tekemistä	aina	53			
	usein	312			
	silloin tällöin	1 354			
	eivät juuri lainkaan	2 381			
Riippuva muuttuja: Historian arvosana					
	III-typin neliösumma	df	Keskineliö	F	P
Korjattu kokonaisvarianssi	33,68	3	11,23	7,26	0,00
Vakio	43170,65	1	43170,65	27908,37	0,00
Huoltajat seuraavat kotitehtävien tekemistä	33,68	3	11,23	7,26	0,00
Virhe	6335,99	4 096	1,55		
Yhteensä	257470,00	4 100			
Yhteisvarianssi	6369,66	4 099			
Tukeyn testi (Ba, b, c)			N	Osajoukko	
Huoltajat seuraavat kotitehtäviä				1	2
silloin tällöin			1 354	7,72	
usein			312	7,81	
eivät juuri lainkaan			2 381	7,88	
aina			53		8,26

LIITETAULUKKO 7i. Historian ja yhteiskuntaopin aihealueiden harrastaminen ja historian ratkaisuosuus.

	Ryhmä	N			
Olen harrastanut historiaan ja yhteiskuntaoppiin liittyviä asioita	en yhtään	1 745			
	alle vuoden	613			
	1-2 vuotta	640			
	3-4 vuotta	548			
	yli 5 vuotta	595			
Riippuva muuttuja: Historian ratkaisuosuus (%)					
	III-tyyppin neliösumma	df	Keskineliö	F	P
Korjattu kokonaisvarianssi	64382,18	4	16095,54	91,44	0,00
Vakio	9495199,38	1	9495199,38	53945,37	0,00
Harrastus	64382,18	4	16095,54	91,44	0,00
Virhe	727998,41	4 136	176,02		
Yhteensä	11637243,39	4 141			
Yhteisvarianssi	792380,59	4 140			
Tukeyn testi (Ba, b, c)					
Olen harrastanut historiaan ja yhteiskuntaoppiin liittyviä asioita	N	Osajoukko			
		1	2	3	4
en yhtään	1 745	47,74			
alle vuoden	613	48,50			
1-2 vuotta	640		53,20		
3-4 vuotta	548			55,45	
yli 5 vuotta	595				57,88

LIITETAULUKKO 7j. Historian ja yhteiskuntaopin aihealueiden harrastaminen ja historian arvosana.

	Ryhmä	N			
Olen harrastanut historiaan ja yhteiskuntaoppiin liittyviä asioita	en yhtään	1 729			
	alle vuoden	601			
	1-2 vuotta	635			
	3-4 vuotta	543			
	yli 5 vuotta	586			
Riippuva muuttuja: Historian arvosana					
	III-tyyppin neliösumma	df	Keskineliö	F	P
Korjattu kokonaisvarianssi	407,16	4	101,79	69,79	0,00
Vakio	214222,50	1	214222,50	146882,07	0,00
Harrastuneisuus	407,16	4	101,79	69,79	0,00
Virhe	5963,67	4 089	1,46		
Yhteensä	257103,00	4 094			
Yhteisvarianssi	6370,83	4 093			
Tukeyn testi (Ba, b, c)			Osajoukko		
Olen harrastanut historiaan ja yhteiskuntaoppiin liittyviä asioita	N	1	2	3	
en yhtään	1 729	7,53			
alle vuoden	601	7,66			
1-2 vuotta	635		7,99		
3-4 vuotta	543			8,23	
yli 5 vuotta	586			8,30	

LIITETAULUKKO 8a. Koulujen väliset erot yhteiskuntaopin ratkaisuosuuksista ja arvosanoista laskettuna.

Riippuva muuttuja: Yhteiskuntaopin ratkaisuosuus		
Varianssi estimaatti		
Komponentti	Estimaatti	
Var(KOULU)	22,51	
Var(Virhe)	139,80	13,9 %
Menetelmä: Maksimum Likelihood Estimation		
Riippuva muuttuja: Yhteiskuntaopin arvosana		
Varianssi estimaatti		
Komponentti	Estimaatti	
Var(KOULU)	0,10	
Var(Virhe)	1,33	6,7 %
Menetelmä: Maksimum Likelihood Estimation		

LIITETAULUKKO 8b. Koulujen väliset erot historian ratkaisuosuuksista ja arvosanoista laskettuna.

Riippuva muuttuja: Historian arvosana		
Varianssi estimaatti		
Komponentti	Estimaatti	
Var(KOULU)	0,08	
Var(Virhe)	1,48	5,1 %
Menetelmä: Maksimi Likelihood Estimation		
Riippuva muuttuja: Historian ratkaisuosuus		
Varianssi estimaatti		
Komponentti	Estimate	
Var(KOULU)	24,00	
Var(Virhe)	167,68	12,5 %
Menetelmä: Maksimum Likelihood Estimation		

LIITETAULUKKO 9. Yhteiskuntaopin ja historian keskimääräisten ratkaisuosuuksien, arvosanojen, niistä laskettujen odotusarvojen ylärajan ylitysten ja alistusten sekä läksyjen tekemisen koulutason korrelaatiot.

	Yhteiskuntaopin ratkaisu-osuus (%)		Yhteiskuntaopin arvosana	Historian ratkaisu-osuus (%)	Historian arvosana	Ero odotusarvoihin yhteiskuntaoppi	Ero odotusarvoihin historia	Läksyjen tekeminen
Yhteiskuntaopin ratkaisu-osuus (%)	r_{xy}	1	0,55	0,74	0,55	0,80	0,46	0,26
	P (kaksisuuntainen)		0,00	0,00	0,00	0,00	0,00	0,00
	N	4158	4103	4158	4109	4103	4109	4152
Yhteiskuntaopin arvosana	r_{xy}	0,55	1	0,53	0,76	-0,03	0,02	0,34
	P (kaksisuuntainen)	0,00		0,00	0,00	0,10	0,15	0,00
	N	4103	4103	4103	4088	4103	4088	4097
Historian ratkaisu-osuus (%)	r_{xy}	0,74	0,53	1	0,60	0,51	0,71	0,25
	P (kaksisuuntainen)	0,00	0,00		0,00	0,00	0,00	0,00
	N	4158	4103	4158	4109	4103	4109	4152
Historian arvosana	r_{xy}	0,55	0,76	0,60	1	0,13	-0,09	0,30
	P (kaksisuuntainen)	0,00	0,00	0,00		0,00	0,00	0,00
	N	4109	4088	4109	4109	4088	4109	4103
Ero odotusarvoihin yhteiskuntaoppi	r_{xy}	0,80	-0,03	0,51	0,13	1	0,54	0,08
	P (kaksisuuntainen)	0,00	0,10	0,00	0,00		0,00	0,00
	N	4103	4103	4103	4088	4103	4088	4097
Ero odotusarvoihin historia	r_{xy}	0,46	0,02	0,71	-0,09	0,54	1	0,06
	P (kaksisuuntainen)	0,00	0,15	0,00	0,00	0,00		0,00
	N	4109	4088	4109	4109	4088	4109	4103
Läksyjen tekeminen	r_{xy}	0,26	0,33	0,25	0,30	0,08	0,06	1
	P (kaksisuuntainen)	0,00	0,00	0,00	0,00	0,00	0,00	
	N	4 152	4 097	4 152	4 103	4 097	4 103	4 152

Painettu
ISBN 978-952-13-5579-0
ISSN 1798-8918

Verkkojulkaisu
ISBN 978-952-13-5580-6
ISSN 1798-8926

Tässä väitöstutkimuksessa analysoidaan oppilaan arvioinnissa ilmenneitä koulujen välisiä eroja suhteessa oppilaiden arvosanoihin ja Opetushallituksen oppimistulosten seuranta-arvioinnissa osoittamaan osaamiseen. Lisäksi tutkimuksessa tarkastellaan opettajien antamien arvosanojen luotettavuutta osaamisen mittarina.

Tutkimuksessa pyritään vastaamaan seuraaviin kysymyksiin:

- Miten oppimistulosten seuranta-arviointi ja arvosanat vastaavat toisiaan?
- Esiintyykö kahden arvioinnin välillä mahdollisia systemaattisia eroja ja jos esiintyy, miten ne ilmenevät?
- Voidaanko seuranta-arvioinnin perusteella tehdä johtopäätöksiä arvosanojen validiudesta, vai jääkö kahden mittauksen validiusongelma ratkaisematta?
- Voidaanko kahden oppiaineen samanaikaisella arvioinnilla ratkaista kahden mittauksen eli arvosanojen ja seuranta-arvioinnin validiutta koskeva ongelma?

Lopuksi pohditaan vielä sitä, minkälaisia johtopäätöksiä voi tutkimusaineiston perusteella tehdä oppilaiden arvosanoista, arvosanojen antamista ohjaavista normeista ja arvosanojen antamisen käytänteistä.