

Hannu Karhunen

TYÖPAPEREITA / WORKING PAPERS

Mitä tiedetään kansallisten kokeiden vaikutuksista?

362



TYÖN JA TALOUDEN
TUTKIMUS
EST 1971
Labore

Tämä tutkimus on osa EDUCA-lippulaivaa, jota rahoittaa Suomen Akatemia (rahoituspäätökset #358924 ja #358946).

Hannu Karhunen

Tutkimusohjaaja (dos.), Työn ja talouden tutkimuslaitos LABORE
Research Fellow, Aalto yliopisto, EDUCA Lippulaiva
hannu.karhunen@labore.fi

Helsinki, 2026

TIIVISTELMÄ

Suomi on Pohjoismaissa ja kansainvälisesti poikkeus: peruskoulussa ei käytetä standardeitua kansallisia kokeita. Tarkastelen tässä katsauksessa, mitä empiirinen tutkimuskirjallisuus kertoo kansallisten kokeiden vaikutuksista oppimistuloksiin, oppilaiden hyvinvointiin sekä opettajiin ja opetukseen, ja kuinka hyvin tutkimusnäyttö on yleistettävissä Suomeen.

Katsaus perustuu 183 vuosina 2010–2025 julkaistuihin tutkimukseen. Aineiston ytimenä on Ketosen ja Atjosen (2025) systemaattinen katsaus, jota täydentää uusi taloustieteellisiin julkaisuihin kohdistettu haku. Tutkimukset luokitellaan päävastemuuttujan ja tutkimusasetelman vahvuuden mukaan, ja tulokinnassa painottuvat kausaalipäätelmiin pyrkivät kokeelliset ja kvassikokeelliset asetelmat. Tulosten perusteella kansalliset kokeet ja niihin liittyvät kannustin- ja informaatiomekanismit parantavat useissa tutkimuksissa oppimistuloksia ja joissakin tapauksissa myös myöhempiä koulutus- ja työmarkkinatulemia.

Vaikutukset ovat usein suurimpia heikommassa asemassa olevilla oppilailla, mikä tukee tulkin-taa kokeiden mahdollisuuksista vahvistaa arvi-oinnin yhdenvertaisuutta. Samalla kokeet lisää-

vät oppilaiden lyhytaikaista stressiä erityisesti koetilanteen läheisyydessä; vahvaa näyttöä pit-käaikaisista hyvinvointihaitoista ei sen sijaan löydy. Huonosti suunniteltu järjestelmä voi kuitenkin luoda kouluille kannustimia strategiseen käyttäytymiseen, kuten heikkojen oppilaiden va-likointiin pois kokeista.

Katsauksen kokonaiskuva poikkeaa Ketosen ja Atjosen (2025) tulkinnoista: kansallisilla kokeil-la on tutkimusnäytön valossa saavutettavissa merkittäviä hyötyjä oppimistulosten ja yhden-vertaisuuden näkökulmasta. Mahdollinen käyt-töönotto Suomessa edellyttäisi muiden maiden kokemusten huolellista tarkastelua sekä järjes-telmän tietoista suunnittelua.

JEL-koodit: I21, I24, I28

Avainsanat:

kansalliset kokeet
oppimistulokset
oppilasarviointi
koulutuksellinen yhdenvertaisuus
oppilaiden hyvinvointi

ABSTRACT

What Do We Know About the Effects of National Tests?

Finland is an exception both within the Nordic countries and internationally: standardized national tests are not used in basic education. In this review, I examine what the empirical research literature tells us about the effects of national tests on learning outcomes, student well-being, and teachers and instruction, and how well this evidence generalizes to the Finnish context.

The review draws on 183 studies published between 2010 and 2025. At its core is the systematic review by Ketonen and Atjonen (2025), complemented by a new search targeting economics publications. The studies are classified according to their primary outcome and the strength of their research design, with the interpretation weighted toward experimental and quasi-experimental designs that support causal inference. The findings indicate that national tests, and the incentive and information mechanisms associated with them, improve learning outcomes in many studies and in some cases also later educational and labour market outcomes.

The effects are often largest for students in less advantaged positions, which supports the interpretation that national tests can strengthen equity in assessment. At the same time, the tests increase students' short-term stress, particularly close to the time of testing; strong evidence of long-term harm to well-being is, by contrast, lacking. A poorly designed system can, however, create incentives for strategic behaviour by schools, such as selecting low-performing students out of testing.

The overall picture from this review departs from the interpretations of Ketonen and Atjonen (2025): Considering the research evidence, national tests can deliver substantial benefits in terms of learning outcomes and equity. Their potential introduction in Finland would require a careful examination of international experience and deliberate system design.

JEL Codes: I21, I24, I28

Keywords:

national testing
learning outcomes
teacher assessment
educational equality
student well-being

JOHDANTO

Suomessa on viime aikoina käyty vilkasta keskustelua peruskoulun kansallisista kokeista (Lehikoinen ym. 2026; Karhunen ym. 2026; Ketonen ja Atjonen 2025). Toisin kuin muissa Pohjoismaissa ja laajasti kansainvälisesti, Suomessa ei käytetä kansallisia standardoituja kokeita peruskoulun eri vaiheissa (ks. Wallenius 2020). Lähimaat eivät ole luopumassa kansallisista koejärjestelmistään, vaan kehittävät niitä aktiivisesti esimerkiksi Tanskassa ja Ruotsissa. Siksi on perusteltua kysyä, mitä tutkimus kertoo kansallisten kokeiden vaikutuksista.

Kansallisia standardoituja kokeita on perusteltu sillä, että opettajien arviot eivät aina vastaa oppilaiden todellista osaamistasoa, vaan voivat heijastaa esimerkiksi oppilaiden taustaa tai sukupuolta. Kokeiden avulla pyritään seuraamaan ja vahvistamaan arvosanojen yhdenvertaisuutta sekä turvaamaan oppilaiden oikeusturvaa erityisesti jatko-opintoihin haettaessa (Tveit 2019, s. 13).

Huoli meritokratian toteutumisesta näkyy myös eri maiden tarkastusvirastojen raporteissa. Esimerkiksi Ruotsin tarkastusvirasto arvioi vuoden 2018 lakiuudistuksen riittämättömäksi: vaikka opettajien tuli huomioida kansallisten kokeiden tulokset arvosanoja antaessaan, uudistus ei lisännyt arvioinnin yhdenvertaisuutta merkittävästi (Riksrevisionen 2022, s. 48). Ruotsi on sittemmin vienyt parlamentin käsiteltäväksi lakiesityksen peruskoulun päättökokeesta (slutprov), mikä vahvistaisi kansallisten kokeiden asemaa päättöarvioinnissa.

Tarkastelen tässä katsauksessa, mitä empiirinen tutkimus kertoo kansallisten kokeiden vaikutuksista. Koulutuspoliittinen päätöksenteko tarvitsee luotettavaa tietoa, koska mahdolliset ratkaisut koskevat lapsia ja nuoria. Erotan laadulliset ja kuvailevat tutkimukset niistä, jotka arvioivat syy-seuraussuhteita vakiintuneilla empiirisillä menetelmillä. Kuvaan myös tutkimusnäytön sovellettavuutta Suomeen, sillä esimerkiksi Aasiassa havaitut ääri-ilmiöt eivät välttämättä yleisty suomalaiseseen kontekstiin.

Tarkastelun lähtökohtana on Ketosen ja Atjosen (2025) katsaus, jota on hyödynnetty sekä edunvalvonnassa¹ että eduskunnalle annetuissa asiantuntijalausunnoissa. Katsaukseen ovat vedonneet esimerkiksi Kansallisen koulutuksen arviointikeskus (Karvi)² ja opetus- ja kulttuuriministeriö³. Koska katsauksen johtopäätöksillä on ollut näkyvä rooli kansallisia kokeita koskevassa koulutuspoliittisessa keskustelussa, on perusteltua arvioida, kuinka vahvasti alkuperäinen tutkimusnäyttö tukee niistä tehtyjä johtopäätöksiä. Täydennän katsausta tuoreemmalla taloustieteellisellä tutkimuskirjallisuudella ja kiinnitän erityistä huomiota siihen,

¹ <https://www.opettaja.fi/tyossa/paivi-atjonen-ei-juuri-mitaan-hyvaa/>

² <https://www.eduskunta.fi/fi/asiat-ja-aanestykset/valtiopaivaasiat/asiakirjat/edktunnus/EDK-2026-AK-12491/pdf>

³ <https://www.eduskunta.fi/fi/asiat-ja-aanestykset/valtiopaivaasiat/asiakirjat/edktunnus/EDK-2026-AK-12606/pdf>

missä määrin tutkimusasetelmat mahdollistavat luotettavat kausaalipäätelmät kansallisten kokeiden vaikutuksista.

TEOREETTINEN VIITEKEHYS

Standardoiduilla kansallisilla kokeilla mitataan eri ikäisten oppijoiden osaamista yhteismitallisesti eri oppiaineissa. Niiden keskeinen tehtävä on tuottaa luotettavaa ja vertailukelpoista tietoa oppijoille, huoltajille, opettajille, koulutuksen järjestäjille ja viranomaisille. Kun kokeiden tuloksia hyödynnetään esimerkiksi opiskelijavalinnoissa, puhutaan korkean panoksen (high stakes) kokeista, kuten ylioppilaskokeesta. Sen sijaan PISA ja kansalliset otospohjaiset arvioinnit ovat matalan panoksen (low stakes) kokeita, koska niitä ei käytetä yksilötason valinnoissa (Kilpi ym. 2025).

Kokeiden tuottama informaatio vaikuttaa yksilöiden kannustimiin ja päätöksentekoon. Erityisesti korkean panoksen kokeet voivat lisätä opiskeluun käytettyä aikaa ja vaivaa pitkällä aikavälillä, mikä tekee niiden vaikutusten mittaamisesta haastavaa (Bishop 1998).

Kokeet tuottavat myös tietoa opettajien ja koulujen toiminnan onnistumisesta. Tämän vuoksi niitä käytetään monissa maissa ohjaus- ja arviointivälineenä. Taloustieteessä tätä tarkastellaan päämies-agentti -mallien kautta: oikein suunniteltu tulosvastuu ja kannustimet voivat yhdenmukaistaa eri toimijoiden tavoitteita (Lovenheim ja Turner 2018; Figlio ja Loeb 2011; Holmström ja Milgrom 1991). Liian kapea-alaiset mittarit voivat kuitenkin kaventaa opetusta ("teaching to the test") tai kannustaa tulosten manipulointiin (Deming ja Figlio 2016).

Lazearin (2006) mukaan kokeeseen opettaminen ei kuitenkaan ole yksiselitteisesti haitallista. Hänen teoreettisen mallinsa mukaan tarkasti rajattu ja ennakoitava koesisältö voi lisätä oppimista erityisesti oppilailla, joille oppiminen on lähtökohtaisesti vaikeampaa, mikä käytännössä voi tarkoittaa esimerkiksi heikommin pärjääviä oppilaita tai kouluja. Sen sijaan vahvemmille oppijoille laajempaan tietopohjaan perustuva ja vaikeammin ennakoitava koe luo paremmat kannustimet laaja-alaiseen oppimiseen.

Kansallisilla kokeilla on sekä summatiivinen että formatiivinen rooli. Summatiivisesti ne tuottavat tietoa siitä, kuinka hyvin oppilas on saavuttanut ennalta määrätyn osaamistason mahdollistaen samalla myös koulutusjärjestelmän arvioinnin. Formatiivisesti niiden tuloksia voidaan hyödyntää opetuksen kehittämisessä ja oppilaiden tukemisessa jokapäiväisessä koulutyössä (Kilpi ym. 2025; Tveit 2019).

Eri maat hyödyntävät kansallisia kokeita eri tavoin oppilasarvioinnissa ja opiskelijavalinnoissa. Ruotsissa kokeet tukevat opettajien arvosananantoa, kun taas Norjassa ne on erotettu oppilasarvioinnista ja niitä käytetään pääasiassa järjestelmätason seurantaan. Norjassa toisen asteen valinta perustuu kuitenkin päättötodistukseen, joka yhdistää vuosiarvosanat ja valtakunnallisten päättökokeiden tulokset.

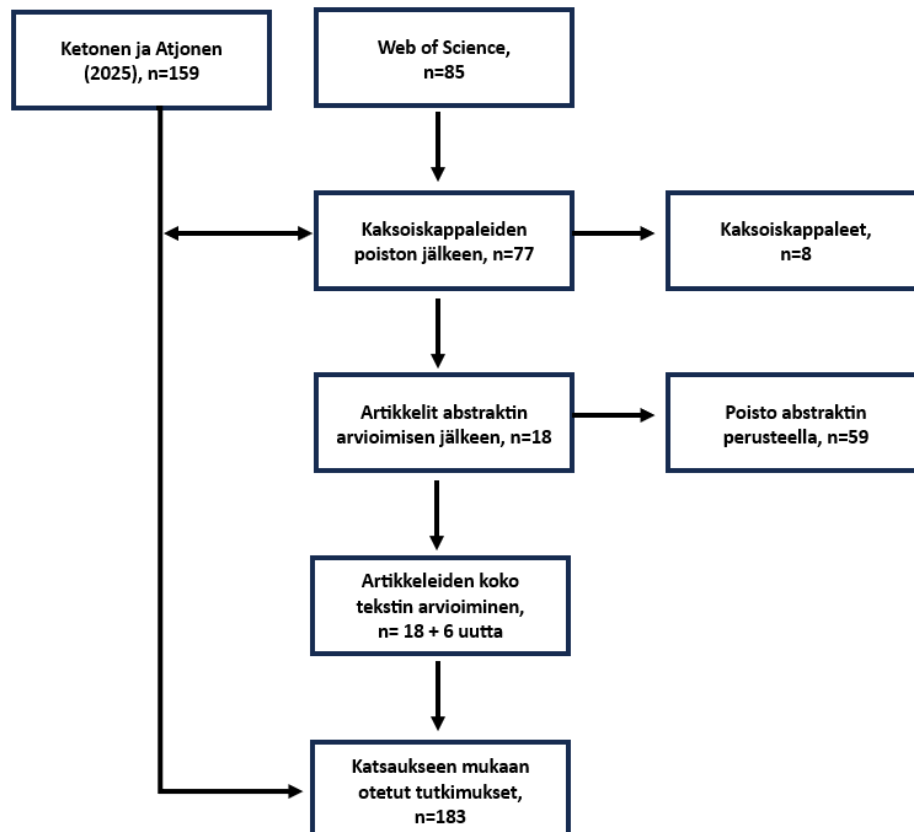
ANALYYSIN TOTEUTTAMINEN

Kirjallisuuskatsaus keskittyy siihen, missä määrin eri tutkimusten tutkimusasetelmat mahdollistavat syy-seuraussuhdetta koskevat kausaalipäätelmät ja kuinka hyvin tulokset ovat yleistettävissä Suomeen.

Tarkastelun lähtökohdaksi on otettu Ketonen ja Atjonen (2025), jotka tarkastelevat empiirisiä tutkimuksia kansallisista kokeista perusopetus- ja lukioikäisillä. Haku kattaa välillä 1.1.2010–1.1.2025 julkaistut tutkimukset, joita haettiin tietokannoista Web of Science, ERIC ja ABI/INFORM Collection. Hakutermit on esitetty alkuperäisjulkaisussa ja sen liitteessä sekä tämän katsauksen liitteessä täydentävän haun osalta.

Tarkasteluun on otettu mukaan kaikki Ketosen ja Atjosen valikoimat 159 artikkelia. Tätä artikkelia varten tehtiin Web of Science -tietokannasta täydennetty haku samoilla kriteereillä, rajattuna taloustieteellisiin julkaisuihin. Haku tuotti 85 tutkimusta, joista 8 oli päällekkäisiä aikaisemman aineiston kanssa. Jäljelle jääneistä 18 luettiin kokonaan abstraktikarsinnan jälkeen. Kokotekstilukemisen perusteella tunnistettiin viitetiedoista vielä kuusi hakuehdot täyttävää tutkimusta, jotka sisällytettiin katsaukseen.

Seuraavan katsauksen aineisto koostuu yhteensä 183 artikkelista, jotka luettiin kokonaisuudessaan ja luokiteltiin katsausta varten. Näistä tutkimuksista 159 on mukana Ketosen ja Atjosen (2025) katsauksessa ja niitä täydentämään löydettiin 24 muuta tutkimusta samalta aikaväliltä. Kuvio 1 kuvaa hakuprosessin vuokaavion. Tutkimuksen liitteessä A kuvataan tarkemmin käytetyt hakuehdot, ja kaikki haussa ilmi tulleet artikkelit.



Kuvio 1. Hakuprosessin vuokaavio

TULOKSET

Aineiston jäsentämiseksi luokittelen tutkimuskirjallisuuden kolmeen teemaan päävastemuuttujan mukaan: (1) vaikutukset oppimistuloksiin ja myöhempiin työmarkkinatulemiin, (2) vaikutukset oppilaiden hyvinvointiin (esim. stressi, asenteet ja motivaatio) sekä (3) vaikutukset opettajiin ja opetukseen (esim. opettajien kokema stressi, käsitykset järjestelmän toimivuudesta, koetehtävien laadusta ja pedagogisista seurauksista). Useissa tutkimuksissa tarkastellaan useita tulemia, jolloin ne luokitellaan tässä järjestyksessä. Vaikutukset muihin tulemiin kuten esimerkiksi eriarvoisuuteen raportoidaan erikseen päätulosten yhteydessä.

Tutkimusnäytön vahvuuden arvioimiseksi tutkimukset luokitellaan kolmeen ryhmään tutkimusasetelman perusteella: (1) kokeelliset ja kvasikokeelliset tutkimukset, (2) kontrollitutkimukset, joissa hyödynnetään laajoja aineistoja ja pyritään muodostamaan verrokkiryhmä (sisältäen pitkittäis- tai poikittaistutkimukset), sekä (3) tutkimukset, joissa verrokkiryhmää ei ole (ml. haastattelut, kyselyt, havainnointi, narratiiviset kirjallisuuskatsaukset, dokumenttianalyysit ja teoreettiset tarkastelut).⁴

⁴ Nykypolitiikassa korostuu näyttöön perustuva päätöksenteko (Parkhurst, 2016). Tutkimusnäytön vahvuutta arvioidaan tiedonhierarkioiden ja esimerkiksi [What Works Clearinghouse](#) (WWC) -standardien ja [Education Endowment Foundation](#) EEF-kehikkojen avulla, joissa keskeistä on tutkimusasetelmien metodologinen laatu.

Taulukko 1. Tutkimuksien lukumäärä tutkimusteeman ja tutkimusasetelman vahvuuden mukaan

		Tutkimuksen päävasteet		
		A) Oppimistulokset ja työmarkkinatulemat	B) Oppilaiden hyvinvointi ja asenteet	C) Opettajat ja opetus
A) Ketonen & Atjonen (2025) -katsaus				
1	Kokeelliset (RCT) tai kvasikokeelliset tutkimusasetelmat	6	2	1
2	Kontrollitutkimukset aineistoihin perustuen, ja joissa rakennetaan verrokkiryhmä (sis. pitkittäis- tai poikittaistutkimukset)	2	3	1
3	Tutkimukset, joissa ei ole verrokkiryhmää (haastattelu, kyselyt, havainnointi, dokumenttianalyysit ja teoreettiset tutkimukset)	14	17	95
Yhteensä		22	22	97
B) Uusi täydentävä haku				
1	Kokeelliset (RCT) tai kvasikokeelliset tutkimusasetelmat	20	0	0
2	Kontrollitutkimukset aineistoihin perustuen, ja joissa rakennetaan verrokkiryhmä (sis. pitkittäis- tai poikittaistutkimukset)	4	0	0
3	Tutkimukset, joissa ei ole verrokkiryhmää (haastattelu, kyselyt, havainnointi, dokumenttianalyysit ja teoreettiset tutkimukset)	0	0	0
Yhteensä		24	0	0

Mukaan valituista 183 tutkimuksesta 46 tarkastelee oppimistuloksia tai työmarkkinatulemia, 22 oppilaiden hyvinvointia ja asenteita ja 97 opettajia ja opetusta. Ketosen ja Atjosen (2025) artikkeleista 18 ei liity suoraan tarkasteltavaan ilmiöön, vaan käsittelee esimerkiksi valmennuskursseja ja yksityisopetusta (esim. Byun 2010), arvosanojen ennustekykyyyn liittyviä kysymyksiä (esim. Cerdeira ym. 2018), tai koulujen arviointiin liittyviä mittausteknisiä seikkoja (esim. Marks 2021). Nämä tutkimukset rajataan tiiviyn vuoksi tarkemman tarkastelun ulkopuolelle. Liitteen taulukot 1 ja 2 tiivistävät kuitenkin keskeiset havainnot kaikista 183 mukaan otetusta tutkimuksesta.

Taulukossa 1 esitetään tutkimusten määrä tutkimusasetelmia ja päävasteita koskien. Taulukon yläosa kuvaa aikaisempaa Ketosen ja Atjosen (2025) katsausta ja alaosa tätä täydentävää uutta hakuja. Huomataan, että aiemmin tehty haku painottui selvästi tutkimuksiin, joissa tutkimuskohteena olivat pääasiassa opettajat ja opetus ja joissa ei ole verrokkiryhmää, mikä lisää riskiä, että tutkimukset eivät pysty tavoittamaan todellisia syy-seuraussuhteita. Täydentävä haku taas painottuu kokeellisiin ja kvasikokeellisiin määrällisiin tutkimusasetelmiin oppimistuloksia ja työmarkkinatulemia koskien. Koska tämän katsausartikkelin tarkoitus on tarkastella kansallisten kokeiden vaikutuksia ja arvioida tutkimusnäytön vahvuutta, katsaus keskittyy niihin tutkimuksiin, joissa on olemassa verrokkiryhmä tai joissa on pyritty rakentamaan sellainen aineistoon perustuen.

A) Oppimistulokset ja työmarkkinatulemat

Aineistoon sisältyvät artikkelit osoittavat, että etenkin lopullisiin arvosanoihin vaikuttavilla kokeilla on merkitystä nuoren tulevaisuuden kannalta, sillä niitä käytetään valintakriteereinä järjestelmässä, jossa kaikki eivät voi päästä opiskelemaan. Machin ym. (2020) näyttävät Englannissa, kuinka niukka hylkäys korkean panoksen englannin kielikokeessa heikentää koulutuspolkuja: juuri ja juuri kokeessa epäonnistuminen laskee korkeakoulutukseen siirtymisestä 2,5–4 prosenttiyksikköä. Niin ikään Yhdysvalloissa Ahn (2014) näyttää kvasikokeellisella tutkimusasetelmalla, että matematiikan kokeessa niukasti hylätyt oppilaat jatkavat noin 5 prosenttiyksikköä epätodennäköisemmin vaativammalle korkeakoulupolulle kuin niukasti läpäisseet.

Kvasikokeellisista 24 tutkimuksesta neljä tarkastelee koepäiville ajoittuvien satunnaisten negatiivisten shokkien pitkän aikavälin vaikutuksia. Ylioppilaskoepäivälle sattuvat saastekertymät (Ebenstein ym. 2016) tai jalkapallon MM-kisat (Metcalf ym. 2019) heikentävät koetuloksia tilastollisesti merkitsevästi, vaikutuksen ollessa suurempi vähemmistöjen edustajilla ja pojilla. Vastaavasti tutkimusnäyttö osoittaa, että koepäivälle sattunut siitepölyn lisääntyminen (Bensnes 2016) sekä korkea lämpötila (Park 2022) voivat heikentää koesuorituksia. Yksittäiselle kokeelle suuren painon antamiseen liittyikin näin selviä riskejä, jotka tulisi huomioida arviointijärjestelmän suunnittelussa.

Osaamisen taso ja tasa-arvo

Kansallisilla kokeiden vaikutusten luotettava tunnistaminen edellyttää tutkimusasetelmaa, jossa verrataan oppilaita, joita kansalliset kokeet koskevat, muuten samanlaisiin oppilaisiin, joihin kansalliset kokeet eivät koske. Asetelmat, joissa osa nuorista tulee kansallisten kokeiden piiriin ja osa

ei, ovat harvinaisia, koska kansalliset järjestelmät otetaan eri maissa tyypillisesti yhtä aikaa käyttöön kaikille.

Tällainen tutkimusasetelma löytyy kuitenkin Tanskasta, jossa parlamentti päätti vuonna 2006 kansallisten kokeiden käyttöönotosta, mutta toteutus 2009/2010 ei sujunut suunnitellusti. Kansallisten kokeiden käyttöönotto estyi joillakin alueilla tapahtuneen tietojärjestelmien kaatumisen vuoksi, ja näin osa oppilaista ei lopulta tehnyt kansallista koetta.

Andersen ja Skyt Nielsen (2020) käyttivät hyväkseen tietojärjestelmien kaatumisen luomaa luonnollista koetta. Tulosten perusteella kansallinen testaus parantaa oppimistuloksia: lukemisessa vaikutus on +0,1 keskihajontaa kahden vuoden kuluttua ja säilyy vielä neljän vuoden jälkeen (+0,07 SD). Positiiviset vaikutukset ovat suurempia matalan sosioekonomisen taustan, pienituloisten ja maahanmuuttajataustaisten oppilaiden keskuudessa, ja erityisesti kouluissa, joissa tällaisia oppilaita oli paljon. Tutkijat havaitsivat myös, että heikosti menestyneet koulut sekä koulut, joissa oli paljon matalan sosioekonomisen taustan tai maahanmuuttajataustaisten oppilaiden oppilaita, olivat muita todennäköisemmin yhteistyöhaluttomia kansallisten kokeiden järjestämisessä.

Vastaavia tuloksia kansallisten kokeiden osaamiseroja pienentävästä vaikutuksesta saatiin katsauksen muista artikkeleista. Ofek-Shanny (2024) selvitti Israelissa, missä määrin oppilasryhmien väliset osaamiserot heijastavat todellista osaamista ja koemotivaatiota. Aineisto perustuu satunnaistettuun kenttäkokeeseen (RCT), jossa sama oppilas suoritti sekä matalan että korkean panoksen kokeen. Arabien ja juutalaisten välinen ero matalan panoksen testissä on 0,38 keskihajontaa, mutta korkean panoksen testissä vain 0,15 keskihajontaa. Erot selittyvät osittain motivaatiolla pikemminkin kuin osaamisella, mikä viittaa siihen, että matalan panoksen kokeet aliarvioivat vähemmistöjen osaamista. Samantyyppisestä ilmiöstä on viitteitä myös Englannista. Wilson ym. (2011) näyttävät ei-kokeellisella kontrollitutkimuksella, että etniset vähemmistöt etenevät koulussa huomattavasti nopeammin kuin muut oppilaat, ja suurin osa tästä kiinniotosta tapahtuu korkean panoksen testeissä lähes kaikissa kouluissa.

Brunello ym. (2022) tarkastelivat Saksassa kvasikokeellisella korkean ja matalan panoksen testauksen vaikutuksia oppimistuloksiin hyödyntäen osavaltioiden välisiä eroja testauksen ajoittumisessa. He havaitsivat, että korkean panoksen ympäristössä testatut oppilaat saavuttivat 0,17–0,23 keskihajontaa paremmat matematiikan tulokset, mikä vastasi merkittävää osaa yhden kouluvuoden oppimisesta. Vaikutukset olivat suurempia tyttöillä, että matalammin koulutettujen perheiden lapsille, mutta nämä erot eivät ole tilastollisesti merkitseviä.

Yhdysvaltalaisessa tutkimuksessa Papay ym. (2016) tarkastelivat standardoitujen matematiikkakokeiden suoritusluokitusten vaikutuksia, vaikka itse suorituksen luokituksella ei ollut suoria institutionaalisia seuraamuksia. Nämä tulokset osoittavat, että juuri ja juuri "Advanced"-luokitus lisäsi vähävaraisten oppilaiden korkeakouluun hakeutumista noin 5 prosenttiyksikköä verrattuna "Proficient"-luokitukseen. Lisäksi heikommat luokitukset ("Needs Improvement" vs. "Warning") paransivat hieman oppimistuloksia, valmistumista ja jatko-opintoihin hakeutumista, erityisesti oppilailla, joilla ei alun perin ollut suunnitelmia nelivuotisista opinnoista.

Federičová ym. (2017) tutkivat Slovakiassa toteutettua koulureformia, joka siirsi selektiivisten koulujen pääsykokeiden ajankohtaa vuodella eteenpäin. Tämä loi kvasikokeellisen asetelman, jossa yksi ikäluokka kohtasi pääsykokeen nyt ja toinen vasta vuotta myöhemmin, jolloin jälkimmäinen toimii vertailuryhmänä yhdessä tšekkiläisten oppilaiden kanssa, jotka eivät kokeneet muutosta. Tutkimuksessa havaitaan, että korkean panoksen pääsykoe parantaa 10-vuotiaiden IEAn (International Association for the Evaluation of Educational Achievement) TIMSS-arvioinnin (Trends in International Mathematics and Science Study) matematiikan tuloksia edeltävänä vuonna noin 0,2 keskihajontaa, ja vaikutus oli erityisen suuri taitavilla oppilailla. Koska TIMSS on matalan panoksen koe, se ei mittaa samalla tavalla oppilaiden osaamista eikä taktista testiharjoittelua, jolloin koulureformi paljasti, että korkean panoksen koe saattoi lisätä oppilaiden ja vanhempien ponnistelua jo ennen koetta.

Falch ja Naper (2013) näyttivät, että tytöt saavat Norjassa kaikissa oppiaineissa järjestelmällisesti parempia opettaja-arvosanoja kuin pojat suhteessa anonyymeihin kokeisiin. Ruotsissa Jalava ym. (2015) suorittivat satunnaistetun kenttäkokeen (RCT) Tukholman 6.-luokkalaisille tarkoituksenaan tutkia, miten erilaiset kannustimet vaikuttavat matematiikan testituloksiin. Kriteeriperusteinen arviointi ei näyttänyt parantavan suoriutumista. Sen sijaan kilpailulliset kannustimet osoittautuivat tehokkaimmiksi: oppilaiden välinen ranking ja sen yhdistäminen palkintoon nostivat testituloksia (0,33–0,45 SD). Pojat reagoivat kilpailuun voimakkaammin.

Katsaukseen sisältyi myös yksi laaja kansainvälinen tutkimus, jossa tarkasteltiin koejärjestelmiä PISA-arviointiin osallistuneissa maissa kvasikokeellisin menetelmin. Bergbauer ym. (2024) analysoivat yli 2 miljoonan oppilaan aineistoa 59 maasta vuosilta 2000–2015. Tulokset osoittivat, että koejärjestelmien vaikutukset oppimistuloksiin riippuivat merkittävästi maiden lähtötasosta sekä maan koe- ja valvontajärjestelmästä. Standardoidut ulkoiset kokeet paransivat oppimistuloksia erityisesti matalan ja keskitason suoriutumisen maissa, kun taas standardoidut kansalliset kokeet olivat hyödyllisiä lähinnä heikosti suoriutuviissa maissa ja voivat jopa heikentää oppimistuloksia korkean suoritustason maissa. Lisäksi Bergbauer ym. (2024) huomauttivat, että pelkkä sisäinen raportointi ei vaikuta oppimistuloksiin, ja tiukempi opettajien valvonta voi joissakin tapauksissa heikentää oppimista erityisesti korkean suorituskyvyn maissa. Maakohtaisiin vertailuihin tulee kuitenkin suhtautua varoen, sillä ne kaappaavat useita ajassa muuttuvia maakohtaisia tekijöitä.

Alueelliset erot vaikutusten tunnistamisen apuna

Monet tutkimusasetelmat hyödyntävät alueellisia eroja kansallisten kokeiden vaikutusten tunnistamiseen. Asetelmia voidaan löytää erityisesti hajautetun hallinnon järjestelmistä, joissa voidaan hyödyntää kvasikokeellisia tai kontrollitutkimuksellisia lähestymistapoja ja osavalttioiden välisiä eroja keskitettyjen kokeiden vaikutusten tutkimiseen, kuten esimerkiksi Saksassa.

Fidjeland (2023) vertasi korkean panoksen kokeita alhaisen panoksen kokeisiin, joka oli mahdollista johtuen koereformin eri ajoituksesta Norjan eri maakunnissa. Hän havaitsi, että korkean panoksen kokeet paransivat koetuloksia (0,04–0,06 SD) erityisesti alueilla, joissa oppilailla oli todellinen kouluvalinnan mahdollisuus, ja että vaikutus näkyi myös matalan panoksen kansallisissa arviointituloksissa (0,07 SD), mikä viittasi oppimisen paranemiseen, ei pelkkään kokeessa

ponnisteluun. Myös nämä tulokset viittasivat siihen, että opiskelijat, joille arvioinnin kannustimet olivat suurimmat, reagoivat niihin muita voimakkaammin.

Jürges ym. (2010) hyödynsivät kvasikokeellisessa lähestymistavassaan keskitettyjen päättökokeiden vs. hajautettujen kokeiden roolia eri osavaltioissa alakoulun jälkeisessä koulutuksessa. He havaitsivat, että keskitetyt päättökokeet nostivat matematiikan pistemäärä määrällä, joka vastaa noin kolmasosaa kouluvuodesta (0,11 SD). He havaitsivat myös vaikutuksen oppilaiden asenteisiin: matematiikka koettiin tämän seurauksena vaikeampana ja tylsempänä. Tätä voi selittää havainnot, että opettajat määräisivät enemmän kotitehtäviä ja tekivät useammin tarkistuksia tunneilla. Samaa tutkimusasetelmaa hyväksikäyttäen Jürges ym. (2012) osoittivat, että keskitetyt päättökokeet paransivat merkittävästi oppimäärään perustuvaa osaamista (0,25 SD), mutta ne eivät vaikuttaneet matematiikan tuloksiin. Sen sijaan ne vähensivät oppimisen iloa ja itsevarmuutta matematiikassa.

Piopiunik ym. (2013) ja Schwerdt ym. (2017) taas tarkastelivat Saksassa keskitettyjen ja paikallisten päättökokeiden vaikutuksia työmarkkinatuloksiin ja osoittivat, että keskitetyt kokeet voivat parantaa joissain ryhmissä myös taloudellisia tuloksia. Piopiunik ym. (2013) havaitsivat, että keskitetyt kokeet olivat yhteydessä noin 11 prosenttia korkeampiin tuloihin matalimman koulutusasteen suorittaneilla sekä alhaisempaan työttömyyteen (noin 3–4 prosenttiyksikköä), mutta vaikutuksia ei löytynyt korkeammin koulutetuilla. Schwerdt ym. (2017) puolestaan osoittavat kvasikokeellisella tutkimuksellaan, että arvosanojen paraneminen näkyi selvästi suurempina ansioina keskitettyjen kokeiden järjestelmässä (6 % vs. 1,6 %), minkä he tulkitsivat viittaavan siihen, että keskitetyt kokeet tuottavat tarkemman ja työmarkkinoilla arvokkaamman signaalin yksilön osaamisesta.

Vaikutusmekanismit

Katsaukseen sisältyvät tutkimukset antavat myös näyttöä niistä mekanismeista, joilla kansalliset kokeet vaikuttavat oppilaisiin. Yksilötasolla kansalliset kokeet näyttävät vaikuttavan yksilöiden käyttäytymiseen. Jos oppilaan on mahdollista parantaa omaa tilannettaan, tieto heikosta suoriutumisesta parantaa tuloksia.

Beuchert ym. (2020) tutkivat Tanskassa kvasikokeellisella menetelmällä, kuinka vanhemmille annettava standardoitu arviointipalaute 3. luokan matematiikan tuloksista vaikutti sekä oppilaan oppimistuloksiin että motivaatioon ja oppilaan oireiluun. Ensinnäkin negatiivinen palaute paransi myöhempiä oppimistuloksia: 3. luokan matematiikan ”negatiivinen” luokkasignaali paransi 6. luokan matematiikan tuloksia (0,061 SD), ja vaikutukset ulottuivat myös 4. luokan lukutaidon tuloksiin (0,045 SD). Negatiivisen palautteen myönteiset vaikutukset olivat suurimmat heikoimmille oppilailla ja pojilla. Sen sijaan haitoista ei löytynyt näyttöä: hyvinvointimittarit eivät viitanneet systemaattisiin kielteisiin vaikutuksiin. Erityisesti heikoimmin suoriutuvien oppilaiden akateeminen itsevarmuus kasvoi, eivätkä minäpystyvyys ja sisäinen motivaatio laskeneet. Oppilaiden kokemassa opettajan antamassa tuessa ei tapahtunut muutoksia, mikä viittaa siihen, että vaikutukset välittyivät ensisijaisesti vanhempien ja/tai oppilaan lisääntyneen panostuksen kautta.

Toinen kvasikokeellinen tanskalainen tutkimus tarkasteli ylioppilastutkinnon arvosanauudistusta, jossa vanhan 13-asteikon arvosanat muunnettiin uuteen 7-portaiseen asteikkoon. Hvidman ja

Sievertsen (2021) hyödynsivät tätä arvosteluasteikon muutosta tilanteessa, jossa muutos ei heijastanut todellista osaamista, analysoidakseen, miten korkean paineen kokeen kannustimien muutos vaikuttaa 2. ja 3. lukiovuoden arvosanoihin, kansallisten kokeiden tuloksiin sekä myöhempään yliopistoon hakeutumiseen ja sieltä valmistumiseen. He havaitsivat, että arvosteluasteikon muutoksesta syntynyt negatiivinen arvosanashokki johti parempiin myöhempisiin arvosanoihin (0,08 SD), ja että vaikutus näkyi myös ulkoisissa arvosanoista riippumattomissa kokeissa, viitaten todelliseen osaamisen kasvuun. Tämän lisäksi he havaitsivat parantaneen todennäköisyyttä valmistua yliopistosta (2 %-yks.), ja että vaikutukset olivat suurimpia oppilailla, jotka olivat lähellä rajaa päästä yliopistoon. Tulokset tukevat tulkintaa, että summatiivisella, korkean panoksen arvioinnilla voi olla kannustinvaikutuksia, jotka voivat parantaa sekä suorituksia että koulutuksellisia lopputuloksia.

Saksassa toteutettu satunnaistettu kenttäkoe (RCT) tarkasteli, miten palautteen ajoitus (aikaisin vs. juuri ennen koetta) ja muoto (tasosijoitus vs. sijoituksen muutos) vaikuttivat 5.–6.-luokkalaisten suorituksiin korkean panoksen matematiikan loppukokeessa (Fischer ym. 2023). Tulokset osoittivat, että varhainen palaute paransi suoritusta (0,224 SD), kun taas myöhäinen palaute heikensi sitä (-0,126 SD). Voimakkain vaikutus liittyi palautteeseen, joka viestii osaamisen heikentyneen: aikaisin annettuna se lisäsi valmistautumista, mutta juuri ennen koetta annettuna se heikensi suoritusta merkittävästi.

Toisaalta suhteellisen ranking-tiedon antaminen myöhään voi kirjallisuuden perusteella vaikuttaa eri tavoin hyvin ja heikosti menestyviin oppilaisiin. Goulas ja Megalokonomou (2021) osoittivat, että suoriutumistietoa saavat jo aikaisemmin hyvin suoriutuvat kreikkalaiset lukiolaiset paransivat edelleen suoriutumistaan (0,15 SD) viimeisen lukiovuoden kansallisissa kokeissa, kun taas heikommin suoriutuvien oppilaiden tulokset heikkenivät (-0,3 SD). Erityisesti tytöt reagoivat tähän suoriutumistietoa koskevaan palautteeseen negatiivisemmin.

Kansalliset kokeet vaikuttavat myös koulujen kannustimiin ja käyttäytymiseen. Winters ym. (2010) tutkivat Floridassa, mitä tapahtuu kouluille, kun niille asetetaan suoriutumiseen perustuvia sanktioita. Kvasikokeelliset tulokset osoittavat, että sanktioiden kohteena olevien koulujen oppilaat paransivat korkean panoksen kokeiden tuloksia lukutaidossa (0,09 SD) ja matematiikassa (0,20 SD). Samalla myös alhaisen panoksen kokeiden luonnontieteiden tulokset paranivat (0,10 SD). Näyttöä syrjäyttämisaikutuksista ei havaittu, vaan tulokset viittasivat laaja-alaiseen oppimisentulosten nousuun.

Rockoff ja Turner (2010) raportoivat samankaltaisia tuloksia New Yorkista, jossa standardoituja koetuloksia alettiin käyttää koulujen luokittelussa. Kvasikokeelliset, koulutason analyysit osoittivat, että heikon vastuuarvosanan saaminen lisäsi oppilaiden matematiikan ja osin englannin kielen kokeen pisteitä (0,05–0,10 SD) muutaman kuukauden aikajänteellä. Tutkimus ei löytänyt näyttöä testitulosten mahdollisesta manipuloinnista. Tutkimuksen laadullinen seuranta viittasi siihen, että vanhempien arviot koulun laadusta paranivat, mutta opetuskäytännöt muuttuivat koepainotteisemmiksi (mm. vähemmän ryhmytystä, enemmän kokeisiin valmistavaa opetusta).

Aitoihin vaikutuksiin näyttää yhdistyvän myös koulujen ja opettajien strategista käyttäytymistä. Ehkä merkittävin yhdysvaltalainen tutkimus tarkasteli Texasissa kvasikokeellisella asetelmalla koulujen suoritusluokituksiin perustuvan kannustinpaineen vaikutuksia (Deming ym. 2016). Tämän kvasikokeellisen tutkimuksen tulokset osoittivat keskimäärin merkittävän parannuksen matematiikan tuloksissa (jopa 0,5 SD), erityisesti heikoimmissa kouluissa, joissa havaittiin myös lisääntyntä korkeakouluun hakeutumista, tutkinnon suorittamista ja myöhemmin korkeampia ansiotuloja. Samansuuntaisia tuloksia ei kuitenkaan havaita lukutaidon osalta. Vaikka vaikutukset olivat keskimäärin positiivisia, järjestelmä lisäsi myös eriarvoisuutta, koska etenkin ns. ”eliittikouluissa” tehtiin strategista vähemmistöoppilaiden siirtämistä erityisopetukseen, mikä heikensi merkittävästi näiden oppilaiden pitkän aikavälin koulutus- ja työmarkkinatuloksia.

Myös Hofflinger ja von Hippel (2020) osoittavat kvasikokeellisella asetelmalla, että vaikka korkean panoksen testien tulokset paranivat Chilessä matalan panoksen kokeisiin verrattuna (0,2–0,3 SD), niin osa tästä noususta selittyy strategisella valikoitumisella: heikoimpia oppilaita jätettiin kokeiden ulkopuolelle. Ilmiö oli erityisen voimakas heikoimmissa kouluissa, joissa sekä tulosparannukset että manipulointi olivat suurempia.

Vaikka tämä katsaus keskittyi kansallisten kokeiden vaikutukseen, Ketosen ja Atjosengin aineistoon sisällynyt Neves ym. (2012) tutkimus Portugalista osoitti, että yksityiskouluissa oppilaat saivat parempia arvosanoja suhteessa menestykseensä standardoidussa kokeessa kuin oppilaat julkisessa koulussa ja tämä ero oli suurin hyvin menestyvillä oppilailla. Ilmiön katsotaan johtuvan arvosanojen painottamisesta korkeakouluvalinnoissa. Tämä viittaa siihen, että oppilaitosten strateginen käyttäytyminen voi heijastua opettajien antamiin arvosanoihin.

B) Oppilaiden hyvinvointi ja asenteet

Katsaukseen sisältyvä tutkimuskirjallisuus kertoo myös kansallisten kokeiden vaikutuksista oppilaiden hyvinvointiin ja asenteisiin.

Kokeisiin liittyvää itse raportoitua ja mitattua stressiä ja sen voimistumista toisen asteen koulutuksen viimeisinä kouluvuosina käsitellään laajasti Wuthrich ym. (2020) kirjallisuuskatsauksessa, joka valikoitui mukaan tähän katsaukseen. Heissel ym. (2021) osoittivat yhdysvaltalaisella aineistolla, että testiviikko nosti oppilaiden kortisolitasoja merkittävästi (noin +18 %), erityisesti jo valmiiksi kuormittuneissa ympäristöissä. Australialainen tutkimus (Wuthrich ym., 2021) osoitti, että lukiolaisten korkean panoksen loppukokeisiin liittyvä paine lisäsi hyvinvoinnin kuormitusta: itse raportoitu stressi kasvoi hieman (0,2 SD) ja ahdistus kasvoi selvästi (1,7 SD) lukion viimeisen vuoden aikana, erityisesti ennen kokeita. Näissä kahdessa edellä mainitussa tutkimuksessa ei kuitenkaan ole selvää verrokkiryhmää tilanteesta, jotka kuvaisivat hyvinvoinnin kehitystä alhaisen panoksen koe tilanteessa tai vaihtoehtoisesti järjestelmässä, jossa käytetään pääsykokeita korkeakouluvalinnoissa päättöarvioinnin arvosanojen sijaan.

Koestressi voi ilmentyä jyrkemmin nuorten ajattelussa äärimmilleen kilpailullisessa ympäristössä. Wangin (2016) tutkimus Etelä-Koreassa osoitti, että odotuksia heikompi suoriutuminen korkean panoksen kokeessa lisäsi itsetuhoisten ajatusten todennäköisyyttä noin 17 prosentilla, mutta

vaikutus oli lyhytaikainen häviten vuoden kuluessa. Lisäksi tämä ei näkynyt matalan panoksen kokeissa.

Katsaukseen tuli mukaan kaksi ehdot täyttävää tutkimusta Ruotsista, joissa keskityttiin oppilaiden hyvinvointiin, ja joihin on hyvä keskittyä tässä tarkemmin Ruotsin läheisyyden takia. Högberg ja Lindgren (2023) analysoivat Ruotsin vuosien 2010–2014 koulutusreformia kvasikokeellisella lähestymistavalla rakentamalla synteettinen vertailuryhmä Norjasta ja Itävalasta sekä vertailemalla Ruotsin kehitystä muihin EU -maihin. Kokonaisuudessaan laaja reformi koostui useista osista, kuten arvioinnin aikaistamisesta, kansallisen testauksen laajentamisesta, uudesta standardiperustaisesta ja suoritusorientoituneesta opetussuunnitelmasta, inklusiopolitiikasta sekä tiukemmista siirtymäkriteereistä toiselle asteelle. Högberg ja Lindgren (2023) havaitsivat, että oppilaiden itse raportoima kouluun kuulumisen tunne heikkeni voimakkaasti reformin myötä, erityisesti valmiiksi haavoittuvassa asemassa olevien ryhmien keskuudessa. Tähän ruotsalaiseen tutkimukseen liittyy kuitenkin merkittäviä rajoitteita: se tarkastelee laajaa reformikokonaisuutta yksittäisten toimien kuten kansallisten kokeiden käytön sijaan.

Högberg ym. (2021) tarkastelivat samaa reformikokonaisuutta keskittyen vuoden 2012 arviointireformin vaikutuksiin. Reformissa aikaistettiin arvosanoja ja siirrettiin kansallisten kokeiden aloitusikää. Tulosten mukaan reformi lisäsi 7.-luokkalaisten itsensä raportoimaa koulustressiä (0,31 SD), heikensi akateemista itsetuntoa (0,22 SD) ja laski elämäntyytyväisyyttä (0,25 SD). Tuloksia on kuitenkin tulkittava varoen, sillä tutkimuksessa käytetty kontrolliryhmä koki myös muutoksia, ja ei ole selvää miten kansallisten kokeiden aloitusiän vaikutus onnistuttiin erottamaan arvosanojen antamisen aikaistamisesta.

C) Opettajat ja opetus

Edellä mainitun Rockoff ja Turner (2010) sivutuloksen lisäksi, katsaukseen löytyi vain kaksi tutkimusta siitä, miten kansalliset kokeet vaikuttavat opettajiin ja opetukseen; yksi kvasikokeellinen ja yksi kontrollitutkimus, molemmat Yhdysvalloista. Näiden lisäksi Berry ym. (2020) -tutkimus Intiasta osoitti, että arviointiuudistukset eivät välttämättä tuota tulosta, jos opettajat muuttavat vain arviointikäytäntöjään mutteivat opetuskäytäntöjään.

Barrett ja Toma (2013) osoittivat kuinka rehtorit käyttävät luokkakokoa strategisesti suhteessa opettajan aiempaan näyttöön. Oppimistulosarviointien perusteella ns. tehokkaat opettajat saivat johdettavakseen systemaattisesti suurempia luokkia. Tämä viittaa siihen, että rehtorit pyrkivät maksimoimaan koulun kokonaissuoriutumisen sijoittamalla enemmän oppilaita parhaiden opettajien ryhmiin.

Fuchsman ym. (2023) tarkastelivat Yhdysvalloissa kvasikokeellisella asetelmalla standardoidun korkeanpanoksen testauksen poistamisen vaikutuksia opettajien työmarkkinakäyttäytymiseen. Tulokset osoittivat, ettei testauksen poistamisella ole vaikutusta opettajien alanvaihtoon tai siirtymiin koulujen välillä, mutta se vähensi hieman sisäisiä siirtoja ja lisäsi tilastollisesti merkitsevästi, joskin määrältään vähäisesti, aloittelevien opettajien pysyvyyttä.

Taulukko 2 kokoaa lopuksi keskeiset eurooppalaiset tutkimukset, joita voidaan hyödyntää arvioitaessa kansallisten kokeiden mahdollisia vaikutuksia oppimistuloksiin, oppilaiden hyvinvointiin ja asenteisiin sekä työmarkkinatulemiin. Valtaosa tutkimuksista osoittaa, että kansalliset kokeet tai niihin liittyvät piirteet parantavat oppimistuloksia. Samanaikaisesti on näyttöä siitä, että kokeet vaikuttavat negatiivisesti oppilaiden itsensä raportoimaan hyvinvointiin, ja lisäävät stressiä erityisesti kokeiden ajallisen läheisyyden aikana. Pitkäaikaisista negatiivisista vaikutuksista ei kuitenkaan löydy näyttöä, sillä tätä ei tarkasteltu katsaukseen sattuneissa artikkeleissa.

On syytä korostaa, että Ketosen ja Atjosen (2025) katsauksen kautta mukaan sisältyneet tutkimukset, joissa ei ole verrokkiryhmää tarkastelevat laajasti oppilaiden ja opettajien käsityksiä sekä korkean panoksen että matalan kansallisista kokeista. Nämä tutkimukset, jotka perustuvat pääsääntöisesti tutkimuskohteiden omiin arvioihin, luokahuonehavaintoihin ja dokumenttianalyyseihin, pitävät kokeita ja testaamista haitallisina (ks. liitetaulukko 2). Opettajat, joiden työtä ulkoapäin tulevat kansalliset kokeet rajoittavat ja ohjaavat jo puhtaan määritelmällisesti, kokevat niiden heikentävän pedagogisia mahdollisuuksia ja kaventavan opetusta.

Taulukko 2. Katsauksen keskeisimmät kansallisten kokeiden vaikutuksia suorasti tai epäsuorasti tarkastelevat tutkimukset Euroopasta

Maa/Tutkimus	Menetelmä (N)	Korkean panoksen koe (KP)	Matalan panoksen koe (MP)	Heterogeenisuus ja huomiot
Tanska				
Andersen ja Skyt Nielsen (2020); kansallisten kokeiden käyttöönotto.	IV (N≈136 887 oppilasta lukutestissä ja 46 338 mat., 2.–8. lk.)		+0,10 SD (2 v.) ja +0,07 SD (4 v.) lukem.; +0,07 SD (3 v.) matem. testitulokseen.	Kokeilla on suurin vaikutus heikon SES perhetaustan oppilaille.
Beuchert ym. (2020); negatiivisen matematiikan testipalautteen antaminen.	RD (N=192 577) 3.lk. oppilaat ja heidän vanhempansa		+0,061 SD mat. 6. lk. +0,045 SD lukem. 4. lk.; ei näyttöä negatiivisista hyvinvointi-vaikutuksista.	Testipalautteen vaikutus on suurin pojille ja heikosti pärjääville oppilaille.
Hvidman ja Sievertsen (2021); arvosteluasteikon uudistus.	DID (N≈27000, lukion 2. ja 3. vuoden oppilaat)	+0,08 SD nousu myöhemmissä arvosanoissa; +2 %-yks. vaikutus yo-valmistumiseen.		Vaikutus suurempi naisilla ja hyvän peruskoulun keskiarvon omaavilla.
Norja				
Fidjeland (2023); korkean panoksen kokeiden laajentuminen.	Tripla-DID (N≈791 000, peruskoulun 10. lk)	+0,04–0,06 SD parannus KP päättökokeessa.	+0.07 SD parannus kansallisissa MP arviointituloksissa.	Viitteitä, vaikutus isompi heikon SES:n oppilaille matematiikassa.
Ruotsi				
Jalava ym. (2015); erilaisten kannustimien	RCT (N=1 045, 6. Lk.)		+0,33–0,45 SD matem. vaikutus taitavimmilla oppilailla (perustuen	Useammat heikoista jättävät yrittämättä

vaikutus matematiikan testituloksiin.		kilpailulliseen asetelmaan); kriteeriperusteinen arvosana ei vaikutusta.	kilpailullisessa asetelmassa (tilastollisesti merkittävä).
Högberg ym. (2021), laaja reformikokonaisuus, jonka osana kansallisia kokeita siirrettiin aikaisemmaksi.	DID (N≈13 000 5., 7., ja 9 lk. oppilasta)	+0,31 SD koulustressi; -0,22 SD akat. itsetunto; +0,15 SD psykosom. oireet; -0,25 SD elämäntyyt.	Mittarit ovat nuorten itseraportoituja suureita. Kontrolliryhmä myös kokee reformin.
Högberg ja Lindgren (2023), laaja reformikokonaisuus.	DID (N≈13 000) + synteettinen kontrolli.	Vähentää koulukuulumisindeksiä -0,19 SD.	Suurimmat negatiiviset vaikutukset heikon SES:n oppilaille.
Slovakia			
Federičová ym. (2017), reformi siirsi pääsykokeita.	DID (N≈ 5 600, 4.lk) ja verrokkiryhmä Tšekistä.	+0.2 SD matematiikan tuloksissa.	Suurin hyöty huippuoppilaille ja niille, jotka ovat lähellä pääsyrajaa.
Saksa			
Jürges ym. (2010), KP vs. MP, poikkileikkaus.	Regressio (N≈3 195) eri osavaltioiden välillä.	+0,11 SD mat. (TIMSS), lisää itseraportoitua. Tylsyyttä.	Matematiikka koetaan vaikeammaksi; enemmän kotitehtäviä.
Jürges ym. (2012), KP vs. MP Koejärjestelmään, paneeli.	DID (N≈5000 9. lk. ja 10. lk.)	+0.25 SD oppimäärään perustuvaa osaamista.	Vähentää itseraportoitua oppimisen iloa ja motivaatiota.
Brunello ym. (2022), KP vs. MP testausikä, paneeli.	DID (N≈7 900)	+0,17–0,23 SD matematiikkaa.	Vähemmän koulutettujen perheiden lapset hyötyvät eniten.
Piopiunik ym. (2013), KP vs. MP, poikkileikkaus.	Regressio, missä kiinteät vaikutukset (N≈5 500 2.aste)	+11 % vaikutus tuloihin; -4 %-yks. vaikutus työttömyyteen (40 vuotiaana).	
Schwerdt ja Woessmann (2017) KP vs. MP, paneeli.	DID (N≈4 700) yliopistosta valmistuneet	1 SD:n parempi arvosana + 6 % tuloja KP osavaltioissa, +1,6 % MP kokeissa.	Vaikutus on suurempi yksityisellä sektorilla.
Fischer ja Wagner (2023), matematiikan palautteen antaminen varhain vs. myöhään.	RCT (N=330, 5.–6. lk)	Varhainen palaute (päiviä ennen koetta) +0,224 SD; myöhäinen palaute (juuri ennen koetta) -0,126 SD matematiikkaan.	Pojat reagoivat palautteeseen voimakkaammin ja päivittävät itseluottamustaan negatiivisesti (-0,19 SD).

Kreikka			
Goulas ym. (2021), 11 lk., suhteellisen rank- tiedon antaminen viimeisen vuoden alussa.	DID (N≈41 000 lukiolaista)	Oppimistulokset: hyvin valmiiksi pärfäävät +0,15 SD, heikommat oppilaat -0,30 SD.	Tytöt reagoivat palautteeseen negatiivisemmin.
RCT = satunnaistettu kenttäkoe; DID = kvasikokeellinen, erotusten erotus; RD = kvasikokeellinen, pistemäärärajojen epäjatkuvuus; IV = kvasikokeellinen, instrumentaalimuuttuja; Tripla-DID = kolmikertainen DID. <i>Huom.</i> Kaikkien tutkimusten tulokset on avattu liitteen taulukoissa 1 ja 2.			

JOHTOPÄÄTÖKSET

Tämän katsauksen tulokset osoittavat, että kansallisilla kokeilla voi olla merkittäviä positiivisia vaikutuksia nuorten oppimistuloksiin sekä arvioinnin yhdenvertaisuuteen. Samalla tulee huomioida, että vaikutukset voivat erilaisia riippuen oppilaiden taitotasosta, sosioekonomista tai etnisestä taustasta tai sukupuolesta.

Yhdenvertainen arviointi

On kasvavaa kansainvälistä tutkimusnäyttöä siitä, että opettajien antamat arvosanat voivat olla systemaattisesti vinoutuneita sukupuolen, sosioekonomisen ja etnisen taustan mukaan siten, että tytöt, korkeamman sosioekonomisen aseman nuoret ja kantaväestöön kuuluvat saavat parempia arvosanoja samasta suoriutumisesta (esim. Gil-Hernández & Espadafor 2025; Alesina ym. 2024; Lavy & Sand 2018). Vastaavia vinoumia on havaittu myös ylioppilaskokeisiin liittyvissä opettaja-arvioissa ulkomaalaistaustan mukaan (Sahlström & Silliman 2024). Koska koulutuspolkujen eriytyminen Suomessa tapahtuu pitkälti toisen asteen valinnoissa, peruskoulun päättöarvioinnin yhdenvertaisuus on keskeinen kysymys (Härkönen & Sirniö 2020).⁵

Kansallisen koulutuksen arviointikeskuksen (KARVI) mittausten mukaan opettajien antamat arvosanat voivat poiketa kansallisista osaamismittauksista merkittävästi, ja tytöt saavat keskimäärin parempia arvosanoja suhteessa osaamiseensa kuin pojat (Metsämuuronen 2023; Kauppinen & Marjanen 2020). Tutkimusta on kuitenkin rajoittanut vertailukelpoisen arviointitiedon puute.

Tarkastelluissa artikkeleissa osaaminen oli parantunut enemmän heikommassa asemassa olevilla vähemmistöillä, pojilla ja alemmilla sosioekonomisilla ryhmillä. Tältä osin tulokset vastaavat myös kirjallisuuskatsauksen ulkopuolelle jääneessä koulutussosiologisessa kirjallisuudessa saatuja tuloksia, joiden perusteella standardoidut kokeet vähentävät oppimistulosten sosiaalista

⁵ Perusasteen päättöarvosanojen yhteismitallisuus on merkittävä yhdenvertaisuuskysymys, jolla on merkittäviä vaikutuksia lasten oikeuksien toteutumiseen Suomessa. Lasten oikeus yhdenvertaiseen kohteluun on taattu Perustuslain 6 §:ssä sekä YK:n lapsen oikeuksien sopimuksen 28 ja 29 artikloissa. Lisäksi lapsen edun ensisijaisuus on keskeinen periaate kaikessa lapsia koskevassa toiminnassa. Lukiossa opiskelijoiden yhdenvertaisuutta pyritään takaamaan kansallisten kokeiden, eli ylioppilaskirjoitusten avulla, joiden tavoitteena on turvata yhdenvertainen pääsy jatko-opintoihin. Ylioppilaskirjoitukset ovat olleet Suomessa käytössä yli 170 vuoden ajan (Kaarninen ja Kaarninen 2002).

eriarvoisuutta mm. luomalla kouluille kannustimia panostaa heikompien ryhmien osaamisen tukemiseen (van de Werfhorst & Mijs 2010; Bol ym. 2014). Esimerkiksi havainto siitä, että ns. tehokkaat opettajat voivat saada lisää oppilaita opetettavakseen (ks. Barrett ja Toma 2013), voi olla yksi ilmentymä tästä.

Osaamisen parantaminen

Yhdenvertaisen arvioinnin lisäksi kansallisilla kokeilla on tavoiteltu oppimistulosten parantamista. Kansallisten kokeiden tuottama informaatio näyttää tutkimuksen valossa vaikuttavan oppimiseen vaikuttavien toimijoiden toimintaan. Etenkin oppilasarviointiin vaikuttavien kokeiden nähdään vaikuttavan yksilöiden käyttäytymiseen niin, että opiskeluun investoidaan enemmän aikaan ja vaivaa ennen kokeeseen osallistumista. Nuoret voivat olla usein lyhytnäköisiä (esim. Sutter ym. 2019), mutta tieto osaamisesta näyttää vaikuttavan myös vanhempien ajankäyttöä ja muita panostuksia koskevaan päätöksentekoon (ks. Beuchert ym. 2020). Kansallisten kokeiden antamalla tiedolla pyritään myös ohjaamaan kouluja ja opettajien toimintaa. Esimerkiksi opettajien on havaittu korjaavan omaa arviointiaan, saadessaan tietoa antamiensa arvosanojen vääristymistä (Alesina ym. 2024).

Kirjallisuuskatsaukseen sisältyvät artikkelit havaitsivat hyvin samansuuntaisia vaikutuksia. Vaikutus parempaan osaamiseen havaittiin myös katsauksen aineistoon sisältyvissä tutkimuksissa kansallisten kokeiden lisäksi erillisillä otospohjaisilla oppimistulosarvioinneilla. Tämä kertoo, ettei kansallisiin kokeisiin liittyvä oppimistulosten paraneminen selity yksin tai pääasiallisesti vain kyseiseen kokeeseen valmistautumisella, vaan kokeet voivat heijastua koulunkäyntiin panostamiseen ja oppimistulosten paranemiseen laajemminkin.

Opettajat ja oppilaat

Vaikka edellä esitetty kokonaiskuva viimeaikaisen tutkimuskirjallisuuden kansallisia kokeita koskevasta näytöstä eroaa olennaisesti Ketosen ja Atjosen (2025) tulkinnoista, katsaukset osin täydentävät toisiaan. Ketonen ja Atjonen tulkitsevat esim. Barrett ja Toman (2013) tuloksen siitä, että koulut pyrkivät lisäämään hyviä tuloksia saavuttavien opettajien oppilasmääriä strategiseksi optimointikäyttäytymiseksi. Kyse voi olla käänteisesti samasta ilmiöstä, jolla mm. Bol ym. (2014) selittävät kansallisten kokeiden sosiaalisia eroja vähentävää vaikutusta: oikein toteutettu puolueeton arviointi johtaa objektiivisempaan oppilaiden ohjaamiseen eri koulutuspoluille ja suurempaan panostukseen heikommin suoriutuviin ryhmiin. Näin kansalliset kokeet voivat muodostaa mekanismin, joka suojaa oppilaita ja kouluja sellaiselta asiaankuulumattomalta vaikutukselta, joka esim. Neves ym. (2012) tuloksissa heijastui yksityiskoulujen arvosanainflaatiossa, joiden tuloksena vanhemmille muodostui mahdollisuus tosiasiallisesti ostaa lapsilleen paremmat arvosanat ja siten paremmat mahdollisuudet edetä koulutuksessa.

Yhdessä näiden tulosten kanssa Ketosen ja Atjosen (2025) raportoimat havainnot opettajien usein kriittisistä käsityksistä kansallisista kokeista tukevat tämän katsauksen tulkintaa, että opettajan ja oppilaan näkökulmasta tarkasteltuna sama asia voi näyttää hyvin erilaiselta. Oppilaiden osaamisen arvioinnin yhdenvertaisuuden parantaminen kansallisilla kokeilla ja osaamiseroja mahdollisesti

pienentävät käytännöt voidaan opettajien näkökulmasta kokea opetusta kaventavina ja eriarvoistavina toimenpiteinä.

Hyvinvointi ja stressi

Toisaalta kirjallisuus antaa johdonmukaisen kuvan siitä, että kokeet ja niihin liittyvät paineet lisäävät oppilaiden stressiä lyhyellä aikavälillä, erityisesti kokeiden läheisyydessä. Epävarmaa on, onko kansallisilla kokeilla oppilaiden hyvinvointiin sellaisia pidempiaikaisia vaikutuksia, joista katsaukseen sisältyvissä artikkeleista ei löytynyt näyttöä.

Katsauksessa tavoittamatta jäi myös tutkimuskirjallisuus siitä, miltä osin oppilaiden stressi liittyi arvioinnin tapaan tai arviointiin liittyviin panoksiin, kuten ennakoituun menestykseen tulevassa opiskelijavalinnassa. Jatko-opintoihin pääsyn kannalta ratkaisevaan arviointiin liittyy potentiaalisesti stressiä riippumatta arvioinnin tavasta.

LOPUKSI

Tämän katsauksen tulokset poikkeavat selvästi Ketosen ja Atjosen (2025) aiemmista tulkinnoista. Varsin vaatimattomalla täydentävällä kirjallisuushaulla sekä Ketosen ja Atjosen käsittelemien artikkelien tarkemmalla analyysillä tutkimuskirjallisuudesta löytyy merkittävästi näyttöä, jotka viittaavat siihen, että kansallisilla kokeilla voidaan saavuttaa merkittäviä hyötyjä. Tämän näytön valossa ei ole yllättävää, että kansalliset kokeet ovat käytössä laajasti muissa Pohjoismaissa sekä kansainvälisesti.

Miksi Suomessa on päädytty niin erilaisiin ratkaisuihin? Yksi mahdollinen selitys on, että opettajien pedagogista autonomiaa korostetaan vahvasti, ja kansalliset kokeet voisivat haastaa tätä. Mahdollisesti myös vuosituhaten vaihteen PISA-menestys ja siitä johdetut päätelmät ovat hallinneet Suomessa käytävää peruskoulukeskustelua.

Huomiota herättävää on, että toisin kuin muissa Pohjoismaissa, Suomessa kiinnitetään suhteellisen vähän huomiota koulutuksen tietoperusteiseen kehittämiseen ja yhdenvertaisuuden toteutumisen varmistamiseen. Lisäksi eri toimijoiden keskustelua leimaavat vahvat näkemykset peruskoulusta sekä oppimiseen ja hyvinvointiin vaikuttavista tekijöistä. Tämä vaikeuttaa analyttistä ja tutkimustietoon pohjautuvaa dialogia, minkä seurauksena päätöksenteko ei aina perustu tutkimusnäyttöön, vaan osin vakiintuneisiin käsityksiin, omiin kokemuksiin ja asenteisiin.

Kansallisten kokeiden mahdollinen käyttöönotto edellyttäisi tutkimusperusteisen keskustelun lisäksi muiden maiden kokemusten huolellista tarkastelua sekä tietoista ja laadukasta järjestelmän suunnittelua.

KIRJALLISUUS

(Kirjallisuuskatsauksen kaikki viitteet löytyvät tutkimuksen liitteestä)

Ahn, Tom (2014) Regression discontinuity design estimates of the effect of high-stakes sanctions on schools. *Economics of Education Review*, 43, 203–216.

Alesina, Alberto & Michela Carlana & Eliana La Ferrara & Paolo Pinotti (2024) Revealing stereotypes: Evidence from immigrants in schools. *American Economic Review*, 114(7), 1916–1948.

Andersen, Simon Calmar & Nielsen, Helena Skyt (2020) Learning from performance information. *Journal of Public Administration Research and Theory*, 30(3), 415–431.

Barrett, Nathan & Toma, Eugenia F. (2013) Reward or punishment? Class size and teacher quality. *Economics of Education Review*, 35, 41–52.

Bensnes, Simon Sjøbstad (2016) You sneeze, you lose: The impact of pollen exposure on cognitive performance during high-stakes high school exams. *Journal of Health Economics*, 49, 1–13.

Bergbauer, Annika B. & Hanushek, Eric A. & Woessmann, Ludger (2024) Testing. *Journal of Human Resources*, 59(S), S118–S153.

Berry, James & Heissel, Jennifer A. & Kassem, Mariam & Mukherjee, Shobhini & Shotland, Marc (2020) Failure of Frequent Assessment: Deviations from the Curriculum and Learning in India. *Journal of Development Economics*, 145, 102464.

Beuchert, Louise & Eriksen, Tine Louise Mundbjerg & Krogh, Morten (2020) The Impact of Standardized Test Feedback on Remedial Education: Evidence from a Literacy Program. *Economics of Education Review*, 77, 101997.

Bishop, John (1998) The effect of curriculum-based external exit exam systems on student achievement. *The Journal of Economic Education*, 29(2), 171–182.

Bol, Thijs & Jacqueline Witschge & Herman G. Van de Werfhorst & Jaap Dronkers (2014) Curricular tracking and central examinations: Counterbalancing the impact of social background on student achievement in 36 countries. *Social Forces*, 92(4), 1545–1572.

Brunello, Giorgio & Kiss, David (2022) Math Scores in High-Stakes Exams. *Economics of Education Review*, 86, 102213.

Byun, Soo-yong (2010) Does policy matter in shadow education spending? Revisiting the effects of the high school equalization policy in South Korea. *Asia Pacific Education Review*, 11(1), 83–96.

Cerdeira, José Miguel & Luis Catela Nunes & Ana Balcão Reis & Carmo Seabra. (2018) Predictors of student success in Higher Education: Secondary school internal scores versus national exams. *Higher Education Quarterly*, 72(4), 304–313.

Deming, David J. & Figlio, David (2016) Accountability in US Education. *Journal of Economic Perspectives*, 30(3), 33–56.

- Deming, David & Cohodes, Sarah & Jennings, Jennifer & Jencks, Christopher (2016) School Accountability, Postsecondary Attainment, and Earnings. *The Review of Economics and Statistics*, 98(5), 848.
- Ebenstein, Avraham & Lavy, Victor & Roth, Sefi (2016) The Long-Run Economic Consequences of High-Stakes Examinations: Evidence from Transitory Variation in Pollution. *American Economic Journal: Applied Economics*, 8(4), 36–65.
- Falch, Torberg & Naper, Linn Renee & Rønning, Marte (2013) Educational Evaluation and Gender Differences. *Economics of Education Review*, 36, 12–22.
- Federicová, Miroslava & Münich, Daniel (2017) The impact of high-stakes school admission exams on study achievements: Quasi-experimental evidence from Slovakia. *Journal of Population Economics*, 30(4), 1069–1092.
- Fidjeland, Andreas (2023) Using High-Stakes Grades to Track Students: Evidence from a Grading Reform. *Economics of Education Review*, 95, 102434.
- Figlio, David & Susanna Loeb (2011) School accountability. *Handbook of the Economics of Education*, 3, 383-421.
- Fischer, Mira & Wagner, Valentin (2023) Do Timing and Reference Points Matter? *Economics of Education Review*, 95, 102438.
- Fuchsman, Dillon & Sass, Tim & Zamarro, Gema (2023) Testing, Teacher Turnover, and the Distribution of Teachers across Grades and Schools. *Society for Research on Educational Effectiveness. Education Finance and Policy*, 18(4), 654-675.
- Gil-Hernández, Carlos J. & Mar C. Espadafor (2025) Teacher bias by student SES or measurement error in ability? A cautionary tale for observational studies, *Social Science Research*, 103254.
- Goulas, Sofoklis & Megalokonomou, Rigissa (2021) Knowing Who You Are: The Effect of Feedback on Short- and Long-Term Outcomes. *Journal of Economic Behavior & Organization*, 183, 31–53.
- Heissel, Jennifer & Adam, Emma & Doleac, Jennifer & Figlio, David & Meer, Jonathan (2021) Testing, Stress, and Performance: How Students Respond Physiologically to High-Stakes Testing. *Education Finance and Policy*, 16(2), 183–208.
- Hofflinger, Alvaro & von Hippel, Paul (2020) Missing children: How Chilean schools evaded accountability by having low-performing students miss high-stakes tests. *Educational Assessment, Evaluation and Accountability*, 32(2), 127–152.
- Holmström, Bengt & Milgrom, Paul (1991) Multitask principal–agent analyses: Incentive contracts, asset ownership, and job design. *The Journal of Law, Economics, and Organization*, 7, 24–52.
- Hvidman, Ulrik & Sievertsen, Hans Henrik (2021) High-Stakes Grades and Student Behavior. *Journal of Human Resources*, 56(4), 1012–1043.
- Härkönen, Juho & Sirniö, Outi (2020) Educational transitions and educational inequality: A multiple pathways sequential logit model analysis of Finnish birth cohorts 1960–1985. *European Sociological Review*, 36(5), 700–719.

Högberg, Björn & Lindgren, Joakim & Johansson, Klara & Strandh, Mattias & Petersen, Solveig (2021) Consequences of School Grading Systems on Adolescent Health: Evidence from a Swedish School Reform. *Journal of Education Policy*, 36(1), 84–106.

Högberg, Björn & Lindgren, Joakim (2023) From a Crisis of Results to a Crisis of Wellbeing—Education Reform and the Declining Sense of School Belonging in Sweden. *Comparative Education*, 59(1), 18–37.

Jalava, Nina & Joensen, Juanna Schroter & Pallas, Elin (2015) Grades and Rank: Impacts of Non-financial Incentives on Test Performance. *Journal of Economic Behavior & Organization*, 115, 161–196.

Jürges, Hendrik & Schneider, Kerstin (2010) Central Exit Examinations Improve Student Performance: Evidence from TIMSS. *Journal of Population Economics*, 23(2), 655–682.

Jürges, Hendrik & Schneider, Kerstin & Senkbeil, Martin & Carstensen, Claus H. (2012) Assessment Drives Learning: The Effect of Central Exit Exams on Curriculum-Based Achievement and General Cognitive Ability. *Economics of Education Review*, 31(1), 56–65.

Karhunen, Hannu & Kuuppelomäki, Tiina & Sarvimäki, Matti & Suhonen, Tuomo & Uusitalo, Roope (2026) Perusopetuksen oppimistulokset ja mahdollisuuksien tasa-arvo, Eduskunnan tarkastusvaliokunnan julkaisu 1/2026.

Kauppinen, Merja & Marjanen, Jukka (2020) Millaista on yhdeksäsluokkalaisten kielellinen osaaminen? Suomen kielen ja kirjallisuuden oppimistulokset perusopetuksen päättövaiheessa 2019. Karvi. Julkaisut 13.

Ketonen, Laura & Atjonen, Päivi (2025) Mitä empiiristen tutkimusten perusteella tiedetään kansallisista kokeista? *Kasvatus*, 56(2), 238–248.

Kilpi, Nestori, Ouakrim-Soivio, Najat & Minkkinen, Laura & Söyrinki, Sini (2025) Oppimistuloskehitys ja koulutuspoliittiset toimet Pohjoismaissa ja Virossa 2010-luvulta 2020-luvun alkuun. Koulutuksen tutkimuslaitos, Raportteja ja työpapereita, 13.

Lavy, Victor & Sand, Edith (2018) On the origins of gender gaps in human capital: Short-and long-term consequences of teachers' biases. *Journal of Public Economics*, 167, 263–279.

Lazear, Edward P. (2006) Speeding, Terrorism, and Teaching to the Test. *The Quarterly Journal of Economics*, 121(3), 1029–1061.

Lehikoinen, Anita & Nokso-Koivisto, Oskari & Vihriälä, Vesa (2026) *Koulutuspolitiikan korjaussarja – kuopasta uuteen nousuun*. Sitra.

Lovenheim, Michael & Turner, Sarah (2018) *Economics of education*. Worth Publishers. New York.

Machin, Stephen & McNally, Sandra & Ruiz-Valenzuela, Jenifer (2020) Entry Through the Narrow Door: The Costs of Just Failing High Stakes Exams. *Journal of Public Economics*, 190, 104224.

Marks, Gary (2021) Should Value-Added School Effects Models Include Student- and School-Level Covariates? Evidence from Australian Population Assessment Data. *British Educational Research Journal*, 47(1), 181–204.

- Metcalfe, Robert & Burgess, Simon & Proud, Steven (2019) Students' Effort and Educational Resources: The Role of Incentives and Peers. *Journal of Public Economics*, 173, 93–107.
- Metsämuuronen, Jari (2023) Matematiikkaa COVID-19-pandemian varjossa III. Syventäviä analyysseja matematiikan 9. luokan arvioinnista keväällä 2021. Kansallinen koulutuksen arviointikeskus.
- Neves, Tiago & Pereira, Joao & Nata, Gill (2012) One-Dimensional School Rankings: A Non-Neutral Device that Conceals and Naturalises Inequalities. *International Journal on School Disaffection*, 9(1), 7–22.
- Ofek-Shanny, Yuval (2024) Measurements of Performance Gaps. *Journal of Human Resources*, 59, 1121-1150.
- Papay, John & Murnane, Richard & Willett, John (2016) The Impact of Test Score Labels on Human Capital Investment: Evidence from a Regression Discontinuity Design. *Journal of Human Resources*, 51(2), 357–388.
- Park, Jisung (2022) Hot Temperature and High-Stakes Performance. *The Journal of Human Resources*, 57(2), 400.
- Piopiunik, Marc & Schwerdt, Guido & Woessmann, Ludger (2013) Central School Exit Exams and Labor Market Outcomes: A European Perspective. *European Journal of Political Economy*, 30, 20–30.
- Riksrevisionen (2022) Statens insatser för likvärdig betygssättning – skillnaden mellan betyg och resultat på nationella prov (RiR 2022:22). Riksrevisionen.
- Rockoff, Jonah & Turner, Lesley (2010) Short-Run Impacts of Accountability on School Quality. *American Economic Journal: Economic Policy*, 2(4), 119–147.
- Sahlström, Ellen & Silliman, Mikko (2024) The extent and consequences of teacher biases against immigrants. Helsinki GSE Discussion Papers 22.
- Schwerdt, Guido & Woessmann, Ludger (2017) The Information Value of Central School Exams. *Economics of Education Review*, 56, 65–79.
- Sutter, Matthias & Zoller, Claudia & Glätzle-Rützler, Daniela (2019) Economic behavior of children and adolescents—A first survey of experimental economics results. *European Economic Review*, 111, 98-121.
- Tveit, Sverre (2019) Policy Legitimation of Education Assessment Reforms. The cases of Norway and Sweden. Series of dissertations submitted to the Faculty of Educational Sciences University of Oslo, (311).
- Van de Werfhorst, Herman G. & Mijs, Jonathan (2010) Achievement inequality and the institutional structure of educational systems: A comparative perspective. *Annual Review of Sociology*, 36(1), 407–428.
- Wallenius, Tommi (2020) Schools, performance and publicity: Contrasting the policy on publicising school performance indicators in Finland with the other Nordic countries. University of Helsinki Faculty of Educational Sciences. Helsinki Studies in Education, 92.

- Wang, Liang Choon (2016) The effect of high-stakes testing on suicidal ideation of teenagers with reference-dependent preferences. *Journal of Population Economics*, 29(2), 345–364.
- Wilson, Deborah & Burgess, Simon & Briggs, Adam (2011) The Dynamics of School Accountability: The Responses of English Secondary Schools to Performance Tables. *Journal of Population Economics*, 24(2), 407–432.
- Winters, Marcus A. & Trivitt, Julie R. & Greene, Jay P. (2010) The Impact of High-Stakes Testing on Student Proficiency in Low-Stakes Subjects: Evidence from Florida. *Economics of Education Review*, 29(1), 138–146
- Wuthrich, Viviana & Jagiello, Tess & Azzi, Vanessa (2020) Academic Stress in the Final Years of School: A Systematic Literature Review. *Child Psychiatry & Human Development*, 51(6), 986–1015.
- Wuthrich, Viviana & Belcher, Jessica & Kilby, Chris & Jagiello, Tess & Lowe, Catherine (2021) Tracking stress, depression, and anxiety across the final year of secondary school: A longitudinal study. *Journal of School Psychology*, 88, 18–30.

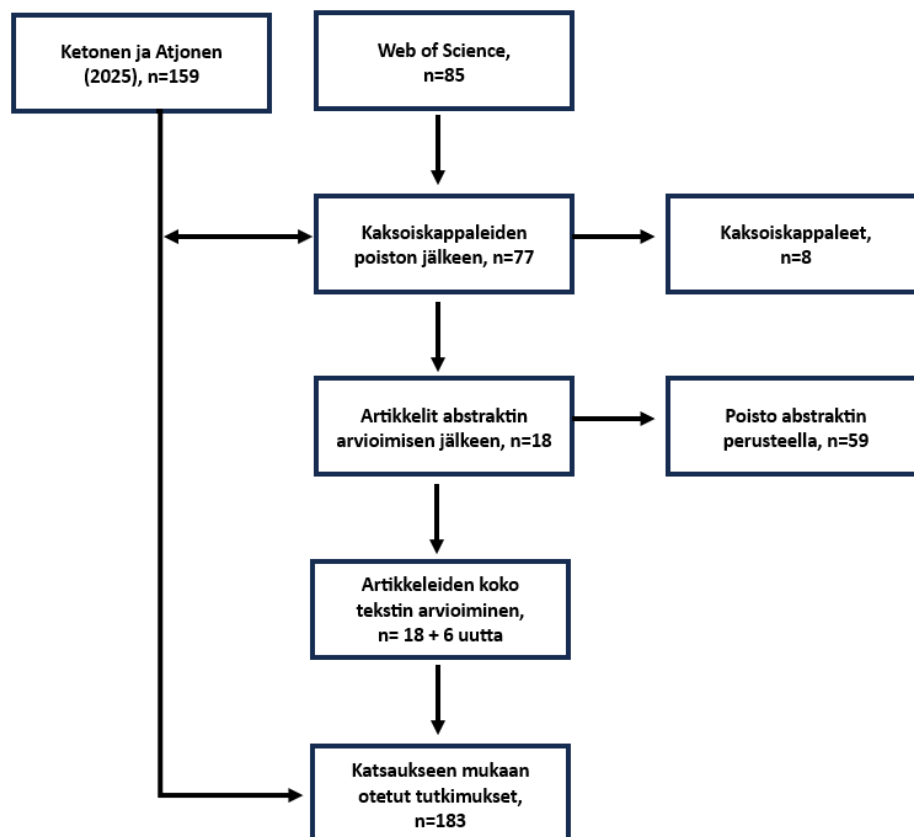
LIITE A. HAKUPROSESSIN EHDOT JA KULKU

Katsaus kattaa 1.1.2010–1.1.2025 julkaistut tutkimukset. Kun artikkeleita lopulta luettiin, niin löydettiin lisäksi viisi relevanttia tutkimusta luettujen tutkimusten käyttämistä viitteistä. Täydentävä haku toteutettiin 21.4.2025 Web of Science tietokannasta seuraaville ehdoilla:

TS=("national assessments" OR "high-stakes assessments" OR "state tests" OR "state testing" OR "standardised testing" OR "standardised tests" OR "high-stakes tests" OR "high-stakes testing" OR "high-stakes exams" OR "high stakes testing" OR "high-stakes grades" OR "national tests" OR "national testing" OR "national exams" OR "central exit exams") AND PY=(2010-2025) AND DT=(Article OR Review) AND WC=(Economics)

Ketonen ja Atjonen (2025) tarkastelevat empiirisiä tutkimuksia kansallisista kokeista perusopetus- ja lukioikäisillä. Mukana olivat muun muassa "high-stakes testing" ja "national testing", mutta ei termejä kuten "ranking list", tai "performance-based accountability". Näin ollen koulujen välisten rankingien sekä tulosvastuullisuuden ja valvonnan tutkimukset rajautuvat katsauksen ulkopuolelle, mikäli nämä eivät tule ilmi artikkeleiden lukemisen yhteydessä. Näitä teemoja koskee laaja kirjallisuus, joka edellyttää erillistä, laajempaa katsausta.

Täydellinen lista täydentävän haun 85 tutkimuksesta on saatavilla tutkijalta pyynnöstä.



Kuvio 1. Täydentävän hakuprosessin vuokaavio

Taulukko 1: Katsauksen vaatimukset täyttävät tutkimukset täydentävästä hausta (24 kappaletta)

	Tutkimus	Maa	Otoskoko ja ikäluokka	Arvioitava politiikka/toimenpide	Tutkimusmenetelmä	Mitattavat päävasteet	Tutkimuksen päätulokset
1	Jürges ym. (2010)	Saksa	TIMSS 1995: ~3 195 oppilasta (7.–8. lk), 566 opettajaa, 137 koulua	Keskitetyt päättökokeet (CEE) vs. hajautetut kokeet ala-asteen jälkeisessä koulutuksessa	Kontrollitutkimus, missä verrataan matematiikan ja biologian tuloksia CEE- ja ei-CEE-osavaltioissa	Matematiikan oppimistulokset (TIMMS), oppilaiden asenteet ja motivaatio sekä opettajien käytännöt	- Matematiikan pistemäärät nousivat ~0.11 SD, joka vastaa noin kolmasosaa kouluvuodesta. - Asenteet: oppilaat, joita päättökoe koski pitivät matematiikasta vähemmän, pitivät sitä tylsempänä ja vaikeampana. - Oppilaat käyttivät tilastollisesti enemmän aikaa opiskeluun kotona CEE-osavaltioissa. - Opettajakäytännöt: enemmän kotitehtäviä, useammin tarkistuksia ja käsittelyä tunneilla.
2	Winters ym. (2010)	USA, Florida	~163 000, 5. luokan oppilasta	Floridan koulujärjestelmä: Verrataan kouluja F-luokiteltuja kouluja D luokiteltuihin kouluihin	Kvasikokeellinen asetelma (RD), joka perustuu siihen, että koulujen luokittelut määräytyivät mekaanisella pisteytyssäännöllä	Matematiikan ja lukemisen tulokset (FCAT)	- Niissä kouluissa olevat oppilaat, jotka saivat juuri ja juuri huonon luokittelun, paransivat korkean panoksen lukemisen tuloksia ~0.9 SD ja matematiikan tuloksia ~0.20 SD. - Myös matalan panoksen tieteen testitulokset paranivat n. 0.10 SD. Ei näyttöä ”crowding-out”-ilmiöstä, vaan näyttöä laaja-alaisen oppimisen kasvusta.
3	Rockoff ym. (2010)	USA, NY	Koulutason aineisto: 985 koulua (luokat 3–8), n. 90 % alueesta 2007–2008. Lisäksi kysely vanhemmille.	Kouluille annettu A–F-vastuuarvosana, joka toi palkintoja (A/B) ja uhkasi sanktioilla (D/F).	Kvasikokeellinen asetelma (RD) Hyödyntää pistemäärärajoja A–F-arvosanoissa	Koulutason matematiikka ja englannin testitulokset	- Heikon arvosanan (D/F) saaneet koulut paransivat oppilaiden suorituksia matematiikassa ja englannissa (0.05–0.10 SD). - Vanhempien arviot koulun laadusta paranivat. - Ei näyttöä ”pelailusta” (esim. testiin osallistumisen manipulointi). - Muutoksia opetuskäytännöissä (vähemmän ryhmätyötä, enemmän testipainotteista opetusta).
4	Wilson ym. (2011)	Englanti	~563 000 7–11- ja ~467 000 11–16-vuotiasta	Dokumentoi vähemmistöjen koulusaavutusten kehitystä ja korkean panoksen GCSE-kokeiden merkitys	Kontrollitutkimus; regressiomallit ja pitkittäistutkimus, (value added analyysi)	Kansalliset testitulokset (englanti, matematiikka, luonnontieteet), GCSE-pisteet	- Kaikki etniset vähemmistöryhmät etenivät keskimäärin paremmin kuin natiivit oppilaat toisen asteen koulussa (ikä 11–16). - Suuri osa parannuksesta tapahtui juuri korkean panoksen GCSE-kokeissa
5	Jürges ym. (2012)	Saksa	~5 000 oppilasta (9.–10. luokkalaiset;	Keskitetyt päättökokeet (CEE) vs. ei-CEE	Kvasikokeellinen asetelma (DID), erot oppimistuloksissa	Matematiikan osaaminen (curriculum vs.	CEEs paransivat merkittävästi oppimäärään perustuvaa osaamista (+0.25 SD), mutta eivät vaikuttaneet matematiikan lukutaitoon (PISA).

			n. 15–16-vuotiaat) PISA-I-Plus aineisto		9. ja 10. luokan välillä (value added analyysi)	literacy), asenteet, motivaatio	• Positiivinen vaikutus löytyi vain ei-akateemisen linjan oppilaille. • Haittapuolena: oppilaat CEE-osavaltioissa raportoivat enemmän ahdistusta, epätoivoa, vähemmän iloa ja motivaatiota matematiikkaan.
6	Piopiunik ym. (2013)	Saksa	~5 500 (SOEP), 2.asteen opiskelijat	Keskitetyt vs. paikalliset koulun päättökokeet (CBEEE)	Kontrollitutkimus; regressiomalli, erot työmarkkinatulemis sa	Kuukausiansiot ja työttömyys	• Keskitetyt loppukokeet yhteydessä korkeampiin tuloihin (11 %) matalimman koulutusasteen (Hauptschule) valmistuneilla. • Keskikoulun (Realschule), lukion (Abitur) ja yliopistosta valmistuneilla ei vaikutusta ansioihin. • Työttömyys on matalampi: n. 4 %-yksikköä pienempi Hauptschule-oppilailla ja 3 %-yksikköä pienempi Abitur-oppilailla.
7	Falch ym. (2013)	Norja	n≈260 000 oppilasarviota, oppilaat 10. luokalta (noin ~ 130 000 16-vuotiaasta)	Arvioidaan opettaja-arvioinnin ja anonymien keskitettyjen päättökokeiden vaikutusta sukupuolten eroihin oppimistuloksissa	Kvasikokeellinen asetelma (DID), hyödynnetään saman oppilaan opettaja-arvosanan ja anonymin koe-arvosanan eroa	Kokeen ja opettaja-arvosanan pistemäärät kolmessa aineessa: matematiikka, englanti ja norja	• Tytöt saavat järjestelmällisesti paremmat opettaja-arvosanat kuin pojat suhteessa anonymeihin kokeisiin kaikissa aineissa (ero on 0.09 SD matematiikassa ja englannissa, 0.02 SD norjassa). • Erot eivät selity kilpailullisuudella tai anonymiteetillä. Opettajan sukupuolella ja kokemuksella on osittaisia yhteyksiä.
8	Ahn (2014)	USA	~40 000	Uusi valmistumisstandardi , jossa oppilaiden tuli läpäistä 5 high stakes End-of-Course (EOC) -koetta saadakseen high school -diplomin	Kvasikokeellinen asetelma (RD). Hyödynnetään Algebra I -kokeen jyrkkää pistemäärärajaa (cutoff)	Matematiikan kurssipolku: valitseeko oppilas Vaativamman matematiikan linjan	• Oppilaat, jotka juuri ja juuri epäonnistuivat Algebra I -kokeessa, olivat noin 5 %-yksikköä vähemmän jatkamaan vaativammalle matematiikan linjalle verrattuna niihin, jotka juuri läpäisivät kokeen. • Poliittikka ohjaa rajatapaussoppilaita pois college-valmiuksiin tähtäävältä matematiikan polulta.
9	Jalava ym. 2015)	Ruotsi	1045 oppilasta Tukholmasta, 6.-luokkalaisia (noin 12-vuotiaita), 47 luokkaa, 17 koulua	Ei rahalliset kannustimet matematiikan koetilanteessa. Neljä käsiteltyä interventiota: T1) A–F kriteeri perusteinen arvosana T2) Luokan sisäinen	Satunnaistettu kenttäkoe (RCT). Oppilaat satunnaistettiin viiteen ryhmään (T1–T4, ja kontrolliryhmä) luokkien sisällä.	Matematiikan testipistemäärä	• Ei-rahalliset kannustimet lisäävät testisuoritusta, mutta vain osalla oppilaista. • Kriteeriperusteinen A–F -arvosana ei lisää suoriutumista. • Kilpailuasetelmat ovat tehokkaimpia, Top-3 ranking (T2) ja Top-3 + palkinto (T4) nostavat testituloksia taitavien oppilaiden tuloksia (0.33–0.45 SD). • Pojat reagoivat kilpailuun (rank-based T2 ja T4)

				ranking T3) Kriteeri perusteinen todistus T4) Luokan sisäinen kilpailu + palkinto			• Tytöt reagoivat sekä kilpailuun että todistukseen. • Keskitason ylemmät oppilaat: kannustimet lisäävät suoritusta, kun taas alimmat 10 % oppilaista kannustimet voivat heikentää suoritusta.
10	Papay ym. (2016)	USA	~1 200–8 300 mallispesifika- tion mukaan; 8. ja 10. luokkalaiset	Standardoiduissa matematiikan kokeissa (alhaisen panoksen) oppilaille annettavat suoritusluokitukset. Nämä etiketit eivät tuoneet virallisia seuraamuksia.	Kvasikokeellinen asetelma (RD) hyödynnetään pistemäärän tarkkoja rajoja, jotka määrittävät labelin, mutta eivät kanna virallisia seuraamuksia.	Valmistuminen, korkeakouluun hakeutuminen vuoden sisällä valmistumisesta, 10. luokan matematiikan pistemäärät, Jatkokoulutussuunn itelmat.	• “Advanced” verrattuna “Proficient” etiketti 10. luokan matematiikassa nosti vähävaraisten oppilaiden korkeakouluun hakeutumisen todennäköisyyttä 5 %-yksiköllä. • “Needs Improvement” vs. “Warning” etiketti lisäsi hieman jatko-opintoihin hakeutumista, paransi 8.luokan matematiikan pistemääriä (0.032 SD) ja valmistumistodennäköisyyttä. Vaikutukset olivat suurimmat niillä oppilailla, jotka eivät alun perin suunnitelleet nelivuotisia jatko-opintoja. • Proficient vs. Needs Improvement -katkokohdan nollahavainto.
11	Ebenstein ym. (2016)	Israel	~415 000 koesuoritusta ja 55 796 oppilasta, 626 koulua.	Korkean panoksen ylioppilaskokeet (Bagrut) ja altistuminen PM2.5-hiukkasille koepäivänä vaikuttaa koetuloksiin ja myöhempiin koulutus- ja työmarkkinatulemiin.	Kvasikokeellinen asetelma; regressio missä opiskelijakiinteät vaikutukset. Sama opiskelija tekee 8– 10 koetta eri päivinä ja oppilaat altistuvat eri saastetasolle riippuen valituista aineista ja koeajankohdista.	Bagrut-koe (0–100), kokeen läpäisy, korkeakoulutus, kuukausiansiot.	• Saaste heikentää koemenestystä selvästi (10 yksikköä PM2.5 hiukkasia laskee koepisteitä 0.40 pistettä). • Suurin vaikutus pojille, heikoille opiskelijoille ja matalan SES-taustan oppilaille. • Laskee todennäköisyyttä saada todistus ja aloittaa kilpailullisessa korkeakoulutuksessa (ei näy ei-kilpailullisessa) ja tämä lopulta näkyy negatiivisena vaikutuksena palkkaan 28 vuoden iässä.
12	Schwerdt ym. (2017)	Saksa	~4 700, seurataan yliopistosta v. 1997 valmistuneita, jotka suorittivat ylioppilaskoke	Keskitetyt valtakunnalliset Abitur-kokeet vs. koulukohtaiset kokeet.	Kvasikokeellinen asetelma (DID), joka hyödyntää osavaltioiden eroja kokeiden keskittämisessä	Kuukausiansiot, työllisyys, työnhakuprosessin tulokset 5 vuotta valmistumisen jälkeen	• Yksi standardipoikkeaman parannus arvosanoissa lisää 6 % ansioita keskitettyjen kokeiden osavaltioissa vs. +1.6 % ansioiden kasvu paikallisiin kokeisiin osallistuneilla. • Vaikutukset suuremmat yksityisellä sektorilla. • Keskitetyt kokeet antavat tarkemman signaalin tuottavuudesta.

			en eri järjestelmissä.				
13	Metcalfe ym. (2019)	Englanti	~ 600,000 opiskelijaa vuodessa	Jalkapalloturnauksen (MM/EM) ajoitus suhteessa ylioppilaskokeisiin	Kvasikokeellinen asetelma, (DID). Turnausvuodet vs. ei-turnausvuodet, sekä kokeet turnauksen aikana vs. sen ulkopuolella	GCSE-arvosanat, tutkintorajan saavuttaminen	• Turnauksen aikana tehtyjen kokeiden tulokset heikompia (−0.017 SD keskimäärin, −0.065 SD jalkapallosta kiinnostuneilla). • Mahdollisuus saavuttaa tutkintoraja laskee 6 %-yksikköä. • Vaikutus suurempi pojilla, erityisesti valkoihoisilla ja matalan SES-taustan oppilailla. Jalkapallosta kiinnostuneilla vetosuhde laskee jopa 28%.
14	Andersen ym. (2020)	Tanska	2.–8. luokan oppilaat lukuvuosina 2009–2010. Oppimistulokset mitataan 2–4 v. myöhem., 136 887 oppilasta lukutestissä ja 46 338 oppilasta matematiikan testissä.	Kansallisten kokeiden käyttöönotto. Ennen tätä Tanskassa oli paikalliset kokeet, mutta Tanskan hallitus päätti 2006 ottaa käyttöön kansalliset kokeet eri luokille.	Kvasikokeellinen asetelma (IV). Kansallisen IT-testausjärjestelmän odottamaton kaatuminen maaliskuussa 2010 esti osaa oppilaista osallistumasta pakolliseen testiin (testauksen puuttuminen on "as-good-as-random").	Standardoidut lukutestin tulokset 2 ja 4 vuotta myöhemmin, matematiikan tulokset 3 vuotta myöhemmin ja testausvelvoitteen Koulutason noudattaminen	• Kansallinen testaus parantaa oppimista merkittävästi. Lukutesti +0.1 SD kahden vuoden kuluttua, matematiikan testi +0.07 SD kolmen vuoden kuluttua (ei aina tilastollisesti merkitsevä pienemmän otoksen vuoksi). • Vaikutus lukemiseen säilyy neljän vuoden kuluttua (0.07 SD) • Ei viitteitä haitasta heikoille oppilaille, vaan päinvastoin. Alhaisen SES, matalan tulotason ja maahanmuuttajataustaisten oppilaiden vaikutukset ovat selvästi suurempia. • Koulut, joissa on paljon alhaisen SES tai maahanmuuttajataustaisia oppilaita, jättävät testit useammin tekemättä.
15	Beuchert ym. (2020)	Tanska	192 577 oppilasta (kolmen 3.luokan kohortin pitkittäisseuranta)	Vanhemmille annettava standardisoitu testipalaute 3. luokan matematiikassa.	Kvasikokeellinen asetelma (RD); Hyödynnetään epäjatkuvuutta neljän luokkarajan ympärillä, joiden mukaan palaute annetaan.	4. ja 6. lk:n lukeminen ja 6 lk:n matematiikka; akateeminen itsevarmuus, sisäinen motivaatio, somaattiset oireet, sairauspoissaolot (4. lk.), oppilaan kokemus opettajan tuesta (6. lk.).	• Negatiivinen palaute parantaa myöhempää oppimista. 3. luokan matematiikan "negatiivinen" luokkasignaali nostaa +0.061 SD 6. luokan matematiikan tulosta (lisäksi +0.045 SD 4. luokan lukemisen tulosta). • Suurimmat vaikutukset heikoimmassa päässä ja pojilla, mutta vaikutukset eivät riipu voimakkaasti perhetaustasta. • Ei näyttöä haitoista. Hyvinvointimittarit eivät viittaa systemaattisiin negatiivisiin vaikutuksiin; alimmassa päässä akateeminen itsevarmuus kasvaa, self-efficacy ja sisäinen motivaatio eivät heikkene.

							• Oppilaiden kokema opettajatuki ei muutu, mikä viittaa siihen, että vaikutukset kulkevat ensisijaisesti vanhempien ja/tai oppilaan oman panostuksen kautta.
16	Berry ym. (2020)	Intia	~15 000 oppilasta, 500 koulussa, joista 400 ala- ja 100 yläkoulua	Jatkuva ja kokonaisvaltainen arviointi (CCE) vs. tavanomainen arviointi	Satunnaistettu kontrolloitu koe (RCT), missä vertaillaan CCE:n ja kontrolliryhmän oppimistuloksia	Testitulokset (matematiikka ja kieli), opettajien käytännöt, koulunkäynti	• CCE ei parantanut oppimistuloksia. Ei vaikutusta testituloksiin. • Opettajat lisäsivät arviointimenetelmiä, mutta eivät muuttaneet opetuskäytäntöjä.
17	Machin ym. (2020)	Englanti	~ 49 000 16-vuotiaasta opiskelijaa	Juuri ja juuri reuttaminen vs. läpäiseminen GCSE English -kokeessa	Kvasikokeellinen asetelma (Fuzzy-RD); Verrataan henkilöitä läpäisyrajan tuntumassa.	Opintojen keskeyttäminen, NEET-status, pääsy ja suoritus 2. asteen koulutuksessa, korkeakouluun pääsy	• Juuri ja juuri reuttaminen GCSE English -kokeessa johtaa merkittäviin negatiivisiin vaikutuksiin koulutuspolulla: 4 %-yks. lisäys keskeyttämiseen, 2 %-yks. lisäys NEET-statuksessa, 6–9 %-yks. lasku pääsyssä 2. asteelle, ja 2.5–4 %-yks. korkeakouluun. • Vaikutukset merkittäviä ja pitkäkestoisia.
18	Goulas ym. (2021)	Kreikka	~41 000 lukiolaista 115 koulussa	Suhteellisen suoritustiedon (rankin) antaminen vs. ei antaminen 11. luokalla	Kvasikokeellinen asetelma (DID); Kreikassa poistettiin rankkitieto osasta aineita. Tämä mahdollisti vertailun ennen ja jälkeen -ryhmien välillä.	12. lk:n arvosanat, yliopistovalinnat, odotetut tulot.	• Palautteen saaminen parantaa huippuoppilaiden suorituksia (+0.15 SD), yliopistovalintoja (+0.11 SD) ja odotettuja tuloja (+0.10 SD). • Heikkojen oppilaiden tulokset heikkenevät (–0.30 SD, –0.31 SD, –0.21 SD). • Naiset reagoivat negatiivisemmin palautteeseen.
19	Hvidman ym. (2021)	Tanska	~27 000 lukio-opiskelijaa	Arvosanauudistus, missä vanhat 13 asteikon arvosanat muunnettiin uuteen 7-pisteasteikkoon (Reformi ei antanut uutta tietoa oppijasta, muutti vain kannustimia)	Kvasikokeellinen asetelma (DID); Hyödyntää muutosta arvostelussa ilman, että se kertoi mitään todellisesta osaamisesta	Seuraavat arvosanat 2. ja 3. lukiovuonna, kansalliset kokeet, työssäkäynti, yliopistoon hakeutuminen ja valmistuminen	• Negatiivinen arvosanashokki johti parempiin myöhempiin arvosanoihin (+0.08 SD), ja vaikutus näkyy myös ulkoisissa kokeissa. • Kasvatti todennäköisyyttä valmistua yliopistosta (+2 %-yks.). • Vaikutus suurempi naisilla ja hyvän peruskoulun keskiarvon omaavilla.
20	Brunello ym. (2022)	Saksa	~7 900 alakoulun (7–19v.) ja ~5 800 yläkoulun (11–	Korkean (high stakes) vs. matalan panoksen (low stakes) testausikä	Kvasikokeellinen asetelma, (DID); Hyödynnetään sitä, että oppilaat ovat	Matematiikan testitulokset peruskoulusta 1.2. ja 4. luokka ja 5., 7.	• Korkean panoksen luokassa testatut oppilaat saavat parempia tuloksia 0.17–0.23 SD. Tämä on iso vaikutus, sillä yhden kouluvuoden oppimisen arvioitu vastaavuus on 0.25–0.33 SD.

			15v.) oppilasta (NEPS)		eri luokka-asteilla testin aikana riippuen osavaltiosta.	ja 9. luokan toistuvat testit.	Vaikutus suurempi tytöillä ja viitteitä siitä, että vähemmän koulutettujen perheiden lapset hyötyvät eniten.
21	Fidjeland (2023)	Norja	Peruskoulun ~791 000 oppilasta (10. luokka)	Arvosanoihin perustuva kouluvalinta (ns. high stakes) vs. lähikoulu malli. Uudistusten vuoksi 10. luokan lopputesti on high-stakes vain niissä maakunnissa, joissa valintauudistus toteutettiin.	Kvasikokeellinen asetelma (TripleDID); Hyödyntää eroja ajassa, alueissa ja koulutarjonnassa, joka syntyy siitä, että reformi toteutettiin eri vuosina eri maakunnissa	Päättökokeen tulokset ("high-stakes") ja kansalliset arviointitulokset, (ns. "low-stakes") kokeet, sukupuoli- ja SES-erot	- Korkean panoksen kokeet paransivat testituloksia (4-6 % SD) erityisesti alueilla, joissa oppilailla on todellinen kouluvalinnan mahdollisuus. - Vaikutus näkyy myös matalan panoksen kansallisissa arviointituloksissa (0.07 SD) mikä viittaa oppimisen lisääntymiseen, ei pelkkään testiponnisteluun. - Ei merkittäviä eroja sukupuolen tai sosioekonomisen taustan mukaan, viitteitä siitä että vaikutus voisi olla vahvempi matematiikassa heikon SES:n oppilailla.
22	Fischer Ym. (2023)	Saksa	7 koulua, 330 5.–6. luokan oppilasta (ikä 11–12).	Oppilaille annettiin henkilökohtainen suoritus palaute muuttamalla kokeellisesti palauteajankohtaa ja sen "kehystä".	Satunnaistettu kenttäkoe (RCT); kahtena ulottuvuutena (ajankohta ja palautteen kehys)	Loppukokeen matematiikkatulokset, itseluottamus, ponnistelun vaikuttavuuden uskomus, minäpystyvyys.	- Varhainen palaute parantaa (+0.224 SD), myöhäinen heikentää (–0.126 SD) matematiikan tulosta. - Palaute, joka kertoo negatiivisesta osaamisen muutoksesta, on kaikkein voimakkain. Aikaisin annettuna se tehostaa valmistautumista, myöhään annettuna se romahduttaa testisuorituksen. - Erityisesti pojat reagoivat voimakkaasti palautteesta.
23	Ofek-Shanny (2024)	Israel	577 oppilasta, 7 koulua, joista osa arabien ja osa juutalaisten.	Testin panostason vaikutus suorituserojen mittaamiseen	Satunnaistettu kenttäkoe (RCT); sama oppilas tekee sekä matalan että korkean panoksen kokeen.	Testitulokset (matematiikka), testimotivaatio, erot sukupuolen ja etnisyyden mukaan	- Arabien ja juutalaisten välinen ero matalan panoksen testissä 0.38 SD, mutta korkean panoksen testissä vain 0.15 SD. - Erot johtuvat osin motivaatiosta, ei osaamisesta, viitaten siihen, että matalan panoksen testit aliarvioivat vähemmistöjen osaamista.
24	Bergbaue r ym. (2024)	59 maata	PISA-aineisto vuosilta 2000–2015. Yli 2 miljoonaa oppilasta. 6 PISA aaltoa (2000, 2003,	Vertaillaan neljää erilaista koe- ja valvontajärjestelmää	Kvasikokeellinen asetelma DID-regressiomalli missä, jossa hyödynnetään maiden pitkittäisiä	PISA-tulokset (matematiikka, lukutaito, luonnontiede)	- Standardoidut ulkoiset testit (SCOMP) nostavat oppimistuloksia, mutta vain maissa, joissa lähtötaso on alun perin matala tai keskitasoinen. - Standardoidut sisäiset testit (SINT) toimivat vain heikommin suoriutuviissa maissa – korkealla tasolla jopa haitallisia.

			2006, 2009, 2012, 2015).		muutoksia koejärjestelmissä		Sisäinen raportointi (IRPT) ei vaikuta oppimistuloksiin, mutta opettajien valvonta voi heikentää oppimistuloksia korkean suorituskyvyn maissa.
--	--	--	-----------------------------	--	--------------------------------	--	--

Taulukko 2: Ketosen ja Atjosen (2025) katsauksen tutkimukset (159 kappaletta)

	Tutkimus	Maa	Otoskoko ja ikäluokka	Arvioitava politiikka/toimenpide	Tutkimusmenetelmä	Mitattavat päävasteet	Tutkimuksen päätulokset
1	Abdallah ym. (2022)	Nigeria	N = 6 349 vastaajaa: 5 600 lukiolaista (20 v ja alle) ja 1 120 luonnontieteiden opettajaa	Korkean panoksen -kokeet ja niiden vaikutus opettajien opetustyyliin	Ei verrokkiryhmää (kysely)	Washback-vaikutus, testipaine, testiahdistus; opettajien opetustyyli	- Korkean panoksen -kokeilla oli merkittävä yhteis- ja yksittäisvaikutus opettajien opetustyyliin. - Suosituin opetustyyli oli asiantuntijan autonomia. - Opettajien ja oppilaiden välillä oli tilastollisesti merkitsevä ero näkemyksissä kokeiden vaikutuksista.
2	Acikalin (2014)	Turkki	Ei otosta, teoreettinen	Uusi sosiaalitiedon opetussuunnitelma	Ei verrokkiryhmää (vertailu- ja keskusteluanalyysi Turkin ja Yhdysvaltojen sosiaalitiedon opetuksesta sekä kirjallisuuskatsaus)	Opetussuunnitelman rakenne, arviointimenetelmät standardoitujen testien vaikutukset oppiaineen sisältöön ja opetuskäytäntöihin	- Uusi opetussuunnitelma painottaa taitoja ja arvoja, mutta koulujen käytäntö on edelleen testikeskeinen. - Standardoidut testit dominoivat ja ovat ristiriidassa taitoperustaisen opetuksen kanssa. - Vaikutuksena opetuksen sisällön supistuminen ja opetuksen siirtyminen testivalmisteluksi.
3	Aksoy (2023)	Turkki	109 yläkoulun opettajaa, 6 eri oppiainetta	Korkean panoksen -kokeiden ja opetussuunnitelman sekä koulutuspolitiikan välinen vastaavuus; siirtymä peruskoulusta toisen asteen koulutukseen	Ei verrokkiryhmää (kysely)	Opettajien näkemykset Korkean panoksen -kokeiden ja opetussuunnitelman yhtenevyydestä, oppikirjojen ja kokeiden vastaavuudesta	- Tulokset osoittavat, että opetus on testikeskeistä ja koulutusjärjestelmä palvelee ensisijaisesti kokeita ei oppimista. - Monissa aineissa kokeet ovat osittain tai huonosti linjassa opetussuunnitelman ja oppimateriaalien kanssa.
4	Alsagoafi (2021)	Saudi-Arabia	103 lukiolaista	Kansalliset korkean panoksen kokeet ja niiden niiden washback-vaikutus	Ei verrokkiryhmää (puolistrukturoidut kyselyt ja haastattelut)	Opiskelijoiden käsitykset kokeiden vaikutuksista, koettu reiluus, testien vaikutus oppimiseen ja	Positiivista washbackia: vähäisempi koepaine, mahdollisuus uusien testien parantamiseen, painotus

						taitoihin, testien antama paine	kriittiseen ajatteluun, ei pelkkään ulkoa opetteluun. Opiskelijat kokivat testit reiluiksi ja pääsyprosessin aiempaa oikeudenmukaisemmaksi, sillä testit tasoittivat opettajan tekemiin kokeisiin liittyvää harhaa.
5	Amrein-Beardsley (2010)	USA, Arizona	3 085 vapaaehtoista opettajaa (pääosin peruskoulu- ja yläkoulun opettajia). Lisäksi 9 opettajaa haastatteluissa ja 8 opettajaa fokusryhmässä	No Child Left Behind -aikakauden korkean panoksen testit. Tavoitteena selvittää, millaisia huijaamisen muotoja opettajat käyttävät ja miten yleisiä ne ovat	Ei verrokkiryhmää (valtakunnallinen online-kysely; teemahaastattelut; fokusryhmä; laadullinen sisällönanalyysi)	Opettajien tietoisuus kollegoiden testihuijaamisesta ja itse raportoitu huijaaminen. Huijaamisen "asteet" 1. aste = tarkoituksellista, 2. aste = tilannesidonnaista, 3. aste = tahatonta/huomaamaton;	- Huijaamista esiintyy kaikilla kolmella "asteella". Yli 50 % vastaajista tiesi kollegoiden syyllistyneen huijaamiseen, ja yli puolet raportoi jonkin asteista huijaamista itse. - Ensimmäisen asteen huijaaminen (esim. vastausten korjaaminen) oli harvinaisempaa, kun taas toisen ja kolmannen asteen huijaus (esim. vihjeiden antaminen, testiin valmentaminen liian tarkasti) oli yleisempää. - Korkean panoksen testit ja rankaisivat rakenteet lisäsivät merkittävästi huijaamisen todennäköisyyttä.
6	Avalos (2020)	USA	4 toisen asteen englannin opettajaa	Korkean panoksen testaus ja sen aiheuttama opetussuunnitelman kaventuminen; vaikutus kirjoittamisen opetukseen moninaisille oppijoille	Ei verrokkiryhmää (19 videotallennettua kirjoitusoppituntia, 8 haastattelua)	Opettajien ihanteelliset vs. todelliset kirjoittamisen opetuskäytännöt; opetuksen rajoitukset; opettajien autonomia	- Opettajat halusivat antaa enemmän aikaa, vähemmän testipainetta - Opettajat tunnistavat oppilaiden tarpeet mutta eivät voi toimia niiden mukaisesti, koska aikataulut, standardit ja vastuullisuusjärjestelmät rajoittavat valintoja.
7	Banks ym. (2015)	Irlanti	748 viimeisen vuosikurssin opiskelijaa 12 eri toisen asteen	Korkean panoksen -kokeet, niiden stressivaikutukset sekä koulukontekstin rooli	Ei verrokkiryhmää (kysely ja fokusryhmä-haastattelut)	Akateeminen stressi, aiempi koulumenestys, opettaja-oppilas- vuorovaikutus, kiusaaminen,	Suurin stressitekijä on kokeiden koettu "koko elämän ratkaisuus". Stressi liittyy sekä

			koulusta; 53 fokusryhmää. Ikäluokka: 16– 18-vuotiaat	stressin lisääjänä tai lieventäjänä		koulun ilmapiiri, kurssivalinnat, työkuorma, tulevaisuushuoli	välittömiin paineisiin että pitkän aikavälin riskeihin. Tytöt ja hyvin menestyvät oppilaat raportoivat selvästi korkeampaa stressiä kuin pojat tai ”alisuoriutujat”.
8	Barrett ym. (2013)	USA, Kentuck y	~23 000 oppilasta, peruskoulu ja lukio, 10 koulupiiriä, Oppilas opettaja- paneeli 2000–2008	Käyttävätkö rehtorit luokkakokoa strategisena välineenä opettajien aiemman mitatun tehokkuuden perusteella	Kontrollitutkimus (kaksivaiheinen analyysi: opettajan tehokkuus estimoidaan teacher value-added - mallilla ja selitetään tuloksella luokkakokoa)	Luokkakoko, oppilaiden oppimistulokset, opettajan tehokkuus.	- Tehokkaat opettajat saavat systemaattisesti suurempia luokkia. Tämä viittaa siihen, että rehtorit pyrkivät maksimoimaan koulun kokonaissuoriutumisen sijoittamalla enemmän oppilaita parhaiden opettajien ryhmiin. - Luokkakoko ei ole tästä syystä eksogeeninen muuttuja ja aiempi luokkakokotutkimus voi olla vinoutunutta.
9	Bensnes (2016)	Norja	~ 69 000 koesuoritusta ja ~25 000 lukiolaista (n. 16–19-v.), vuosilta 2008– 2011	Miten siitepölyaltistus kokeiden aikana vaikuttaa kognitiiviseen suoriutumiseen	Kvasikokeellinen asetelma (ennen- jälkeen) (hyödynnetään sitä, että sama opiskelija tekee kokeita eri päivinä eri siitepölytasojen aikana)	Koetulokset (standardoidut arvosanat) ja jatko- opintovalinnat	- Siitepölyn lisääntyminen heikentää suoriutumista kokeissa (+1 keskihajonta siitepölyä aiheuttaa 2.5 % SD lasku testituloksissa) - Vähentää merkittävästi opiskelijoiden todennäköisyyttä valita STEM-ala.
10	Bolden ym. (2020)	Englanti, Yhdysval lat + TIMMS ja PIRLS maat	Aggregaatti suureet Englanti, USA, TIMMS ja PIRLS maat	Laajat koulutusreformit ja tulosvastuullisuus (kansalliset testit, opetussuunnitelmauudi stukset, tarkastusjärjestelmät, No Child Left Behind (USA))	Ei verrokkiryhmää (maakohtaisten aikasarjojen ja niiden trendien esittäminen graafisesti ja argumentointi miten ne ovat vaikuttaneet)	Oppimistulosten kehitys (lukeminen, matematiikka, luonnontieteet)	Oppimistulokset ovat pysyneet yllättävän vakaina pitkällä aikavälillä suurista investoinneista ja reformeista huolimatta, joka tutkimuksen mukaan osoittaa, että koulutusreformit eivät ole nostaneet oppimista globaalisti.
11	Burton (2012)	USA	23 palkittua opettajaa, joilla yli 10 vuoden opetuskokemus	Tavoite oli selvittää, kuinka hyvin opettajien kokemukset ja näkemykset vastaavat luonnontieteiden	Ei verrokkiryhmää (verkkokysely, jossa 23 opettajaa vastasi 28 avoimeen kysymykseen ja sisällönanalyysi)	Opettajien vastausten yhteensopivuus reformiraporttien sisältöjen kanssa seuraavissa teemoissa.	- Reformiraporttien ja opettajien näkemyksen vastaavuus oli keskimäärin kohtalainen - Korkean panoksen kokeet vaikuttavat haitallisesti

				opetuksen reformiraporttien reformitavoitteita			tutkimuksellista opetusta ja oppilaiden uteliaisuuden tukemiseen.
12*	Byun (2010)	Etelä-Korea	~6 000 perhe–oppilasparia, 7.–9. luokalla (peruskoulun yläaste, 13–15-v.)	Etelä-Korean vuodesta 1974 alkaen käyttämä uudistus, joka poisti lukion korkean panoksen yhteisvalintakokeet ja otti käyttöön satunnaisen sijoittelun oppilaiden kotialueen lukioihin	Kontrollitutkimus Samankaltaistaminen (Propensity Score Matching)	Kuukausittaiset kotitalouksien menot varjokoulutukseen	- Varjokoulutusmenot vähenivät reformin ansiosta - Vaikutus on suurin matalan SES:n perheille, lieventäen koulutuksellista eriarvoisuutta.
13	Cairns (2021)	Australia	15 historian opettajaa ja opettajien yhdistyksen edustaja	Tarkastelee standardisoituja loppuvuoden ylioppilastyyppejä kokeita	Ei verrokkiryhmää (kriittinen politiikka-analyysi ja Diskurssianalyysi)	Analysoidaan laadullisesti, miten kokeiden vaatimukset ohjaavat opetusta, mahdollista Exam washback – kokeiden vaikutusta, opettajien autonomiaa, työhyvinvointi ja toimijuutta, opetussuunnitelmien kaventumista.	- Opettajien mukaan korkean panoksen kokeet ohjaavat opetusta voimakkaasti - Pedagoginen luovuus ja opiskelijoiden syvälinen historiallisen ajattelun kehittäminen kärsivät - Kokeet lisäävät eriarvoisuutta sillä hyvin resursoitujen koulujen pystyvä hyödyntämään intensiivistä opetusta.
14	Camphu ijsen ym. (2023)	Chile ja Norja	Kyselyaineisto: Chilestä 1 225 ja Norjasta 1 306 perusopetuksen opettajaa Haastattelut: Chilestä 28 opettajaa ja Norjasta 13 opettajaa	Tarkastelee standardoidut testejä ja testipohjaista tulostavasta opettajien näkökulmasta	Ei verrokkiryhmää (kyselyt ja haastattelut)	Opettajien uskomukset ja käsitykset testien validiteetista, hyödyllisyydestä opetuksen näkökulmasta ja oikeudenmukaisuus.	- Opettajat molemmissa maissa suhtautuvat kriittisesti testeihin: testit eivät mittaa oppimista kunnolla, eivät kuvaa opettajan laatua, koetaan epäoikeudenmukaisiksi. - Chilessä opettajat kriittisempiä testien hyödyllisyydestä ja vaikutuksista työhön, kun taas - Norjassa kritiikki taas testitulosten julkisesta levittämisestä.
15	Camphu ijsen	Norja	3 046 artikkelia, 155 alueellisesta ja	Valtakunnalliset standardoidut kokeet	Ei verrokkiryhmää (sisällönanalyysi)	Median kehukset	Media korostaa kilpailua, vertailua ja tulostavasta

	ym. (2022)		paikallisesta sanomalehdestä väliltä 2004–2018	osana testipohjaista vastuullisuus järjestelmää Norjassa			• Vertailee kouluja ja tekee ranking-listoja • Alussa enemmän kritiikkiä, alueellisia eroja.
16	Camphu ijsen ym. (2021)	Norja	37 haastattelua Ja 4 keskeistä politiikkadokumentaattia	Testipohjainen tulosvastuu osana koulutusreformia ja hallintomallin muutosta	Ei verrokkiryhmää (politiikka-analyysi)	Politiikan kehitys, käyttöönoton syyt, odotukset ja perustelut, käytetyt ohjauskeinot	• PISA-tulokset (2000) toimivat ”shokkina”, joka käynnisti reformin. • Tulosvastuu institutionalisoitui nopeasti, vaikka oli aiemmin kiistanalainen (mekanismeina näkyvyys, vertailu, hallinnollinen kontrolli) • Koulutuksen laatu ja tasa-arvo on yhä enemmän määritelty testitulosten kautta.
17	Carter (2019)	USA	6 opettajaa, alakoulu (3.lk,8–9-vuotiaat)	Korkean panoksen -kokeet ja koulutusjärjestelmän rakenteellinen arvio	Ei verrokkiryhmää (monitapaustutkimus)	Opetuskäytännöt ja pedagogiikka, oppilaiden kokemus ja osallistuminen ja koulutuspolut	• Standardoidut testit ohjaavat opetusta kapeaksi, Oppilaita luokitellaan ja leimataan testitulosten perusteella ja erityisesti vähemmistöoppilaat kärsivät.
18*	Cerdeira ym. (2018)	Portugali	23 632 korkeakouluopiskelijaa	Korkeakouluvalinnan kriteerit: kumpi ennustaa paremmin opiskelumenestystä lukion sisäiset arvosanat vai kansalliset pääsykokeet	Kontrollitutkimus (regressio, value-added lähestymistapa)	Bachelor-tutkinnon lopullinen keskiarvo	• Lukion sisäiset arvosanat ennustavat paremmin opiskelumenestystä kuin kansalliset kokeet. • Vanhempien koulutustaso ennustaa vain vähän kun suoritus vakioitu.
19	Clarke 2023	Englanti	18 oppilasta yhdestä koulusta (14–15-v.)	Oppilaiden kokemukset tulosvastuun vaikutuksesta hyvinvointiin	Ei verrokkiryhmää (ryhmähaastattelut)	Hyvinvointi, tunteet, mieliala, merkityksellisyys	• Hyvät arvosanat yhdistyvät oppilaiden kokemuksissa hyvinvointiin. • Testikulttuuri aiheuttaa painetta, stressiä ja ahdistusta.
20	Coldwell ym. (2017)	Englanti	20 alakoulua, 40 alakoulun rehtoria ja 20 yläkoulun johtajaa sekä	Korkean panoksen kansalliset kokeet jotka liittyvät siirtymään alakoulusta yläkouluun	Ei verrokkiryhmää (tapaustutkimukset, haastattelut ja fokusryhmät oppilaille)	Opettajien ja johtajien käsitykset testien roolista, oppimistulosten luotettavuudesta, vaikutuksesta	• Tuloksellisuus johtaa opetuksen kaventumiseen ja kokeeseen opettamiseen.

			oppilasfokusryhmiä			opetukseen ja opetussuunnitelmaan	• Yläkoulut usein testaavat oppilaat uudelleen koska eivät luota alakoulujen mittauksiin.
21	Colombo ym. (2013)	USA	20 yläkoulun opettajaa	Testipohjainen tulosvastuu (No Child Left Behind) ja opettajien ajatukset	Ei verrokkiryhmää (puolistrukturoidut haastattelut)	Opettajien tavoitteet, keinot ja usko	• Standardoidut testit mittaavat kieltä, eivät sisältöosaamista ja johtavat epärealistisiin tavoitteisiin. • Vaikuttaa negatiivisesti opettajien tunteeseen vaikuttaa asioihin.
22	Cornelius ym. (2022)	Australia	Yksi syrjäinen tutkimuskoulun (n. 270 oppilasta) rehtori ja 6 johtajaa	Testipohjainen arviointi ja tulosvastuu	Ei verrokkiryhmää (autoetnografia)	Opettajien ja johtajien kokemukset, oppilaiden hyvinvointi ja osallistuminen	• Testit sivuuttavat sosiokulttuuriset tekijät. • Testiin opettaminen ja opetuksen kaventuminen. • Vähemmän autonomiaa opettajalle.
23	Cuff ym. (2018)	Englanti	1141 18-vuotiaiden yksikköä ja 550 16-vuotiaiden yksikköä. Oppilaiden määriä eri raportoida.	Arviointiuudistus (2010–2011)	Ei verrokkiryhmää (kuvaileva, aikasarja tuloksista 2010–2015)	Oppilaiden suoriutuminen kokeissa ja suoritustason muutokset ajan yli	• Näyttöä Sawtooth-efektistä, suoriutuminen paranee asteittain reformin jälkeen. • Parannukset eivät välttämättä tarkoita todellista oppimista vaan testiin tottumista ja strategioiden oppimista.
24	Daliri-Ngametua ym. (2024)	Australia	2 koulua (alakoulu ja yläkoulu), 27 opettajaa ja 7 koulun johtajaa	Kansallinen standardoitu testi ja testipohjainen tulosvastuu	Ei verrokkiryhmää (tapaustudkimus)	Opetuskäytännöt ja päätöksenteko kouluissa, opettajien työ ja kokemukset sekä koulun rakenteet ja käytännöt	• Testit sulautuneet arjen käytäntöihin ohjaten opetussuunnitelmia ja vaikuttaen henkilöstö- ja resurssipäätöksiin • Opetus kaventuu ja testikeskeisyys lisääntyy
25	Dawadi (2019)	Nepal	247 oppilasta (14-16-v.) ja 6 oppilasta ja heidän vanhempansa	Kansallinen korkean panoksen koe ja vanhempien osallistuminen	Ei verrokkiryhmää (kyselyt ja haastattelut)	Vanhempien osallistumisen muodot ja intensiteetti, testiin valmistautuminen ja vanhempien vaikutus oppimiseen ja hyvinvointiin	• Vanhempien osallistuminen on erittäin vahvaa • Kova paine ja kontrolli • Eriarvoisuus vanhempien koulutuksen mukaan • Oppiminen kapeutuu testisuoritukseen

26	Dawadi (2021)	Nepal	6 oppilasta ja heidän vanhempansa	Kansallinen korkean panoksen koe ja washback-ilmiö	Ei verrokkiryhmää (oppilaiden suulliset päiväkirjat ja puolistrukturoidut haastattelu)	Oppilaiden oppimisstrategiat ja testiin valmistautuminen, washbackin luonne, motivaatio	• Oppiminen painottuu ulkoa opetteluun • Motivaatio on kaksijakoinen, voi lisätä työskentelyä, mutta myös kaventaa oppimista ja lisätä stressiä
27	Deming (2016)	USA, Texas	~887 000 oppilasta (useita kohortteja), lukioikäiset (9.–12. luokka) joiden seuranta n. 25-v. asti	Koulut jaettiin osavaltiossa neljään julkiseen ryhmään oppilaiden standardoitujen kokeiden tulosten perusteella: 1) Low performing; 2) Acceptable; 3) Recognized; 4) Exemplary;	Kvasikokeellinen asetelma. Vertaillaan saman koulun eri kohortteja, joilla eri tasoinen paine syntyen koulujen luokittelusta. Tämä on mahdollista, koska pass/failed-osuudet, jotka määrittivät koulujen ”ryhmät”, muuttuivat ajassa kolmessa alimmassa ryhmässä.	Lyhyen aikavälin: testitulokset matematiikassa ja lukemisessa, valmistuminen, opintopisteet. Pitkän aikavälin: korkeakouluun kirjautuminen, tutkinnon suorittaminen, ansiot.	• Keskimäärin 0.5 SD vaikutus matematiikan tuloksiin, ei vaikutusta lukemiseen. • Vaikutukset suurimpia heikoimmissa (”Low performing”) kouluissa, missä iso 0.2 SD parannus matematiikassa, 1.9 p.p. lisäys yliopistoon menossa (+1.27 p.p. BA-tutkinto) ja 298 dollarin lisäys ansioissa 25-vuotiaana. • Havaitaan myös koulujen tekemää manipulointia vauraammissa ja lähtötasoltaan korkeammissa kouluissa (”Recognized”), missä tehdään heikosti suoriutuvien oppilaiden strategista siirtämistä erityisopetukseen, mistä seurautuu iso negatiivinen vaikutus oppimistuloksiin, tutkintoon ja ansioihin.
28	Doecke ym. (2010)	Australia	Opettajien ryhmäkeskustelut yhdessä koulussa (otoskokoa ei raportoitu)	Kansallinen standardoitu testi (NAPLAN) ja testipohjainen tulosvastuu	Ei verrokkiryhmää (opettajien haastattelut ja fokusryhmät, teoreettista pohdintaa)	Opettajien kokemukset ja käsitykset, opettajaidentiteetti ja pedagogiikka, opetuksen sisältö ja arviointikäytännöt, sekä oppilaiden oppimisen ja hyvinvoinnin ymmärtäminen	• Testit muuttavat opetusta merkittävästi: teaching to the test, opetuksen kaventuminen • Tulosvastuu kaventaa opetusta ja redusoivat oppimisen mittareiksi, samalla kun opettajat pyrkivät säilyttämään inhimillisen ja kokonaisvaltaisen näkemyksen oppimisesta

29	Dowley ym. (2022)	Australia	206 oppilasta, vuosiluokat 7 ja 9, neljästä koulusta	Kansallinen standardoitu testi (NAPLAN), Jota verrataan oppilaan kokemuksi koulun sisäisiin testeihin.	Ei verrokkiryhmää (36-kohdan verkkokysely SurveyMonkey)	Arvostus, odotukset, motivaatio ja tunteet	• Oppilaat arvostavat ja odottavat enemmän sisäisiltä kokeilta kuin NAPLANilta • Ulkoiset testit ovat tylsiä, ja aiheuttavat stressiä, mutta koskevat vain pientä ryhmää.
30	Dymoke (2012)	Uusi-Seelanti ja Englanti	6 koulua; 24 opettajaa, 13 luokkahuonehavaintoa (14-18-v.). 18 NZ-opettajaa.	Tarkastellaan, miten korkeanpanoksen testijärjestelmät vaikuttavat runouden opetukseen ja pedagogiikkaan.	Ei verrokkiryhmää (haastattelut, luokkahuonehavainnot ja dokumenttianalyysi)	Runouden asema opetuksessa ja arvioinnissa, opettajien pedagogiset valinnat, opettajien kokemat mahdollisuudet ja rajoitteet	• Runouden asema heikkenee, vähemmän aikaa ja tilaa opetuksessa • Opetuksen kaventuminen, koska vähemmän luovaa kirjoittamista • Opettajat kokevat menettävänsä autonomiaa
31	Ehren ym. (2013)	USA, NY	9 alakoulua (K–5), haastateltuina rehtorit ja kaksi opettajaa/koulu	Monimittarinen tuloksellisuus järjestelmä (valtion ja osavaltion)	Ei verrokkiryhmää (dokumenttianalyysi, haastattelut ja havainnointi)	Koulujen reaktiot Tavoitteisiin, Opetussuunnitelmaan ja opetukseen	• Standardoidut testit ohjaavat toimintaa eniten • Selvä opetuksen kaventuminen • Keskittyminen helppoihin mittareihin
32	Eklöf ym. (2013)	Ruotsi	1 189 oppilasta, 9. luokalta, 27 koulun otos (16-v.)	Arvioidaan oppilaiden kokemuksia kansallinen luonnontieteen testin tärkeydestä, panostuksesta ja koestressistä.	Ei verrokkiryhmää (kyselytutkimus)	Koettu tärkeys, panostus, testijännitys	• Testi koettiin kohtalaisen tärkeäksi, mutta ei yhtä tärkeäksi kuin arvosanat • Oppilaat pitivät kansallista koetta vähemmän tärkeänä kuin koulun omia kokeita tai kurssiarvosanoja • Testijännityksen tärkein ennustaja oli oppilaan usko omaan osaamiseensa
33	Elo ym. (2021)	Suomi	1 lukio, 9 opettajaa	Ylioppilaskoe ja oppiainelähtöinen kulttuuri	Ei verrokkiryhmää (9 opettajan puolistrukturoidut haastattelut ja taustaja dokumenttiaineisto)	Opettajien kokema autonomia, lukion profiilin vaikutus autonomiaan, ylioppilaskirjoitusten vaikutus autonomiaan	• Opettajat kokevat laajaa autonomiaa • Koulun profiili (poikkitieteisyys) voi lisätä opettajien vapautta • Ylioppilaskoe ohjaa opetusta
34	Federičová ym. (2017)	Slovakia	TIMSS-aineisto, missä Slovakiasta ~5 000–5 600 ja Tšekistä ~4 200–4	Vuoden 2009 reformi siirsi pääsykokeet 4. luokan lopusta 5. luokan loppuun. Tämän	Kvasikokeellinen asetelma (DID). Kontrolliryhmänä pääsykokeet muuten	TIMSS-matematiikkatulokset	• Pääsykokeiden läheisyys nosti oppilaiden matematiikkatuloksia 0.2 SD. Suurin hyöty

			600 oppilasta/ vuosi. (4. luokka eli 10-vuotiaat)	perusteella valikoidutaan 8-vuotisiin akateemisiin kouluihin.	yhtä varhaisessa vaiheessa tekevät tšekkiläiset oppilaat, 5. lk.		huippuoppilaille ja niille, jotka lähellä pääsyrajaa. • Viitteitä siitä, että vaikutus syntyy oppilaiden ja vanhempien lisääntyneestä panostuksesta.
35	Fisher ym. (2011)	USA	Neljä lukiota, 4 haastattelua, 17 tuntihavaintoa, 9 opettajaa	Kun iso koulu jaetaan pienempiin "oppimisyhteisöihin", niin voiko se parantaa opetusta tilanteessa missä korkeanpanoksen testausta.	Ei verrokkiryhmää (luokkahuonehavaintoj a, syvähaastatteluja)	Opetuskäytännöt ja suunnittelu, Opettajien käsitykset	• Opetus keskittyy pinnalliseen sisältöön • Korkeanpanoksen testaus ohjaa opetusta (teaching to test) • Pienemmät oppimisyhteisöt eivät ratkaise ongelmaa.
36	Friedric h ym. (2022)	USA	21 opettajaa yhteensä: 3 alakoulusta, 10 yläkoulusta, 8 lukiosta; opettajat 6 eri koulupiiristä (K–12)	Hätäetäopetukseen siirtyminen COVID-19- pandemian vuoksi sekä siihen liittyvät muutokset, kuten testivaatimusten väliaikainen keskeyttäminen.	Ei verrokkiryhmää (verkkokysely, jossa avoimia kysymyksiä, pragmaattinen konstruktivistinen sisällönanalyysi)	Oppilaiden osallistuminen ja tavoitettavuus, teknologinen saavutettavuus ja tasa- arvo, koulun ja koulupiirin tuki opettajille, testauksen vaikutus opetukseen	• Standardoitujen testien poistaminen vähensi painetta ja antoi enemmän vapautta opetuksen suunnitteluun. • Haasteena oppilaiden sitoutumisen lasku ja vaikeus pitää yhteyttä oppilaisiin. • Epäyhtenäinen tuki koulupiireiltä.
37	Fuchsm an ym. (2020 /2023)	USA, Georgia	Opettaja- aineistoa luokissa 1–8 vuosilta 2007– 2018. Yhteensä 453 955, joista 91 529 uran alkuvaiheessa.	Valtakunnallisen testauksen osittainen poistaminen, 2011 testauksen poistaminen 1.–2. luokilta ja 2017 testauksen poistaminen 6.–7. luokkien luonnontieteistä ja yhteiskuntaopista.	Kvasikokeellinen asetelma (DID). Vuosien 2011 ja 2017 muutoksen kohdanneita opettajia verrataan vastaaviin opettajiin, joita muutos ei koskettanut.	Opettajan päätös vaihtaa ammattia, koulun vaihto, piirin vaihto, luokka- asteen vaihto, koulun laadun jakauman (value- added) muutokset	• Keskimäärin testauksen poisto ei opettajapoistumaan keskimäärin, mutta testauksen poisto laski aloittelevien opettajien poistumaa 7 % alaluokilla ja 21 % 6.–7. luokille. • Pieni vaikutus luokka-asteen vaihtamiseen (-0.5 %-yks.) • Ei merkittäviä muutoksia koulun laadun jakaumassa.
38	Furuta (2022)	132 maata	132 maan kansalliset koulujärjestelmät	Kansallisten oppimistulosten arviointien käyttöönotto sekä maailmanlaajuinen yleistyminen	Ei verrokkiryhmää (tapahtumahistoria- analyysi 1960–2011)	Milloin maat ottavat käyttöön ensimmäisen kansallisen arvioinnin, mikä on kansallisen arvioinnin taso, miten kansallisiin arviointeihin	• Kansalliset arvioinnit yleistyvät nopeasti 1990 jälkeen erityisesti perusopetuksessa. • Korkean panoksen kokeita käyttävät ja varhain oppilaita eriyttävät maat ovat selvästi

						vaikuttavat korkean panoksen kokeet ja tulosvastuu	vähemmän todennäköisiä ottamaan käyttöön kansallisia arviointeja. Muodostaa globaalin trendin, jossa arvioinnit nähdään osana "Education for All" -tavoitteiden seurantaa ja tasa-arvoideologiaa.
39	Gannon (2019)	Australia	181 opettajaa	NAPLAN testijärjestelmä ja tulosvastuu	Ei verrokkiryhmää (kysely)	Opettajien käsitykset NAPLANista, opetuskäytännöt, vaikutus (washback), ajankäyttö, koetut seuraukset, arviointitiedon hyödyllisyys	- Opetus muokkautuu testin mukaan, pedagogiikka kaventuu - Ajankäyttöön vaikutus epäselvä - Järjestelmä ja vertailutieto lisäävät opettajien painetta - Positiivista että enemmän fokusta kirjoittamiseen, mutta tämä opetuksen kaventumisen kustannuksella.
40	Giambo (2017)	USA, Florida	~100 oppilasta Alakoulusta (4.–6. lk.) ja yläkoulusta	Korkeanpanoksen testauksen käyttö valmistumiseen ja koulujen arviointiin	Ei verrokkiryhmää (kysely)	Oppilaiden käsitykset testistä, testivalmistelun strategiat, motivaatio ja käyttäytyminen, koettu testin relevanssi ja oikeudenmukaisuus	- Negatiiviset ja neutraalit asenteet ja testivalmistelu on pinnallista. - Testi koetaan epäoikeudenmukaiseksi. - Motivaatioon ristiriitainen vaikutus.
41	Giourou kakis ym. (2010)	USA, NY	3 lukiota (n. 15–18-vuotiaat) ja 4 opettajaa	Korkeanpanoksen testaus ja kokeeseen valmistautuminen	Ei verrokkiryhmää (kysely, havainnot, haastattelut)	Opetuskäytännöt, testivalmistelun muodot, opettajien käsitykset testauksesta	- Opettajat opettivat "testilähtöisesti" (teaching to the test) - Opettajat kamppailivat auttaa oppilaita läpäisemään koe ja samalla tukea pitkäjänteistä oppimista ja lukijaksi kasvamista
42	Gonzalez ym. (2016)	USA, Texas	145 opettajaa (peruskoulu, yläkoulu, lukio)	Korkeanpanoksen testaus ja sen vaikutus stressiin ja ammatilliseen pystyvyyteen	Ei verrokkiryhmää (kysely ja ryhmähaastattelut)	Ammatillinen pystyvyys ja työperäinen stressi	- Vaikutus stressiin, mutta ei vaikutusta ammatilliseen pystyvyyteen. - Vaikutus opetuksen kaventumiseen.

43	Goodacre (2023)	Englanti	3 tapaustutkimusoppilasta	GCSE-arviointi, kansallinen opetussuunnitelma, korkean panoksen testaus	Ei verrokkiryhmää (tapaustutkimus ja kriittinen politiikka-analyysi)	Oppilaiden edistyminen, kouluosoiutumisen, arviointikäytännöt, eriarvoisuusvaikutukset	- Mittaava järjestelmä peittää alleen rakenteellisen epätasaa- - Korkea panostus ja numeroihin perustuva arviointi tuottavat epäinhimillisiä, hegemonisia käytäntöjä, jotka uusintavat luokka-, rotu- ja sukupuolieroja
44*	Guill ym. (2019)	Saksa	n ≈ 5 669 (PIRLS), 4 685 (TIMSS), 17 759 (NAS); alakoulun oppilaat (n. 9–10 v.)	Korkean panoksen koulusiirtymä (alakoulusta yläkouluun), erityisesti sitovat vs. ei-sitovat suositukset (tracking-järjestelmä)	Kontrollitutkimus (logistinen regressio, eri osavaltioiden vaihtelu, mutta ei vaihtelua ajassa)	Todennäköisyys osallistua yksityisopetukseen, ja Korkean panoksen suositusten esiintyminen koulussa	- NAS 2011 -aineiston mukaan oppilaat korkean panoksen testien osavaltioissa käyttävät yksityisopetusta. PIRLS- ja TIMSS-aineistoissa ei löydy vastaavaa. - Yksityisopetusta käyttävät heikommin suoriutuvat, matalamman SES:n ja useammin maahanmuuttajataustaiset oppilaat.
45	Gunnulf sen ym. (2018)	Norja	176 opettajaa, alempi toinen aste.	Kansalliset testit matalan panoksen kontekstissa	Ei verrokkiryhmää (kysely)	Opettajien käytännöt (testitulosten käyttö), asenteet testejä kohtaan, rehtorin fasilitointi.	- Rehtorin aktiivinen fasilitointi on vahvasti yhteydessä opettajien testitulosten käyttöön ja positiivisempiin asenteisiin. - Politiikan tavoitteiden ja käytännön välillä on kuilu.
46	Göloğlu ym. (2021)	Turkki	13 opettajaa yläkoulusta	Korkean panoksen testaus osana koulutuspolitiikkaa	Ei verrokkiryhmää (puolistrukturoidut haastattelut ja sisällönanalyysi)	Opetuskäytännöt, arviointikäytännöt, oppilaiden ja opettajien motivaatio, viestintä (opettaja–oppilas, opettaja–vanhemmat)	- Korkean panoksen testaus ohjaa opetusta kohti testipainotteisuutta ja perinteisiä menetelmiä. - Heikentää motivaatiota, lisää stressiä ja kaventaa opetusta.
47	Hardy ym. (2018)	Australia	Yksi koulu. Haastateltavina koulun rehtori ja Head of Curriculum (HoC).	Dataohjattu koulutuspolitiikka	Ei verrokkiryhmää (haastattelut)	Opetuskäytännöt, datan rooli päätöksenteossa, oppilaiden identiteetti motivaatio	- Data tekee ”keskitasoiset” oppilaat näkyviksi interventioiden kohteina, mutta samalla redusoiivat heidät datapisteiksi. - Opetusta ohjaa vahvasti testidata ja vertailu.

							Data voi tukea oppimista ja vuorovaikutusta, mutta myös kaventaa käsitystä oppimisesta ja tehdä oppilaista.
48	Heissel ym. (2021)	USA, New Orleans	93 oppilasta, luokat 3–8 (n. 8–15 v.), pääosin matalan SES-taustan oppilaita	Korkean panoksen standardoidut testit (Louisianan osavaltion vuosittainen koe)	Kontrollitutkimus (kenttätutkimus missä tehdään fysiologinen mittausta ja seurataan kohdehenkilöitä baseline arvoon verrattuna)	Kortisolitasot (stressi), kortisolireaktiivisuus, testisuoriutuminen. Kolmen viikon mittainen otos: baseline-viikko, low-stakes-testiviikko, high-stakes-testiviikko.	- Testiviikko nostaa stressitasoja (n. +18 % ennen testiä). - Korkean stressin ympäristöissä elävät oppilaat näyttivät voimakkaampia stressireaktioita. - Sekä liian suuri että liian pieni stressireaktio liittyy heikompaan suoriutumiseen. - Testitulokset heijastavat osin stressiä ja stressinsietokykyä eikä pelkkää osaamista.
49	Hoadley Ym. (2016)	Etelä-Afrikka	14 koulua (matala SES), 69 opettajan haastattelu	Laajamittainen standardoitu testaus	Ei verrokkiryhmää (haastattelut)	Opettajien tulkinnat testituloksista, pedagogiset käytännöt, resurssien kohdentaminen	- Testit tekevät opetuksen, opettajat ja oppilaat “näkyviksi” ja lisäävät eriyttämistä ja resurssien kohdentamista. - Testit aiheuttavat painetta ja ambivalenssia
50	Hofflinger ym. (2020)	Chile	~1.96 milj. 4. luokan oppilasta vuosilta 2005–2013 (9–10-v.)	Chilen kansallinen korkeiden panosten testi (4., 8., 10. lk.) ja SEP-reformi asteittainen käyttöönotto.	Kvasikokeellinen asetelma (DID) paneeliaineisto ja (Regressio missä otetaan huomioon kiinteät koulu ja vuosi vaikutukset reformin yli).	Testien väliin jättämisen todennäköisyys, vaikutukset koulun luokitteluun, vaikutus kansallisiin 4.luokan matematiikan ja lukemisen tuloksiin.	- Kansalliset testitulokset nousivat (0.2–0.3 SD), mutta osa noususta johtui strategisesta manipuloinnista: heikot oppilaat jätettiin testien ulkopuolelle. - Suurempi vaikutus köyhissä kouluissa, mutta niissä myös suurempi tulosten manipulointi.
51	Holm (2017)	Tanska	3 koulua, 3 luokkaa; seuranta 6 vuoden ajan (luokat 1–6, n. 6–12 v.)	Kansallinen standardoitu lukutesti osana tulosvastuupolitiikkaa	Ei verrokkiryhmää (haastattelut, havainnointi ja dokumenttiaineisto)	Opettajien tulkinnat oppilaiden lukutaidosta; testitulosten käyttö; käsitykset testien validiteetista.	- Opettajat rakentavat kahta rinnakkaista osaamiskäsitystä (testipohjainen vs. arjen havaintoihin perustuva), joista jälkimmäistä pidetään luotettavampana - Testien validiteettia kyseenalaistetaan erityisesti

							monikielisten oppilaiden kohdalla ja testituloksia käytetään varoen ja kriittisesti.
52*	Holmes ym. (2018)	Englanti	2933 oppilasta (n. 16 v., GCSE); lisäksi 43 asiantuntijaa (PhD mat.) arvioimassa tehtäviä.	GCSE-kokeiden vaikeustason vertailtavuus eri koejärjestäjien (exam boards A, B, C) välillä	Kontrollitutkimus (ennustettu vaikeus vs. todellinen vaikeus tai ns. psykometrinen tapaustutkimus)	Tehtävien ja kokeiden vaikeus, testisuoriutuminen, korrelaatio arvioidun ja todellisen vaikeuden välillä	- Eri asiantuntija-arviot korreloivat melko vahvasti todellisen vaikeuden kanssa ($r \approx 0.67-0.76$) - Koejärjestäjien välillä oli merkittäviä vaikeuseroja.
53	Hong ym. (2010)	Etelä-Korea & USA	44 korealaista ja 21 yhdysvaltalaisista yläkoulun opettajaa	Arvioi luovuuden edistäminen opetuksessa ja korkean panoksen testauksen konteksti rajoitteena	Ei verrokkiryhmää (kysely ja maiden vertailu)	Luovuuden määritelmät ja käsitteelliset piirteet (, opettajien käsitykset luovuuden opettamisesta ja koetut esteet	- Opettajien käsitykset luovuudesta olivat yksilötasolla kapeita. - Korkean panoksen testaus rajoittaa luovuuden edistämistä
54	Huddleston (2014)	USA	Kirjallisuuskatsaus, Alakoulun ja yläkoulun nivelvuodet, joihin kohdistetaan testiperusteinen luokallejäänti	Testiperusteinen luokallejäänti (test-based grade retention) eri osavaltioissa ja kaupungeissa	Ei verrokkiryhmää (Maxwellin (2006) kirjallisuuskatsausmenetelmä)	Oppimistulokset (lyhyt vs. pitkä aikaväli), koulupudokkuus, sosioemotionaaliset vaikutukset, eriarvoisuusvaikutukset	- Lyhyen aikavälin oppimistulokset voivat parantua, mutta vaikutukset heikkenevät ajan myötä - Luokalle jättäminen lisää merkittävästi koulupudokkuuden riskiä - Toimenpiteet kohdistuvat suhteettomasti heikoimmassa asemassa oleviin (matala SES, vähemmistöt)
55	Huss ym. (2011)	USA	104 opettajaa (52 % vastausaste); middle school (n. 10–15 v. oppilaat)	Korkean panoksen testaus ja accountability-järjestelmät suhteessa middle school -mallin toteutumiseen	Ei verrokkiryhmää (kysely)	Opettajien arviot middle school -mallin komponenttien toteutumisesta, vaikutukset opetukseen, ajankäyttöön, autonomiaan, testauksen koetut vaikutukset	- korkean panoksen testaus kaventaa opetusta (vähemmän valinnaisia ja rikastavia sisältöjä) ja vähentää opettajien autonomiaa - Korkean panoksen testauksella on myös positiivisia vaikutuksia (esim. parempi fokus ydinaineisiin)
56	Högberg ym. (2023)	Ruotsi	~18 500 oppilasta Ruotsista. (PISA-mittausten	Tutkimus tarkastelee Ruotsin 2010–2014 koulutusreformia,	Kvasikokeellinen asetelma (DID)	Oppilaiden kuulumisen tunne (itsearvio) ja	Koulukuulumisen tunne romahti 2012–2015, poikkeuksellisen voimakkaasti ja suurimmat

			poikkileikkuasaineistot, 2000–2018)	joissa korostuu tulosvastuullisuus ja inklusiopolitiikka	Synteettinen kontrolli rakennettiin Norjasta ja Itävalasta (niiden painotettu keskiarvo)	eriarvoisuus ryhmien välillä	negatiiviset vaikutukset kohdistuivat heikommassa asemassa oleviin.
57	Högberg ym. (2021)	Ruotsi	~13 000 oppilasta HBSC-kyselystä 2010–2014, 11–15 v., luokat 5, 7 ja 9.	Arviointi- ja arvosanareformi 2012: Arvosanojen antamista aikaistettiin luokalta 8 (~14 v.) luokille 6 ja 7. Samalla kansalliset kokeet siirrettiin 5.luokalta 6:lle luokalle	Kvasikokeellinen asetelma (DID). Käsittelyryhmä: 7. luokan oppilaat (ensimmäiset, joita reformi koski). Kontrollit: 5. ja 9. luokka	Itse raportoitu koulustressi, akateeminen itsetunto, psykosomaattiset oireet, elämäntyytyväisyys	Reformin jälkeen itse raportoitu koulustressi kasvoi 7.luokalla 0.31 SD., akateeminen itsearvio laski 0.22 SD, psykosomaattiset oireet lisääntyivät 0.15 SD ja elämäntyytyväisyys väheni 0.25 SD.
58	Ioannidou ym. (2022)	Englanti & Wales	51 opettajaa ja 3 haastattelua	Käytännön luonnontieteen opetus ja summatiivinen arviointi	Ei verrokkiryhmää (kysely ja tapaustutkimus)	Opettajien käsitykset käytännön opetuksesta, arviointikäytännöt	Korkean panoksen testaus ohjaa opetusta
59	Journell (2010)	USA	6 hallinnon opettajaa kolmesta lukiosta	Testauksen vaikutus opetukseen ja poliittiset tapahtumat	Ei verrokkiryhmää (etnografinen tutkimus)	Opettajien halukkuus ja kyky sisällyttää ajankohtaisia poliittisia tapahtumia	Korkean panoksen testaus kaventaa opetusta
60	Kiomrs ym. (2011)	Iran	53 toisen asteen opettajaa	Testien washback-vaikutus (eli kokeiden vaikutus opetukseen)	Ei verrokkiryhmää (kyselylomake opetuskäytännöistä ja opettajien laatimien kokeiden analyysi)	Opettajien arviointiosaamisen taso ja opetuskäytännöt	- Opettajien arviointiosaaminen on heikkoa - Kaikki opettajat mukauttavat opetustaan kokeisiin
61	Klein (2016)	Israel	60 koulua (30 ala- ja 30 yläkoulua) ja 300 opettajaa (5 per koulu)	Korkean panoksen testaus ja tilivelvollisuus	Ei verrokkiryhmää (kysely)	Vanhempien, valvojan, rehtorin ja ainevastaavan osallistuminen, opettajan työpanos, opetustuntien määrä, epäeettinen toiminta	Korkean panoksen testaus ohjaa voimakkaasti koulun toimintaa ja resurssien kohdentamista, usein opetuksen laajuuden kustannuksella.
62	Krause ym. (2022)	Alankomaat	23 videonauhottettua maantiedon oppituntia	Korkean panoksen kokeiden vaikutus opetukseen	Ei verrokkiryhmää (videohavainnointia 23 oppituntia, tehtävien ja kysymysten analyysi)	Tehtävien kognitiivinen taso, tehtävien alkuperä, tehtävien purku, opettajan kysymysten taso	- Testaus rajaa ajattelun tasoa opetuksessa ja ohjaa opetusta voimakkaasti - Oppikirjat dominoivat opetusta - Korkeamman ajattelun tavoitteet eivät toteudu opetuksessa

63	Lambert ym. (2015)	Australia	15 draamaopettajaa, 13 entistä opiskelijaa	Neoliberalismi koulutuksessa, tilivelvollisuus	Ei verrokkiryhmää (puolistrukturoidut haastattelut)	Opettajien kokemukset, opetuksen muutokset, oppiaineiden asema	- Tulosperusteinen kulttuuri hallitsee, opetus kaventuu ja eriarvoisuus kasvaa - Data ohjaa opetusta ("dataphilia")
64	Larsson ym. (2020)	Ruotsi	5 ruotsin kielen opettajaa (lukio)	Kansallinen standardoitu koe politiikkana	Ei verrokkiryhmää (puolistrukturoidut haastattelut)	Opettajien opetuskäytännöt suhtautuminen kokeeseen sopeutuminen vs. vastustus	- Testit muokkaavat opetusta, mutta vaihtelee osa-alueittain "Opetetaan testiin" lisääntyä - Kansallinen koe ei vaikuta opetukseen tasaisesti.
65*	Leckie ym. (2019)	Englanti	~503 000 oppilasta, 3 098 toisen asteen koulua	Korkean panoksen koulutilivelvollisuusjärjestelmä ja oppilastaustan huomiointi jättäminen	Kontrollitutkimus (lineaarinen regressiomalli, johon lisätään kontrolleja)	Koulun suoriutumispisteet, sijoitukset (ranking), luokittelut, oppilaiden edistyminen eri ryhmissä	- Oppilastaustan huomioiminen muuttaa merkittävästi tulkintoja koulun laadusta (32 % kouluista vaihtaa luokkaa ja noin 20 % kouluista siirtyy rankingissa yli 500 sijaa). - Ilman taustamuuttujia mittari suosii hyväosaisia oppilasryhmiä opettavia kouluja ja rankaisee heikommassa asemassa olevia.
66	Leite ym. (2018)	Portugali	45 opettajaa kolmesta toisen asteen koulusta	Kansallinen opetussuunnitelma ja sen kontekstualisointi suhteessa kansallisiin kokeisiin	Ei verrokkiryhmää (fokusryhmähaastattelut, aineistolähtöinen sisällönanalyysi)	Opettajien käsitykset, motivaatio, käytännöt, hyödyt, rajoitteet	Opettajat ovat motivoituneita kontekstualisoimaan opetusta, koska se lisää oppilaiden motivaatiota ja ymmärrystä, mutta kansalliset kokeet ja laaja opetussuunnitelma rajoittavat tätä.
67	Leite ym. (2019)	Portugali	236 perus- ja toisen asteen opettajaa	Kansallinen opetussuunnitelma vs. sen kontekstualisointi eri kouluasteilla	Ei verrokkiryhmää (kysely)	Opettajien käyttämät opetuskäytännöt ja erot koulutusasteiden välillä	- Opetus on perinteisiä, että kontekstualisoivia käytäntöjä. - Ylemmillä asteilla korostuvat perinteiset, sisältökeskeiset menetelmät, jota selitetään kansallisten kokeiden paineella
68	Leo (2023)	USA	Yksi lukio 18 nykyistä ja 15 entistä oppilasta, 6 opettajaa ja 2	Korkean panoksen kokeet ja niiden vaikutus	Ei verrokkiryhmää (haastattelut)	Opetuskäytännöt, oppilaiden kokemukset, opettajien toiminta, "agency" (toimijuus)	Kokeet dominoivat opetusta ja heikentävät oppilaiden tavoitteita, mutta opettajat ja oppilaat löytävät rajattuja tapoja

			hallinnon edustajaa	englanninoppijoihin sekä opettajiin			vastustaa ja neuvotella järjestelmää.
69	Lewis ym. (2014)	Australia	Yksi peruskoulu ja 32 opettajaa ja johtajaa.	NAPLAN korkean panoksen kokeet siihen liittyvä rahoitus, maine ja tavoitteet	Ei verrokkiryhmää (haastattelut)	Opettajien käytännöt, identiteetti, päätöksenteko, koulun toiminta	• Testitulokset ohjaavat voimakkaasti toimintaa • Rahoitus, maine ja tavoitteet rakentuvat testien ympärille, josta syntyy "performatiivinen" opettajuus ja koulukulttuuri
70	Li ym. (2020)	Kiina	25 toisen asteen opettajaa (12 koulua)	Integrated writing - opetus ja opettajien käsitykset korkean panoksen koeympäristössä	Ei verrokkiryhmää (haastattelut)	Opettajien käsitykset, pedagogiikka, suhde opetussuunnitelmaan	• Erotetaan kolme käsitystyyppiä: testiorientoitunut, prosessorientoitunut, kehityksellinen ja vain viimeinen linjassa opetussuunnitelman kanssa.
71*	Li ym. (2018)	USA	~24 000 oppilasta, 1 353 opettajaa, 254 koulua	Testiin valmistautuminen ja valtion kokeiden tulokset	Kontrollitutkimus	Testiin valmistautuminen ja testitulokset	• Heikosti suoriutuneet oppilaat saivat selvästi enemmän testivalmistelua. • Hyödyttää vähemmistöoppilaita suhteellisesti enemmän, mutta vaikutus ei kavenna oppimiseroja käytännössä.
72	Lidar ym. (2017)	Ruotsi	407 yläkoulun opettajaa	Kansallisten kokeiden vaikutus opetukseen	Ei verrokkiryhmää (kysely ja temaattinen analyysi)	Opetuksen sisällöt, opetustraditiot, muutokset opetuksessa	• Opetus jakautuu kahteen traditioon (Academic & Moral) ja kokeet ohjaavat opetusta kohti soveltavaa ja yhteiskunnallista sisältöä.
73	Lilja ym. (2020)	Ruotsi	46 opettajaa, 6. ja 9. luokalta	Eettiset kompetenssit, joita opettajat pitävät tärkeinä (osana kansallisia uskonnonopin kokeita)	Ei verrokkiryhmää (14 ryhmähaastattelua)	Eettiset taidot ja niiden kohteet (itse, toinen, yhteiskunta)	• Opettajille neljä ydinkompetenssia: ymmärtäminen, toimiminen, verbalisoiminen, sitkeys. • Opettajilla laajempi käsitys kuin opetussuunnitelmassa.
74	Lu ym. (2012)	Kiina	4 provinssin opetussuunnitelmat ja lukion biologian päättökokeet	Opetussuunnitelman ja standardoitujen kokeiden välinen yhdenmukaisuus	Ei verrokkiryhmää (sisällönanalyysi ja Porterin alignment-malli)	Alignment-indeksi, sisältö ja kognitiivinen taso	• Kaikki kokeet heikosti linjassa opetussuunnitelman kanssa • Kokeet painottavat matalampia kognitiivisia tasoja

75	Lundqvist ym. (2021)	Ruotsi	10 opettajaa	Opetussuunnitelma, varhaisempi arviointi ja kansalliset kokeet	Ei verrokkiryhmää (haastattelut)	Opettajien opetuskäytännöt, arviointikäytännöt, käsitykset reformeista, ja sopeutuminen muutokseen	- Opettajat mukauttavat opetusta lisäämällä läpinäkyvyyttä ja arviointia, mutta kokeet myös lisäävät stressiä. - Reformit tukevat ammatillista oppimista ja yhteistyötä.
76	MacQuarrie ym. (2021)	Skotlanti	31 opettajaa, aineisto kerätty opettajakonferenssissa	Arviointireformin implementointi vähemmistökielisessä opetuksessa	Ei verrokkiryhmää (fokusryhmätutkimus)	Opettajien käsitykset reformista, arvioinnin tarkoitus, vaikutukset opetukseen, oppilaisiin ja vanhempiin	- Opettajat epävarmoja reformin tarkoituksesta - Pelko "teaching to the test" - Mahdollisia hyötyjä yhtenäisyys, mutta riskejä oppimiselle opettajan autonomialle
77	Malsbary (2016)	USA	Neljän opettajan haastattelu, n. 50 osallistujaa kokouksesta	Opettajien vastarinta standardoitua testausta vastaan ("test refusal") ja opettajat politiikan tekijöinä	Ei verrokkiryhmää (etnografinen tapaustutkimus)	Opettajien merkityksenanto testivastarinnalle, vaikutukset oppilaisiin (kuuluminen, luottamus), opettajien ammatillinen toimijuus	- Opettajien vastarinta toimii politiikan tekemisen muotona - Standardoitu testaus koettiin haitalliseksi erityisesti maahanmuuttaja- ja monikielisille oppilaille
78	Mammadov (2019)	Turkki	Kahdeksan lahjakasta yläkoulun oppilasta eri koulutyypeistä	Lahjakkaiden oppilaiden koulutus- ja sosiaaliset kokemukset sekä "gifted"-leiman vaikutukset	Ei verrokkiryhmää (fenomenologinen tutkimus)	Oppilaiden kokemukset koulusta (boredom, opetus), testistressi, sosiaaliset suhteet, stigmatisaatio, coping-strategiat	- Oppilaat kokevat tylsyyttä ja riittämätöntä haastetta koulussa. - Korkean panoksen testit aiheuttavat stressiä ja burnoutia "gifted"-leima tuottaa sosiaalista stigmaa.
79	Chak (2024)	Kiina	50 lukiolaista (17–19 v) ja 6 haastattelua	Korkean panoksen kokeen (washback) vaikutus englannin oppimisstrategioihin	Ei verrokkiryhmää (kysely ja puolistrukturoidut haastattelut)	Oppimisstrategioiden käyttö, strategiapreferenssit, oppijoiden käsitykset, testipaineen vaikutus	- Varjokoulutus (tutorial schools) ohjaa strategioita enemmän kuin koulut tai koe. - Opiskelijat kokevat ristiriidan aidon kielenoppimisen ja kokeeseen valmistautumisen välillä.
80	Marinho ym. (2017)	Portugali	Kaksi koulua, 6 matematiikan opettajaa ja kaksi koordinaattoria.	Matematiikan summatiivisen arvioinnin käytännöt ja opettajien uskomukset	Ei verrokkiryhmää (tapaustutkimus)	Arviointikäytännöt (testit), opettajien uskomukset, arvioinnin tarkoitus, opetuskäytännöt	- Summatiiviset kokeet dominoivat. - Testit nähdään objektiivisina, tehokkaina ja motivoivina,

							mutta vaikutus kohdistuu lähinnä hyvin oppilaisiin. • Arviointi keskittyy luokitteluun, ei oppimisen tukemiseen.
81*	Marks (2015)	Australia	Longitudinal Study of Australian Children (~4000 oppilasta NAPLAN-testitulokset luokilta 3, 5 ja 7.)	Koulusektorien (julkinen, katolinen, yksityinen) vaikutus oppimistuloksiin.	Kontrollitutkimus (regressiot ja koulukiinteät vaikutukset, value added malli)	Lukeminen, matematiikka, kirjoittaminen, oikeinkirjoitus, kielioppi.	• Erot koulujen välillä katoavat lähes kokonaan, kun kontrolloidaan sosioekonominen tausta ja erityisesti aiempi kyky. • Pieniä "value-added" vaikutuksia yksityiskouluissa.
82*	Marks (2021)	Australia	NAPLAN-aineisto, ~ 1.5 miljoonaa oppilasta (vuodet 3→5 ja 7→9),	Pitäisikö value-added -malleihin lisätä oppilas- ja koulutason kontrollimuuttujia	Kontrollitutkimus (regressiot ja koulukiinteät vaikutukset, value added malli)	Lukeminen, matematiikka, kirjoittaminen, oikeinkirjoitus, kielioppi.	• Aiempaan osaamiseen perustuva malli selittää suurimman osan oppimistuloksista. • Lisämuuttujat (SES, demografia, koulutason muuttujat) eivät paranna selitysvoimaa, mutta heikentävät mallien vakautta ja johdonmukaisuutta.
83	Martin (2012)	Kanada	Politiikka- ja opetussuunnitelm adokumentit 1990–2000-luvulla	Neoliberaalien koulutusreformien ja yrityslogiikan ("corporate culture") vaikutus koulutukseen ja oppilaan rooliin	Ei verrokkiryhmää (dokumenttianalyysi ja teoreettinen tulkinta)	Analyysin kohteena diskurssit: koulutuksen tavoitteet, oppilaan rooli, työelämäorientaatio	• Koulutus politisoituu talouslogiikan mukaisesti: oppilaat rakennetaan "tuleviksi työntekijöiksi", koulutus sidotaan suoraan talouskasvuun.
84	May (2016)	USA	Yksi alakoulu, sen oppilaat (luokat 1–4), opettajat ja opettajaopiskelijat, sekä osittainen vertailukoulu	Taidekasvatuksen (musiikki/viuluopetus) rooli koulureformissa ja vaikutukset oppimiseen sekä hyvinvointiin	Ei verrokkiryhmää (tapaustutkimus ja pitkittäinen seuranta)	Kognitiivinen kehitys, lukeminen, keskittyminen, sosiaaliset taidot, itsetunto, opettajien ja vanhempien kokemukset	• Viuluopetus yhdistyy parempaan keskittymiseen, yhteistyötaitoihin ja itsetuntoon. • Taide voi toimia koulureformin välineenä erityisesti heikommassa asemassa oleville oppilaille.
85	McCarthy (2024)	Englanti	Neljä toisen asteen oppilasta (15–18 v) yhdestä koulusta, havainnointia ja haastatteluja	Korkean panoksen kokeiden rooli oppilaiden arjessa ja reaktiot kokeiden peruuntumiseen COVID aikana	Ei verrokkiryhmää (havainnointi ja puolistrukturoidut haastattelut)	Oppilaiden kokemukset, asenteet kokeisiin, motivaatio, identiteetti, "exam-oriented habitus", reaktiot.	• Kokeet muokkaavat voimakkaasti oppilaiden ajattelua ja toimintaa. • kokeiden peruminen aiheutti ristiriitaisia reaktioita (ilo, pettymys, merkityskriisi).

							• Oppilaat kyseenalaistivat kokeisiin perustuvan järjestelmän ja sen oikeudenmukaisuuden.
86*	McCoy ym. (2024)	Irlanti	Pitkittäisaneisto ("Growing Up in Ireland") (~ 5 375 oppilasta, 17–18v.) ja 10 koulun tapaustutkimus	Varjo koulutus (yksityisopetus) korkean panoksen koejärjestelmässä: eriarvoisuus ja hyvinvointi	Kontrollitutkimus ja täydentävä Ei verrokkiryhmää (logistinen regressio ja fokusryhmähaastattelut)	Varjo koulutukseen osallistuminen, sosioekonomiset tekijät, aiempi osaaminen, koulutusodotukset, hyvinvointi, opiskelijoiden kokemukset	• Laajalle levinnyt (~50–60%) ja sosiaalisesti eriytynyt (SES, vanhempien koulutus, tulot. • Erityisesti keskitason ja korkeiden odotusten oppilaat käyttävät sitä "investointina". • Oppilaat eivät suoraan yhdistä sitä stressiin, vaikka se lisää kuormitusta ja vähentää vapaa-aikaa.
87*	Meeter ym. (2023)	Alankomaat	~ 180 000 toisen asteen opiskelijaa	Korkean panoksen kansallisten kokeiden vs. koulun GPA:n kyky ennustaa korkeakouluun jäämistä	Kontrollitutkimus (pitkittäisanalyysi, logistinen regressio)	Ensimmäisen vuoden retention (jatkaako opiskelija), GPA, kansalliset koearvosanat	• GPA ennustaa retention selvästi paremmin kuin kansalliset kokeet • Kokeet eivät selitä lisävarianssia GPA:n lisäksi • Kokeissa epäonnistuneiden kohdalla GPA säilyy vahvana ennustajana.
88	Michaelides (2014)	Kypros	25 valmistunutta lukiolaista	Oppilaiden käsitykset korkean panoksen kansallisista kokeista ja niiden rooli.	Ei verrokkiryhmää (puolistrukturoidut haastattelut ja sisällönanalyysi)	Oppilaiden käsitykset: kokeiden oikeudenmukaisuus, ymmärrys pisteytyksestä, koettu vaikutus oppimiseen, stressi, yksityisopetus, koulun toiminta	• Oppilaat pitävät järjestelmää pääosin oikeudenmukaisena ja luotettavana. • Kokeet ohjaavat voimakkaasti opetusta ja oppimista. • Yksityisopetus nähdään välttämättömänä, joka lisää eriarvoisuusriski.
89	Misco ym. (2011)	USA, Ohio	204 opettajaa (23 % vastausaste) ja 9 syvähaastattelua	Korkean panoksen testauksen ja standardien vaikutus opetukseen	Ei verrokkiryhmää (kysely ja haastattelut)	Opettajien käsitykset: opetuksen sisältö, pedagogiikka, ajankäyttö, testikeskeisyys.	• Standardeja pidetään pääosin hyödyllisinä, mutta korkean panoksen testaus johtaa testikeskeisyyteen ja opetuksen kapeutumiseen.
90	Moore-Anderson (2022)	Englanti	Teoreettinen/pedagoginen artikkeli	Arvioinnin kehittäminen ja miten mitata	Ei verrokkiryhmää (teoreettinen/pedagoginen kehysartikkeli)	Oppilaiden ajattelun laatu	• Lyhyet vastauskysymykset mittaavat lähinnä pinnallista tietoa.

				ajattelua yli lyhyiden vastausten			Syvämpi ymmärrys edellyttää laajoja, avoimia tehtäviä.
91	Mora (2011)	USA	30 oppilasta (latinalaisamerikka lainen middle school, 11–14v.)	Korkean panoksen testauksen vaikutus oppilaiden kokemaan tylsyyteen ja koulukokemukseen	Ei verrokkiryhmää (etnografinen pitkittäinen tapaustutkimus 2.5 vuoden ajalta, havainnointi ja haastattelut)	Oppilaiden kokema tylsyys, sitoutuminen, käsitykset opetuksesta, suhtautuminen kouluun ja jatko-opintoihin	Korkean panoksen testaus johtaa opetuksen kaventumiseen ja testikeskeiseen, opettajajohtoiseen pedagogiikkaan, mikä lisää oppilaiden kokemaa tylsyyttä.
92	Morgan (2016)	USA	Ei empiiristä otosta; politiikka-artikkeli	Korkean panoksen standardoitujen testien käyttö koulujen ja opettajien arvioinnissa	Ei verrokkiryhmää (teoreettinen/politiikka-analyysi)	tarkastelussa vaikutukset: oppiminen, opetuksen laatu, opettajien yhteistyö, oikeudenmukaisuus, arvioinnin validiteetti	- Korkean panoksen testit tuottavat harhaanjohtavaa tietoa oppimisesta. - Kaventavat opetusta, lisäävät huijaamista ja epäeettisiä käytäntöjä.
93	Murchan (2018)	Irlanti	Laaja politiikka-aineisto (1990–2017): raportit, viralliset dokumentit, mediadata	School-Based Assessment (SBA) -reformin käyttöönotto junior tasolla ja siihen liittyvä politiikan implementaatio	Ei verrokkiryhmää (dokumenttianalyysi ja temaattinen analyysi)	Analyysin kohteena: sidosryhmien reaktiot, politiikan implementaatio, opetuksen ja arvioinnin muutokset ja roolit	- Reformi kohtasi voimakasta opettajavastustusta. - Paikalliset kontekstit ja opettajien rooli voivat estää kansainvälisesti inspiroitujen reformien toteutumisen.
94	Murphy ym (2012)	Englanti ja Wales	997 oppilasta ja kaksi arviointijärjestelmää: Englanti (SATS, korkean panoksen testaus) vs. Wales (opettaja-arviointi)	Oppilaiden käsitykset luonnontieteiden arvioinnista (KS2) ja sen vaikutus oppimiseen, motivaatioon ja hyvinvointiin.	Ei verrokkiryhmää (verkkokysely ja lasten osallistaminen kanssatutkijoina)	Asenteet arviointiin, oppimismotivaatio, testimieltymykset, hyvinvointi, koettu hyöty	- Oppilaat suhtautuvat positiivisesti arviointiin ja testaamiseen (erityisesti ei-SATS), mutta kokevat merkittävää stressiä. - Arviointijärjestelmästä riippumatta näkemykset samankaltaisia
95	Musolen ym. (2010)	USA	148–214 opettajaa ja koulun johtoa	Korkean panoksen testauksen (NCLB) vaikutus yläkoulun pedagogiikkaan, opetukseen ja koulun toimintafilosofiaan	Ei verrokkiryhmää (kysely ja avoimet vastaukset)	Opetuskäytännöt, ajankäyttö, opetussuunnitelma (laajuus vs. kapeus), testiin valmistautuminen	Testaus johtaa opetuksen kaventumiseen, testivalmennuksen lisääntymiseen ja oppilaslähtöisten käytäntöjen vähenemiseen.
96	Nehring ym. (2015)	Pohjois-Irlanti	4 toisen asteen koulua ja tarkempi analyysi yhdestä	Korkean panoksen testauksen ja 21. vuosisadan taitojen	Ei verrokkiryhmää (kysely, haastattelut)	Oppimistehtävien kognitiivinen vaativuus,	Ulkoiset kokeet heikentävät syvällistä oppimista ja ohjaavat matalan tason taitoihin.

			koulusta (≈670 oppilasta, 28 % matala SES). (oppilaat, opettajat, johto)	(deeper learning) välinen jännite koulutasolla		arviointikäytännöt, koulun käytännöt	• Projektit, portfolioit ja paikallinen arviointi lisäävät korkeatasoista oppimista. • Koulun sisäiset käytännöt voivat kompensoida testipaineita.
97	Neves ym. (2012)	Portugali	2,64 miljoonaa kansallista koetta (2002–2010), kaikki toisen asteen koulut (public vs. private)	Koulu-rankinglistat (league tables) ja niiden rooli eriarvoisuuden tuottamisessa ja esittämisessä	Kontrollitutkimus (regressio, missä selitetään kansallisten kokeiden ja opettajien tekemien arviointien välistä eroa yksityisissä ja julkisissa kouluissa)	Kansalliset koepisteet, koulun sisäiset arvosanat, erot näiden välillä (internal vs. exam scores), ranking-sijoitukset, public vs. private -erot.	• Ranking listat perustuvat käytännössä kansallinen kokeen ja opettajan arvion yhdistelmään. • Rankingit näyttävät yksityiset koulut parempina, vaikka ero ei perustu pelkästään osaamiseen: opettajien tekemät arvioinnit ovat enemmän ylöspäin harhaisia yksityisissä kouluissa kansallisiin mittauksiin verrattuna.
98	Njiru ym. (2021)	Kenia	9 matematiikan opettajaa ja 26 oppilasta (toinen aste)	Oppilaiden ranking vaikutus oppilaiden ääneen ja toimijuuteen matematiikan opetuksessa	Ei verrokkiryhmää (haastattelut)	Oppilaiden ääni ja toimijuus, resurssien jakautuminen, ryhmäroolit, opettajien päätökset, motivaatio ja osallistuminen	• Ranking tuottaa epätasa-arvoa luokahuoneessa: parhaat oppilaat saavat enemmän resursseja. • Ranking voi lisätä kilpailua ja motivaatiota, mutta heikentää tasapuolista osallistumista ja oppimiskokemusta.
99	Noboa (2013)	USA, Texas	16 opettajaa	Rakenteellisten tekijöiden (standardit, testaus, curriculum) vaikutus opetussisältöihin ja kulttuurisesti relevanttiin pedagogiikkaan	Ei verrokkiryhmää (haastattelut)	Opetussisällöt latino kontekstissa, opetusmenetelmät, opettajien autonomia, rakenteelliset rajoitteet, oppilaiden kiinnostus	• Opetusta ohjaavat vahvasti curriculum-standardit ja korkean panoksen testit. • Opettajilla halua sisällyttää kulttuurisesti relevanttia sisältöä, mutta aika-, resurssi- ja testipaineet estävät sen.
100	Nortvedt (2018)	Norja	Ei yksilötason otosta; kansallinen koulutusjärjestelmä ja PISA-aineisto (2000–2015),	PISA:n vaikutus koulutuspolitiikkaan ja matematiikan opetukseen	Ei verrokkiryhmää (politiikka-analyysi ja PISA-tulosten ja reformien tarkastelu)	Kansalliset politiikkamuutokset (testit, curriculum, teacher education), PISA-tulokset (keskiarvot,	• PISA aiheutti Norjassa “PISA-shokin” (2001) johtaen kansallisiin testeihin, arviointijärjestelmään ja curriculum-reformeihin.

			politiikkadokumen- tit ja reformit			trendit, equity), oppimistulosten jakaumat (low/high achievers), oppilasasenteet	- Kansalliset kokeet suunniteltiin PISA-viitekehityksen mukaisiksi. - Huomio siirtynyt rankingeista heikkoihin ja huippuoppilaisiin, mutta järjestelmä säilyttää silti vahvan tasa-arvo fokuksen.
101	O’Sullivan ym. (2020)	Australia ja Englanti	33 opettajaa (18 Australiasta ja 15 Englannista)	Korkean panoksen testauksen, standardoinnin ja koulutuspolitiikan vaikutus opettajien autonomiaan, identiteettiin ja pedagogiikkaan	Ei verrokkiryhmää (syvähaastattelut)	Opettajien kokemukset: autonomia, työkuorma, pedagogiset valinnat, suhde testaukseen, ammatillinen identiteetti, hyvinvointi.	- Korkean panoksen testaus ja standardointi johtavat opetuksen kaventumiseen ja opettajien autonomian heikkenemiseen. - Opettajat kokevat voimakasta painetta, stressiä ja työuupumusta.
102	Oh ym. (2016)	USA	Ei empiiristä otosta; kansallinen koulutuspolitiikka konteksti (politiikkadokumentit, teoria).	Kansallisen opetussuunnitelman ja standardoidun testauksen merkitys ja vaikutukset opetukseen ja arviointiin	Ei verrokkiryhmää (teoreettinen kirjallisuuteen perustuva analyysi)	Tarkastellaan opetuksen tavoitteita, opettajien tulosvastuuta, arviointikäytäntöjä	Kansallinen opetussuunnitelma voisi tuoda selkeyttä, yhtenäisyyttä ja parempaa tavoitteiden ja arvioinnin linjausta, mutta liian jäykkä standardointi johtaa opetuksen kaventumiseen, testikeskeisyyteen ja paikallisen kontekstin sivuuttamiseen.
103	Ömür ym. (2020)	Turkki	12 opettajaa yläkoulusta	Korkean panoksen testauksen vaikutus opettajien ammatilliseen identiteettiin ja opetuskäytäntöihin	Ei verrokkiryhmää (syvähaastattelut)	Opettajien kokema paine, autonomia, työkuorma, opetusmenetelmät, motivaatio, opettajaidentiteetti	- Korkean panoksen testaus muokkaavat merkittävästi opettajan työn luonnetta ja heikentävät ammatillista autonomiaa, muuttaen opettajan roolin pedagogisesta asiantuntijasta mekaaniseksi ”teknikoksi”. - Opetusta muuttuu testikeskeiseksi, lisää työstressiä ja kaventaa koulutuksen tarkoitusta.
104	Park (2022)	USA, NY	~1 000 000 opiskelijaa; yli 4,5 miljoonaa	Lämpötilan vaikutus korkean panoksen testisuoritukseen ja	Kvasikokeellinen asetelma (ennen-jälkeen)	Koepisteet (z-score), läpäisy, valmistuminen (high school graduation),	Korkea lämpötila heikentää suoritusta merkittävästi (-0.009

			koesuoritusta; korkean panoksen Regents-kokeet	pitkän aikavälin koulutuloksiin	Hyödynnetään sitä, että opiskelijaa seurataan yli ajan ja yhdistetään säädataan testipäiviä koskien	opettajien arviointikäytännöt	SD per °F ja jopa -13 % SD 90°F vs. 75°F). - Vähentää läpäisyä ja valmistumista ja vaikutukset ovat pysyviä. - Opettajat kompensoivat nostamalla arvosanoja kuumina päivinä.
105	Puspitasari (2024)	Indonesia	8 opettajaa yläkoulusta	Korkean panoksen - kokeiden washback ja sosio-kulttuuriset tekijät	Ei verrokkiryhmää (puolistrukturoidut haastattelut)	Opettajien kokemukset, pedagogiset valinnat, sosiaalinen paine, opetuskäytännöt	Washback ei selity pelkällä opetuksella, vaan sosiaaliset ja kulttuuriset tekijät (paine, kollektivismi) muokkaavat voimakkaasti opetusta
106 *	Putwain ym. (2015)	Englanti	1433 oppilasta 18 yläkoulusta (luokat 10–11), kolme mittauskertaa.	Opettajien tekemä pelottelu, niihin perustuvat viestit ja niiden tulkinta (challenge vs. threat)	Kontrollitutkimus (hyödyntää kyselyaineistoa ja tietoa, kolme mittauskertaa)	Akateeminen minäpystyvyys, pelottelun tulkinta	Pelottelu vaikuttavat eri tavoin: haasteena tulkittuna lisäävät motivaatiota ja minäpystyvyyttä, uhkana tulkittuna heikentävät niitä.
107 *	Putwain ym. (2014)	Englanti	347 oppilasta kahdesta koulusta	Pelotteluun perustuvat opettajaviestit ja niiden vaikutus motivaatioon ja koeksuorituksiin	Kontrollitutkimus (hyödyntää kyselyaineistoa ja tietoa, kolme mittauskertaa)	Pelotteluun perustuvat opettajaviestit, GCSE- koepisteet	Useammat ja uhaaviksi koetut viestit liittyvät heikompaan sisäinen motivaatioon ja heikompiin koetuloksiin.
108	Ramsay-Jordan (2020)	USA	4 opettaja opiskelijaa	Korkean panoksen testauksen vaikutus kulttuurisesti responsiivisen opetuksen toteuttamiseen	Ei verrokkiryhmää (haastattelut)	Opettajaopiskelijoiden kokemukset: autonomia, opetuskäytännöt, mahdollisuudet toteuttaa opetussuunnitelma.	Korkean panoksen testaus ja pacing guides rajoittavat voimakkaasti opettajien mahdollisuuksia toteuttaa kulttuurisesti responsiivista opetusta.
109	Rappley ym. (2018)	Japani, USA, UK, Australia	Japani: NHK:n koululaistutkimukset (~557–1350 per kierros) ja PISA 2015 (15-v.) otokset Japanista, Yhdysvalloista, Iso-Britanniasta, Australiasta	Tutkimus ei arvioi yksittäistä reformia vaan tarkastelee pitkän aikavälin muutoksia koepainetta koskien Japanissa suhteessa muihin maihin ja epäilee länsimaiden	Ei verrokkiryhmää (kuvataan trendejä ja keskiarvoja)	PISA-väittämät: huoli kokeista, jännitys, jännitys oppiessa, huoli huonoista arvosanoista; opiskeluajan paineen lähteet (vanhemmat, opettajat), kilpailu yliopistopaikoista	- Japanin “exam hell” -stereotypia on empiirisesti virheellinen. - Japanilaisten oppilaiden ahdistus ei ole korkeampi kuin länsimaissa. - Stressitasot ovat samankaltaisia tai joissain tapauksissa matalampia ja lisäksi

				stereotyyppistä selitystä ("exam hell")			opiskelu-aika ei ole poikkeuksellisen korkea.
110	Alqahtani ym. (2024)	Saudi-Arabia	42 opettajaa (K12)	Tarkastelee kahta kilpailevaa pedagogista lähestymistapaa: opettaja- ja oppilaskeskeistä opetusta	Ei verrokkiryhmää (kyselytutkimus WhatsApp ryhmien kautta)	Kvantitatiivinen kyselytutkimus ja avoimet vastaukset	Opettajat pitävät molempia lähestymistapoja tärkeinä, mutta käytännössä opetuksessa on ristiriita: monet luulevat käyttävänsä yhtä lähestymistapaa, mutta toimivat toisin.
111	Redman (2013)	USA, Arizona	Neljä K12-opettajaa	Tarkastellaan "sustainability education" -integraatiota USA:n K-12-kouluissa	Ei verrokkiryhmää (haastattelut)	Opettajien kokemukset: esteet, pedagogiset käytännöt, opetuksen toteutus	- Standardoidut testit ja opetussuunnitelmapaineet estävät siirtymistä kokeellisiin ja oppilaskeskeisiin menetelmiin. - Opettajat ovat kuitenkin motivoituneita, ja oppilaiden sitoutuminen kasvaa merkittävästi.
112	Redon (2020)	Chile	99 opettajaa ja 2 fokusryhmää (perus- ja toisen asteen koulut)	Tutkimus tarkastelee opettajien kokemuksia kansalaiskasvatuksesta Chilen koulutusjärjestelmässä	Ei verrokkiryhmää (haastattelut ja fokusryhmäaineisto)	Opettajien diskurssit: neoliberalismi, kilpailu, yhteisöllisyys, opetuksen merkitys	Kilpailu, standardoidut testit ja byrokratia heikentävät yhteisöllisyyttä ja kansalaiskasvatusta sekä kaventavat opetuksen merkitystä.
113	Rivas ym. (2020)	Argentiina, Brasilia, Chile, Kolumbia, Meksiko, Peru, Uruguay	134 haastattelua + yli 500 politiikkadokumenttia ja artikkelia; politiikka-analyysi 2000–2015	Koulutuksen regulatorisen hallinnan muutos (governance turn): miten opetus, arviointi ja opetussuunnitelma ovat tulleet tiukemman sääntelyn piiriin	Ei verrokkiryhmää (politiikka-analyysi, dokumenttianalyysi, asiantuntijakysely ja puolistrukturoidut haastattelut)	Opettaja-arviointi, opetussuunnitelma, testaus, koulun autonomia	Latinalaisessa Amerikassa on tapahtunut "regulatory governance turn", jossa testaus, standardit ja arviointi ovat lisänneet opetuksen kontrollia ja johtaneet opetuksen standardoitumiseen ja "teaching to the test" -ilmiön kasvuun.
114	Roda (2017)	USA	Seitsemän K–8 koulua joista 47 yksilöhaastattelua ja 13 fokusryhmää	Neoliberaalien politiikkavaikutusten ilmentyminen pidennettynä koulupäivänä (=ELT)	Ei verrokkiryhmää (syvähaastattelut, fokusryhmät ja havainnointi)	ELT:n toteutus, opetuksen sisältö, testipainotus, koulutuksellinen eriarvoisuus, lähikoulu-muut -koulujen erot	ELT toteutuu epätasa-arvoisesti: muut kouluissa lisäaika käytetään rikastavaan opetukseen, mutta lähikouluissa pääosin testiharjoitteluun ja interventioihin.

							Korkean panoksen testaus, koulushoppailu ja segregaatiot vahvistavat eroa
115	Saeki ym. (2015)	USA	Kirjallisuuskatsaus CCSS- ja tulosvastuupolitiikoista	Korkean panoksen kokeet ja Common Core -politiikan psykologiset ja opetukselliset vaikutukset	Ei verrokkiryhmää (narratiivinen kirjallisuuskatsaus)	Opettajien hyvinvointi (stressi, motivaatio, moraali), opetuskäytännöt, oppiminen	Testipohjainen tulosvastuu lisää opettajien stressiä, burnoutia ja motivaation laskua sekä johtaa opetuksen kaventumiseen, teaching to the test -ilmiöön ja epäoptimaalisiin opetuskäytäntöihin.
116	Sanzo ym. (2011)	USA	10 menestyvää middle school -rehtoria	Menestyvien rehtorien johtamiskäytännöt tulosvastuu ympäristössä	Ei verrokkiryhmää (haastattelu)	Johtamiskäytännöt: yhteistyö, ammatillinen kehitys, opetuksen johtaminen, avoimuus	Menestyvät rehtorit korostavat jaettua johtajuutta, systemaattista ammatillista kehitystä, vahvaa pedagogista johtajuutta ja avointa kulttuuria, jotka tukevat oppimistuloksia tulosvastuujärjestelmässä
117	Schoenbach ym. (2010)	USA	Ei ole	Lukutaidon kehittäminen sisältöaineissa	Ei verrokkiryhmää (Kuvaileva kehittämisartikkeli)	Lukustrategiat, opettajien lukuprosessit, oppilaiden lukutaidon kehittäminen	Sisältöalueisiin sidotut lukustrategiat voivat parantaa oppilaiden lukutaitoa ja ymmärtämistä eri oppiaineissa.
118	Scott (2012)	Englanti	267 vastausta opiskelijakyselystä, 14 opiskelijahaastattelua kolmesta oppilaitoksesta	Resit-politiikan eli uusintakokeiden vaikutus oppimiseen ja arvioinnin validiteettiin (A-levels)	Ei verrokkiryhmää (kysely, pitkittäiset haastattelut ja asiantuntijahaastattelut)	Opiskelijoiden oppimisstrategiat, motivaatio, resit-käyttäytyminen, oppimisen syvyys, arvioinnin oikeudenmukaisuus	Resit-järjestelmä (uusintakokeiden mahdollisuus) parantaa arvosanoja, mutta johtaa pinnalliseen oppimiseen, "teaching to the test" -strategioihin, taktiikkapohjaiseen opiskeluun ja heikentyneeseen oppimisen laatuun.
119	Corkin ym. (2018)	USA	24 matematiikan opettajaa (K-12), jotka osallistuivat intensiiviseen koulutukseen	Oppilaslähtöisen opetuksen toteutuksen esteet ja mahdollistajat	Ei verrokkiryhmää (puolistrukturoidut haastattelut ja teemallinen analyysi)	Opettajien kokemukset, opetuskäytännöt, esteet, fasilitaattorit	Oppilaslähtöisen opetuksen toteutus estyy erityisesti testauskulttuurin, resurssipulan, köyhyyden, hallinnollisten rajoitteiden ja opettajien teoreettisen ymmärryksen puutteen vuoksi.

120 *	Sieverding ym. (2019)	Egypti	SYPE-paneeliaineisto (2009 & 2014) ~10 000–15 000 nuorta ja 24 syvähaastattelua	Yksityisopetuksen (private tutoring) syyt ja dynamiikka	Kontrollitutkimus ja haastattelut	Tutoringin käyttö, kustannukset, syyt (exam pressure, school quality, teacher incentives)	Yksityisopetus on rakenteellisesti sidottu korkean panoksen kokeisiin, heikkoon koulun laatuun ja opettajien kannustimiin, ja siitä on tullut lähes pakollinen osa koulutusjärjestelmää.
121	Smyth ym. (2012)	Irlanti	897 alemman toisen asteen oppilasta, 748 ylemmän asteen oppilasta ja ryhmähaastattelut	Korkean panoksen kokeen vaikutus opiskelijoiden kokemuksiin	Ei verrokkiryhmää (kysely ja syvähaastattelut)	Stressi, työmäärä, oppimisstrategiat, opiskelijakäsitykset opetuksesta	Korkean panoksen kokeet lisäävät stressiä, kaventavat opetusta ja muuttavat opiskelijoiden oppimiskäsityksiä instrumentaalisemmiksi.
122	Solanki Ym. (2020)	USA ja UK	16 opettajaa ja standardoidut kokeet (Regents & GCSE)	Korkeamman vs matalamman ajattelun (HOT vs LOT) osuus testeissä ja opetuksessa	Ei verrokkiryhmää (testien item-analyysi ja luokkahuone havainnot)	HOT vs LOT -tehtävien määrä, opetuksen kognitiivinen taso	Testit ja opetus painottuvat matalan tason ajatteluun ja tästä syystä kriittinen ajattelu jää vähäiseksi.
123 *	Somers ym. (2011)	USA	Neljä suurta IES:n rahoittamaa moniosavaltiolista satunnaistettua kokeellista tutkimusotosta.	Miten osavaltioiden tason testit eroavat valtion kokeilutestien tuloksista	Satunnaistettu kontrolloitu koe (RCT). Neljä erillistä rct:tä.	Testitulokset (state vs. study tests), vaikutusarviot	Paikalliset osavaltioiden testit sopivat yleisen oppimisen mittaamiseen, mutta eivät spesifien taitojen arviointiin.
124	Soto-Peña ym. (2024)	USA	Kaksi koulua ja niiden opettajat ja oppilaat; havainnot, haastattelut, dokumentit	Culturally sustaining pedagogy (CSP) ja sosiokulttuurinen kompetenssi DLBE-ohjelmissa	Ei verrokkiryhmää (haastattelut ja havainnot)	Opettajien ja oppilaiden kokemukset, opetuskäytännöt, curriculum, testauspaineet	Korkean panoksen testaus ja standardoitu lukujärjestys ohjaavat opetusta pois oppilaiden kokemuksista ja kriittisestä pedagogiikasta, mikä johtaa valkoisuutta ylläpitävien käytäntöjen uusintamiseen.
125	Spina (2017)	Australia	6 koulua; opettajat, rehtorit, oppilaat; usean vuoden etnografinen aineisto	6 koulua; opettajat, rehtorit, oppilaat; usean vuoden etnografinen aineisto	Ei verrokkiryhmää (haastattelut)	Opettajien työ, opetussisältö, kirjoittamisen opetustavat, oppilaiden kokemukset	Testaus ja datapaine ohjaavat opetusta kapeaan, tekniseen kirjoittamiseen ja heikentävät luovuutta sekä oppilaiden oppimiskokemusta.

126	Sporre (2019)	Ruotsi	Aineistona opetussuunnitelma, kansalliset kokeet 2013, opettajaohjeet; haastattelut testien laatijoista (2 kpl)	Etiikan opetuksen arviointi kansallisten kokeiden avulla (peruskoulu, luokat 6 ja 9)	Ei verrokkiryhmää (dokumenttianalyysi ja haastattelut)	Eettinen kompetenssi, erityisesti argumentointi, käsitteiden ja mallien käyttö	- Kansalliset kokeet mittaavat kapeasti argumentatiivista ja käsitteellistä osaamista. - Laajemmat eettiset taidot jäävät mittaamatta, ja tämän takia kansallisiin kokeisiin perustuva arviointi on haitallinen etiikan opetukselle.
127	Strandler (2017)	Ruotsi	10 kokenutta opettajaa alakoulun 6. luokalta.	Tulosperusteiset koulutusreformit: kansalliset kokeet, arvioinnin standardointi ja arvosanojen käyttöönotto	Ei verrokkiryhmää (haastattelut ja luokahuonehavainnot)	Opettajien toiminta, opetuskäytännöt, merkityksenanto, opetuksen sisältöpainotukset	- Reformit siirsivät opetusta kohti helposti mitattavaa, sisältökeskeistä opetusta. - Avoimempi, oppilaslähtöinen ja kriittinen oppiminen väheni. Tämä kaventaa oppiaineen monimuotoisuutta ja voi heikentää kriittistä ajattelua.
128	Stroup ym. (2011)	USA	Yliopisto-opiskelijat (STEM-opiskelijat opettajankoulutuksessa)	Psykometriset arviointikäytännöt ja tilastollisen mittaamisen (erityisesti keskiarvon ja testauksen) käyttö opetuksessa	Ei verrokkiryhmää (teoreettinen tutkimus sisältäen demonstraatioperustaisen lähestymistavan)	Opiskelijoiden ajattelutavat, tilastollinen ymmärrys, käsitykset mittaamisesta ja arvioinnista	- Keskiarvon käyttö voi vääristää oppimisen ilmiötä: se ei välttämättä kuvaa kenenkään ajattelua. - Standardoidut testit rakentuvat tilastoista, mutta tämä voi katkaista yhteyden todelliseen oppimiseen ja merkityksenmuodostukseen.
129	Thompson ym. (2012)	USA	Analyttinen katsaus ja hyödyntää laajoja sekundäärilähteitä Kohteena K–12-ikäiset afroamerikkalaiset oppilaat.	Korkean panoksen standardoidut testit (erityisesti No Child Left Behind Act, NCLB), standardoidut testit, tulosvastuu ja niiden vaikutus afroamerikkalaisiin oppilaisiin.	Ei verrokkiryhmää (narratiivinen kirjallisuuskatsaus)	Oppimistulokset (testitulokset, valmistumisasteet), oppilasasenteet (apatiisuus), kurinpitotoimet, koulutusjärjestelmän toiminta	- Korkean panoksen testaus ei ole merkittävästi parantanut oppimistuloksia, mutta sillä on useita haitallisia vaikutuksia afroamerikkalaisiin oppilaisiin. ”Narsistinen” koulutusjärjestelmä, joka keskittyy testituloksiin todellisen oppimisen sijaan.
130	Thompson ym. (2014)	Australia	Oppilaat vuosiluokilla 3, 5, 7 ja 9 (NAPLAN-kansallinen)	Korkean panoksen standardoidut testit (NAPLAN) ja niihin	Ei verrokkiryhmää (teoreettinen analyysi ja media raporttien tarkastelu)	Miten korkean panoksen testaus muuttaa opettajuuden ehtoja; miten digitaalinen data	Korkean panoksen testaus muuttaa opetusta: opettajan työ arvioidaan yhä enemmän datan (testitulosten) kautta. Tämä

			testaus, ei yksittäistä otoskokoa)	liittyvä dataperusteinen arviointi		ja tietokannat muokkaavat "hyvän opettajan" käsitettä; minkälaisia datamanipulaation strategioita opettajat ja koulut käyttävät	johtaa "datan manipulointiin" (ei vain huijaamiseen), joka on looginen seuraus järjestelmästä. Opettajuus siirtyy moraaliseen "hyvästä opettajasta" kohti "jaettua perustuvaa opettajaa" (dividuated teacher).
131	Town ym. (2015)	USA	Artikkelissa kuvattu luokkahuonekokeilu yläkoulun algebran oppilaiden kanssa	Kuvaileva opetuskokeilu: oppilaat keräävät dataa ajoneuvojen liikkeestä analysoivat pisteitä, muodostavat line of best fit -suoran ja esittävät tulokset ja keskustelevat slope ja intercept -merkityksistä	Ei verrokkiryhmää (opetuskokeilu ja kvalitatiivinen havainnointi)	Oppiminen, osallistuminen, yhteistyö, käsitteellinen ymmärrys	Konkreettinen, dataan perustuva oppiminen parantaa käsitteellistä ymmärrystä (esim. kulmakerroin = nopeus), lisää oppilaiden aktiivisuutta ja yhteistyötä sekä tukee syvempää matemaattista ajattelua verrattuna pelkkään symboliseen opetukseen.
132 *	Troncoso ym. (2016)	Chile	~202 605 oppilasta (13–16 v.) ja siirtymä peruskoulusta toisen asteen alkuun, 2 438 koulua	Koulun vaikuttavuuden mittaaminen (value-added -mallit) ja koulutuksen tilivelvollisuusjärjestelmä (accountability)	Kontrollitutkimus (lisätään eri tason kontroleja)	Oppimistulokset, oppimisen edistyminen, koulu-, luokka- ja alue-efektit, sosioekonomisten tekijöiden vaikutus	Perinteiset 2-tasoiset mallit yliarvioivat koulun vaikutusta, jopa 21 % koulutason varianssista selittyy todellisuudessa luokka- ja paikallistason eroilla.
133	Tveit (2018)	Norja ja Ruotsi	Aineistona 13 (Norja) ja 31 (Ruotsi) politiikkadokumenttia sekä asiantuntijahaastattelut (ministeriöt ja virastot)	Kansallisten kokeiden rooli koulutuspolitiikassa (2000–2017): sertifiointi, ohjaus (governance) ja oppimisen tukeminen	Ei verrokkiryhmää (dokumenttianalyysi ja asiantuntijahaastattelut)	Arvioinnin tarkoitukset (certify, govern, support), politiikan muutokset ajan yli, roolien väliset ristiriidat	- Kansallisten kokeiden tavoitteet ovat muuttuneet ja laajentuneet ristiriitaisiksi. - Useiden tarkoitusten yhdistäminen (erityisesti formative + summative) aiheuttaa jännitteitä ja heikentää politiikan selkeyttä ja toimivuutta.
134	Van Hover ym. (2012)	SUA	1 luokka (n. 21 oppilasta, 9. lk), joista osa erityisen tuen oppilaita	Yhteisopetus (co-teaching) inklusiivisessa luokassa korkean panoksen testauksen kontekstissa	Ei verrokkiryhmää (havainnot, haastattelut, opetusaineistot ja reflektiot)	Opettajien yhteistyö, opetuksen toteutus, testipaineen vaikutus, opetuksen sisältö ja tavoitteet	Yhteisopetus toimii hyvin yhteistyönä, mutta korkean panoksen testaus ohjaa opetusta vahvasti kohti faktamuistamista ja "teaching to

							the test” ilmiötä, mikä rajoittaa syvällistä oppimista.
135	Vetter ym. (2012)	USA	Kaksi lukion englannin luokkaa (n. 48 oppilasta, grades 9–12)	Reflektiivinen arviointikäytäntö: daybook defense (suullinen reflektio kirjoituspäiväkirjasta)	Ei verrokkiryhmää (oppilasvihat (240 merkintää), videoidut keskustelut ja sisällönanalyysi)	Oppilaiden lukija- ja kirjoittajaidentiteetti, reflektiivinen ajattelu, osallistuminen	Reflektiivinen arviointi vahvistaa oppilaiden identiteettiä lukijoina ja kirjoittajina sekä lisää tietoisuutta omista oppimisprosesseista.
136	Vosgano ff ym. (2011)	Australia	16 oppilasta, 9. lk (n. 14-vuotiaat), kuulovammaisia; vertailu koulun ja valtion keskiarvoihin	Inklusiivinen koulutus ja standardoidut testit	Ei verrokkiryhmää (testitulosten vertailu 2005–2007)	Matematiikan ja luonnontieteiden osaaminen, kielitaidon vaikutus, tehtävätyyppien vaikeus, osa-alueiden erot	- Suurin osa kuulovammaisista oppilaista suoriutui heikommin kuin ikätoverinsa. - Kielelliset vaatimukset selittävät merkittävän osan suoriutumiserosta
137	Wahleit hner (2018)	USA, Kalifornia	Viisi lukion englannin opettajaa (case) + 171 opettajan kysely	Kirjoittamisen opetus, opettajien osaaminen ja koulutus, koulutuspolitiikka	Ei verrokkiryhmää (kysely ja haastattelu)	Opettajien pedagoginen osaaminen, opetuskäytännöt, politiikan paineet, oppilaiden kirjoitustaidot	- Opettajien osaaminen ja konteksti vaikuttavat ratkaisevasti opetukseen. - Korkean panoksen testaus kaventaa kirjoittamisen opetusta.
138	Wang (2016)	Etelä-Korea	~1800 toisen asteen opiskelijaa (kaksi pitkittäisaineistoa)	Korkean panoksen koe ja sen vaikutus mielenterveyteen	Kontrollitutkimus (lineaarinen regressio)	Itsetuhoiset ajatukset (dikotominen: ”Tunnetko joskus itsetuhoisia ajatuksia ilman selvää syytä?”, elämäntyytyväisyys, testisuoriutuminen suhteessa odotuksiin	- Odotuksia heikompi suoriutuminen lisää itsetuhoisia ajatuksia noin 17 %. - Vaikutus heikkenee ajan myötä, vuoden kuluttua ei merkitsevä. - Placebo: Ei vaikutusta matalan panoksen kokeissa. - Kyse ei ole huonoista oppilaista: Ilmiötä voi näin selittää odotusten ja toteuman suhde.
139	Wang (2020)	Kiina	Seitsemän yläkoulun fokusoppilasta (13–16 v.)	Maaseutu–kaupunki -migratio ja urbaani koulutus, testaus ja sosiaalinen integraatio	Ei verrokkiryhmää (haastattelut)	Oppimisen sopeutuminen, suoriutuminen, kieli, sosiaalinen integraatio, identiteetti, stressi,	- Maaseudulta muuttaneet vähemmistöoppilaat kohtaavat merkittäviä oppimis-, kieli- ja integraatiohaasteita; - Korkean panoksen testaus lisää painetta ja kilpailua.
140	White ym. (2012)	Australia	Yksi koulu; 10 matematiikan	NAPLAN-kansallisen testauksen käyttö opetuksen ohjauksessa;	Ei verrokkiryhmää	Opettajien valitsemat opetusstrategiat, opetuskäytännön	Pelkkä testiin harjoittelu kaventaa opetusta, mutta NAPLAN-datan käyttäminen

			opettajaa, vuosiluokat 7 ja 9	datan hyödyntäminen ammatillisissa oppimiskeskusteluissa	(opettajien kyselyt ja haastattelut, tuntihavainnot)	muutokset, opettajien asenteet/varmuus, oppilaiden numeriiikan oppiminen ja testitulosten kehitys	yhteisissä ammatillisissa keskusteluissa auttoi opettajia kohdentamaan opetusta oppilaiden virhekäsityksiin ja lukutaitovaatimuksiin.
141 *	Wiberg (2019)	Ruotsi	4 090 oppilasta, 8. lk (TIMSS 2015; 150 koulua) ja rekisteridata (arvosanat ja kansalliset kokeet 6. ja 9. luokalta)	TIMSS-matematiikan tuloksia verrataan erikseen sekä kansallisiin testituloksiin että opettajan antamiin arvosanoihin	Kontrollitutkimus (lineaarinen regressio)	Matematiikan osaaminen (TIMSS), arvosanat, kansallisten kokeiden tulokset, sosioekonomiset taustatekijät, koulukonteksti	- TIMSS-tulokset korreloivat vahvasti arvosanojen ja kansallisten testien kanssa, mutta sosioekonominen tausta selittää merkittävän osan vaihtelusta. Koulutason vaikutus on suhteellisen pieni. - Kansalliset kokeet ja opettajan arvosanat korreloivat vahvasti.
142 *	Wiberg ym. (2019)	Ruotsi	4 090 oppilasta, 8. lk (TIMSS 2015; 150 koulua) ja rekisteridata (arvosanat ja kansalliset kokeet 6. ja 9. luokalta)	TIMSS-luonnontieteen tuloksia verrataan erikseen sekä kansallisiin testituloksiin että opettajan antamiin arvosanoihin	Kontrollitutkimus (lineaarinen regressio)	TIMSS luonnontieteiden osaaminen, arvosanat, kansalliset testitulokset, sosioekonominen tausta	- TIMSS-tulosten ja kouluarvioinnin välinen yhteys on keskivahva. - TIMSS ja kansalliset mittarit ovat yhteydessä, mutta yhteys on vain kohtalainen. Arvosanat ja kansalliset kokeet korreloivat keskenään vahvemmin.
143	Wilkinson ym. (2020)	Englanti	30 oppituntia yhdessä yläkoulussa	Opettajien viestintä korkean panoksen kokeista (GCSE) luokahuoneessa	Ei verrokkiryhmää (havainnointitutkimus ja temaattinen analyysi)	Opettajien viestintätäylyt, motivaatioviestit, kontrolli vs. tuki, arviointiin liittyvät viestit	- Opettajat käyttävät kokeisiin liittyvää viestintää moniin tarkoituksiin: motivointiin, käyttäytymisen kontrollointiin ja ohjeistamiseen. - Viestintä sisältää sekä painostavia että tukevia elementtejä.
144	Winter (2017)		Fokusoidut keskustelut (5 kpl) ja kaksi yksilöhaastattelua	Curriculum policy reform teknisen tilivelvollisuuden kontekstissa	Ei verrokkiryhmää (fokusoidut keskustelut ja puolistrukturoidut haastattelut)	Opettajien kokemukset, opetuskäytännöt, curriculum-muutokset, dataohjaus, opettajan autonomia, oppilaiden asema	- Tekninen tilivelvollisuus kaventaa opetusta, lisää testiohjausta ja muuttaa opettajien ja oppilaiden rooleja. - Data ja arviointi määrittelevät oppilaita ja opettajia sekä heikentävät ammatillista autonomiaa.

145	Wright ym. (2020)	USA	896 oppilasta (6.–8. lk), kaksi middle school -koulua	Korkean panoksen kirjoitustestin (7. lk) vaikutus kirjoitusmotivaatioon eri luokka-asteilla	Ei verrokkiryhmää (kysely ja ryhmävertailut)	Kirjoitusmotivaatio: asenteet, uskomukset kirjoittamisesta ja uskomukset itsestä kirjoittajana	7.-luokkalaisten motivaatio oli selvästi heikompaa kuin 6. ja 8. luokilla. Erityisesti asenteet ja kirjoittamisen arvo laskivat testivuonna.
146	Wuthrich ym. (2021)	Australia	638 opiskelijaa (lukion viimeinen vuosi, n. 16–18 v)	Korkean panoksen loppukokeisiin liittyvä paine lukion viimeisenä vuonna	Kontrollitutkimus (selitetään painetta, lineaarinen regressio 4 mittauskertaa)	Stressi, ahdistus, masennus ja moderaattorit.	Stressi kasvoi vähän (0.2 sd) ja ahdistus paljon (1.7 sd) vuoden aikana, erityisesti ennen kokeita.
147	Wuthrich ym. (2020)	17 eri maata	60 tutkimusta; kohderyhmä: lukion viimeiset 1–2 vuotta (n. 16–19 v)	Akateeminen stressi ja korkean paineen kokeet lukion loppuvaiheessa.	Ei verrokkiryhmää (systemaattinen kirjallisuuskatsaus)	Stressi, ahdistus, masennus, testiahdistus ja riskitekijät	- Noin 1/6 opiskelijaa kokee erittäin korkeaa stressiä. - Stressi kasvaa kohti kokeita ja yksilölliset tekijät (esim. testiahdistus, perfektionismi) keskeisiä.
148	Xiao (2017)	Kiina	3 lukioluokkaa (~180 oppilasta), 3 opettajaa	Formatiivinen arviointi testikeskeisessä (high-stakes) koulutusjärjestelmässä	Ei verrokkiryhmää (haastattelut)	Oppiminen, palautteen käyttö, oppilaiden osallistuminen, opettajien strategiat	Summatiivisia kokeita voidaan käyttää formatiivisesti (test follow-up), mutta vaikutus riippuu opettajan tavasta toteuttaa palautetta ja oppilaiden aktiivisuudesta.
149	Xiao ym. (2023)	Kiina	~9800 oppilasta, 15-vuotiaat (PISA 2009)	Formatiivisen arvioinnin vaikutus oppimistuloksiin (lukutaito)	Kontrollitutkimus	Lukutaito, lukumotivaatio, formatiiviset käytännöt	- Formatiivisella arvioinnilla pieni mutta positiivinen kokonaisvaikutus oppimistuloksiin. - Motivaatio toimii keskeisenä välittävänä tekijänä.
150	Yazıcılar ym. (2019)	Turkki	5 matematiikan opettajaa (lukio, 9. lk; eri koulutyypit)	Opetussuunnitelman toteutus ja muokkaaminen keskitetyn järjestelmän ja korkean panoksen testien kontekstissa	Ei verrokkiryhmää (haastattelut, havainnot, dokumenttianalyysi)	Opettajien adaptaatiokäytännöt, syyt adaptaatioon, opetuksen muutokset, autonomia	Opettajat mukauttavat opetussuunnitelmaa systemaattisesti oppilaiden tason ja korkean panoksen testien mukaan, vaikka virallisesti heidän tulisi noudattaa sitä sellaisenaan.
151	Yedigöz-Kara ym. (2022)	Useita maita	10 laadullista tutkimusta (opettajat)	Eri maiden englannin opetuksen opetussuunnitelmauudistukset (CLT-reformit)	Ei verrokkiryhmää	Implementaation esteet eri tasoilla, opettajien toiminta ja uskomukset	Opetussuunnitelmauudistukset epäonnistuvat usein, koska järjestelmät (macro–micro) eivät ole koordinoituja.

							• Erityisesti testipolitiikka, tuen puute ja opettajien valmiudet estävät toteutuksen.
152 *	Yell ym. (2012)	USA	Ei perinteistä otosta; oikeustapauksia ja koulutuspolitiikkaa koskeva analyysi	Korkean paineen kokeet ja niiden vaikutus erityisoppilaisiin (students with disabilities)	Ei verrokkiryhmää (oikeustapausanalyysi (case law review) ja politiikka-analyysi)	Oikeudelliset periaatteet: syrjimättömyys, FAPE, testien validiteetti, mukautukset	• Kokeet ovat laillisia myös erityisoppilaille, mutta vain tietyin ehdoin. • Käytännössä ne asettavat merkittäviä haasteita erityisoppilaille ja voivat rajoittaa valmistumista.
153	Yin ym. (2015)	Kiina	Yksi lukion kemian luokka (54 oppilasta, ~16–17 v) ja yksi opettaja	Formatiivisen arvioinnin integrointi korkean panoksen testauskontekstissa	Ei verrokkiryhmää (yhteistoiminnallista toimintatutkimusta)	Oppilaiden ajatteluprosessit, oppiminen, opettajan käytännöt ja käsitykset, palautteen käyttö	• Formatiivinen arviointi paljastaa oppilaiden ajattelun, muuttaa opetusta ja parantaa oppimista • Toteutus on mahdollista myös testikeskeisessä järjestelmässä, mutta vaatii merkittävää opettajan ajattelun muutosta ja tukea.
154 *	Young (2020)	Kiina	18 lukiolaista (17–18 v), vuoden seuranta	Varjokoulutus (korkean panoksen testauksen kontekstissa)	Ei verrokkiryhmää (seuranta)	Oppilaiden kokemukset, merkityksenanto, oppimiskäsitykset, valtasuhteet	• Yksityisopetus vahvistaa testijärjestelmän tuottamaa ”oppressiota” lisäämällä pinnallista oppimista, auktoriteettisuhteita ja suoriutumiskeskeisyyttä.
155	Zavacky ym. (2017)	USA	Ei empiiristä otosta; perustuu aiempaan tutkimuskirjallisuuteen ja politiikkasuosituksiin (K12)	Välitunnin (recess) vähentyminen korkean panoksen testauksen seurauksena	Ei verrokkiryhmää (narratiivinen katsaus ja politiikka-analyysi)	Fyysinen aktiivisuus, kognitio, käyttäytyminen, akateeminen suoriutuminen	• Välitunti tukee oppimista, käyttäytymistä ja hyvinvointia. • Testaus on vähentänyt välitunteja, mikä voi heikentää oppimistuloksia ja hyvinvointia.
156	Zhang ym. (2013)	Kiina	733 opettajaa (81 % vastasuprosentti) (13-15 v. oppilaat epäsuorasti) ja 9 opettajahaastattelua	Opettajien uskomukset opetussuunnitelmauudistuksen kontekstissa	Ei verrokkiryhmää (Kysely ja puolistrukturoidut haastattelut)	Opettajien pedagogiset uskomukset, käsitykset oppimisesta, roolit, kontekstitekijät	• Opettajien uskomukset ovat moniulotteisia: sekä konstruktivistisia että perinteisiä. • Korkean panoksen testaus ja kulttuuri rajoittavat uudistuksen toteutumista.

157	Zhang (2021)	Useita	Ei yksittäistä otosta; 20 tutkimusta (vanhemmat eri koulutusasteilla)	Vanhempien toiminta korkean panoksen kielitestauksen kontekstissa	Ei verrokkiryhmää (systemaattinen kirjallisuuskatsaus)	Vanhempien osallistumiskäyttäytyminen (PI), testivalmistelustrategiat, vaikutustekijät	Korkean panoksen testaus muuttaa vanhempien käyttäytymistä: lisää testivalmisteluun liittyvää osallistumista, mutta vaikutukset oppimistuloksiin ovat ristiriitaisia.
158	Zohar ym. (2017)	Israel	20 yläkoulun luonnontieteiden opettajaa, kokeneita	Korkean panoksen vs. korkeamman ajattelun taitojen (HOT) opetus	Ei verrokkiryhmää (haastattelut)	Opettajien näkemykset: HOT-opetus, testipaine, oppilaiden erot	Korkean panoksen testaus lisää HOT-tehtäviä, mutta muuttaa opetuksen mekaaniseksi ja heikentää syvällistä oppimista sekä kasvattaa eriarvoisuutta.
159	Özerk ym. (2012)	Norja ja Uusi-Seelanti	Perustuu dokumentteihin, aiempaan tutkimukseen ja asiantuntijahaastatteluihin	Kansalliset arviointijärjestelmät (National Standards NZ; National Testing Norway) ja niiden vaikutus alkuperäiskansojen koulutukseen (Māori, Sámi)	Ei verrokkiryhmää (dokumenttianalyysi ja asiantuntija haastattelut)	Koulutusjärjestelmän muutokset, opetuksen sisältö (teaching to the test), kieli ja kulttuuri, oppilaiden asema, opettajien toiminta, koulupolitiikan vaikutukset	- Kansalliset testit kaventavat opetusta, lisäävät "teaching to the test" -ilmiötä ja uhkaavat alkuperäiskansojen kieliä ja kulttuuria - Järjestelmät voivat lisätä eriarvoisuutta ja heikentää kulttuurisesti relevanttia opetusta.

* Näitä tutkimuksia, joissa oli kontrolliryhmä (tai joka pyrittiin tekemään) ei sisällytetty artikkelin tekstiin.

KIRJALLISUUS

Katsaukseen sisällytetyt uudet artikkelit (24 kpl)

Ahn, Tom (2014) Regression discontinuity design estimates of the effect of high-stakes sanctions on schools. *Economics of Education Review*, 43, 203–216.

Andersen, Simon Calmar & Nielsen, Helena Skyt (2020) Learning from Performance Information. *Journal of Public Administration Research and Theory*, 30(3), 415–431.

Bergbauer, Annika B. & Hanushek, Eric A. & Woessmann, Ludger (2024) Testing. *Journal of Human Resources*, 59(S), S118–S153.

Berry, James & Heissel, Jennifer A. & Kassem, Mariam & Mukherjee, Shobhini & Shotland, Marc (2020) Failure of Frequent Assessment: Deviations from the Curriculum and Learning in India. *Journal of Development Economics*, 145, 102464.

Beuchert, Louise & Eriksen, Tine Louise Mundbjerg & Krogh, Morten (2020) The Impact of Standardized Test Feedback on Remedial Education: Evidence from a Literacy Program. *Economics of Education Review*, 77, 101997.

Brunello, Giorgio & Kiss, David (2022) Math Scores in High-Stakes Exams. *Economics of Education Review*, 86, 102213.

Ebenstein, Avraham & Lavy, Victor & Roth, Sefi (2016) The Long-Run Economic Consequences of High-Stakes Examinations: Evidence from Transitory Variation in Pollution. *American Economic Journal: Applied Economics*, 8(4), 36–65.

Falch, Torberg & Naper, Linn Renee & Rønning, Marte (2013) Educational Evaluation and Gender Differences. *Economics of Education Review*, 36, 12–22.

Fidjeland, Andreas (2023) Using High-Stakes Grades to Track Students: Evidence from a Grading Reform. *Economics of Education Review*, 95, 102434.

Fischer, Mira & Wagner, Valentin (2023) Do Timing and Reference Points Matter? *Economics of Education Review*, 95, 102438.

Goulas, Sofoklis & Megalokonomou, Rigissa (2021) Knowing Who You Are: The Effect of Feedback on Short- and Long-Term Outcomes. *Journal of Economic Behavior & Organization*, 183, 31–53.

Hvidman, Ulrik & Sievertsen, Hans Henrik (2021) High-Stakes Grades and Student Behavior. *Journal of Human Resources*, 56(4), 1012–1043.

Jalava, Nina & Joensen, Juanna Schroter & Pallas, Elin (2015) Grades and Rank: Impacts of Non-financial Incentives on Test Performance. *Journal of Economic Behavior & Organization*, 115, 161–196.

Juerges, Hendrik & Schneider, Kerstin (2010) Central Exit Examinations Improve Student Performance: Evidence from TIMSS. *Journal of Population Economics*, 23(2), 655–682.

Juerges, Hendrik & Schneider, Kerstin & Senkbeil, Martin & Carstensen, Claus H. (2012) Assessment Drives Learning: The Effect of Central Exit Exams on Curriculum-Based Achievement and General Cognitive Ability. *Economics of Education Review*, 31(1), 56–65.

Machin, Stephen & McNally, Sandra & Ruiz-Valenzuela, Jenifer (2020) Entry Through the Narrow Door: The Costs of Just Failing High Stakes Exams. *Journal of Public Economics*, 190, 104224.

Metcalfe, Robert & Burgess, Simon & Proud, Steven (2019) Students' Effort and Educational Resources: The Role of Incentives and Peers. *Journal of Public Economics*, 173, 93–107.

Ofek-Shanny, Yuval (2024) Measurements of Performance Gaps. *Journal of Human Resources*, 59, 1121–1150.

Papay, John P. & Murnane, Richard J. & Willett, John B. (2016) The Impact of Test Score Labels on Human Capital Investment: Evidence from a Regression Discontinuity Design. *Journal of Human Resources*, 51(2), 357–388.

Piopiunik, Marc & Schwerdt, Guido & Woessmann, Ludger (2013) Central School Exit Exams and Labor Market Outcomes: A European Perspective. *European Journal of Political Economy*, 30, 20–30.

Rockoff, Jonah E. & Turner, Lesley J. (2010) Short-Run Impacts of Accountability on School Quality. *American Economic Journal: Economic Policy*, 2(4), 119–147.

Schwerdt, Guido & Woessmann, Ludger (2017) The Information Value of Central School Exams. *Economics of Education Review*, 56, 65–79.

Wilson, Deborah & Burgess, Simon & Briggs, Adam (2011) The Dynamics of School Accountability: The Responses of English Secondary Schools to Performance Tables. *Journal of Population Economics*, 24(2), 407–432.

Winters, Marcus A. & Trivitt, Julie R. & Greene, Jay P. (2010) The Impact of High-Stakes Testing on Student Proficiency in Low-Stakes Subjects: Evidence from Florida. *Economics of Education Review*, 29(1), 138–146

Atjonen ja Ketonen (2025) katsauksen artikkelit (159 kpl)

Abdallah, Muhammad & Taiwo, Aliyu & Ukamaka, Akuche (2022) Influence of High-Stakes Tests on Teaching Styles of Science Teachers in Senior Secondary Schools in Nigeria. *African Educational Research Journal*, 10(3), 291–301.

Acikalin, Mehmet (2014) Future of Social Studies Education in Turkey. *Journal of International Social Studies*, 4(1), 93–102.

Aksoy, Erdem (2023) Teaching or Testing, Which Matters More? The Transition among Education Levels in Turkey. *International Journal of Contemporary Educational Research*, 10(2), 470–483.

- Alsagoafi, Ahmad (2021) Exploring Saudi Students' Perceptions of National Exams: A Washback Study. *Revista Românească pentru Educație Multidimensională*, 13(1), 213–234.
- Amrein-Beardsley, Audrey & Berliner, David C. & Rideau, Sharon (2010) Cheating in the First, Second, and Third Degree: Educators' Responses to High-Stakes Testing. *Education Policy Analysis Archives*, 18(14), 1–36.
- Avalos, Mary A. & Perez, Xuchilt & Thorrington, Vanessa (2020) Comparing Secondary English Teachers' Ideal and Actual Writing Practices for Diverse Learners: Constrained Professionalism in Figured Worlds of High-Stakes Testing. *Reading & Writing Quarterly*, 36(3), 225–242.
- Banks, Joanne & Smyth, Emer (2015) "Your whole life depends on it": Academic stress and high-stakes testing in Ireland. *Journal of Youth Studies*, 18(5), 598–616.
- Barrett, Nathan & Toma, Eugenia F. (2013) Reward or punishment? Class size and teacher quality. *Economics of Education Review*, 35, 41–52.
- Bensnes, Simon Sjøbstad (2016) You sneeze, you lose: The impact of pollen exposure on cognitive performance during high-stakes high school exams. *Journal of Health Economics*, 49, 1–13.
- Bolden, David & Tymms, Peter (2020) Standards in Education: Reforms, Stagnation and the Need to Rethink. *Oxford Review of Education*, 46(6), 717–733.
- Burton, Erin Peters & Frazier, Wendy Michelle (2012) Voices from the Front Lines: Exemplary Science Teachers on Education Reform. *School Science and Mathematics*, 112(3), 179–190.
- Byun, Soo-yong (2010) Does policy matter in shadow education spending? Revisiting the effects of the high school equalization policy in South Korea. *Asia Pacific Education Review*, 11(1), 83–96.
- Cairns, Rebecca (2021) Exams Tested by COVID-19: An Opportunity to Rethink Standardized Senior Secondary Examinations. *Prospects*, 51(1–3), 331–345.
- Camphuijsen, Marjolein K. & Levatino, Antonina (2022) Schools in the Media: Framing National Standardized Testing in the Norwegian Press, 2004-2018. *Discourse: Studies in the Cultural Politics of Education*, 43(4), 601–616.
- Camphuijsen, Marjolein K. & Møller, Jorunn & Skedsmo, Guri (2021) Test-Based Accountability in the Norwegian Context: Exploring Drivers, Expectations and Strategies. *Journal of Education Policy*, 36(5), 624–642.
- Camphuijsen, Marjolein K. & Parcerisa, Lluís (2023) Teachers' beliefs about standardised testing and test-based accountability: Comparing the perceptions and experiences of teachers in Chile and Norway. *European Journal of Education*, 58(1), 67–82.
- Carter, Amanda VandeHei (2019) Look How Far We Haven't Come: The Possible Implications of Current Educational Context and Practices for Young Black Males. *Education Sciences*, 9(2), 142.

- Cerdeira, J. M. & Nunes, L. C. & Reis, A. B. & Seabra, C. (2018) Predictors of student success in Higher Education: Secondary school internal scores versus national exams. *Higher Education Quarterly*, 72(4), 304–313.
- Clarke, T. (2023) Do Scores "Define" Us? Adolescents' Experiences of Wellbeing as "Welldoing" at School in England. *Review of Education*, 11(1).
- Coldwell, M. & Willis, B. (2017) Tests as boundary signifiers: Level 6 tests and the primary secondary divide. *Curriculum Journal*, 28(4), 578–597.
- Colombo, M. & McMakin, D. & Jacobs, C. & Shestok, C. (2013) Hopefulness for Teachers of ELLs in the Era of NCLB. *Multicultural Perspectives*, 15(2), 81–87.
- Cornelius, K. & Mackey-Smith, K. (2022) Improving Educational Outcomes: Why Don't Remote Schools in Australia Measure Up? *Issues in Educational Research*, 32(3), 925–942.
- Cuff, B. M. P. & Meadows, M. & Black, B. (2019) An investigation into the Sawtooth Effect in secondary school assessments in England. *Assessment in Education-Principles Policy & Practice*, 26(3), 321–339.
- Daliri-Ngametua, R. & Wescott, S. & McKay, A. (2024) On the Uses and Use of NAPLAN: The Hidden Effects of Test-Based Data-Centric Accountabilities. *Journal of Education Policy*, 39(6), 843–860.
- Dawadi, S. (2020) Parental Involvement in National EFL Test Preparation. *RELC Journal: A Journal of Language Teaching and Research*, 51(3), 427–439.
- Dawadi, S. (2021) Factors Influencing Washback of a High-Stakes English-as-a-Foreign-Language Test. *TESL-EJ*, 25(3).
- Deming, D. J. & Cohodes, S. & Jennings, J. & Jencks, C. (2016) School Accountability, Postsecondary Attainment, and Earnings. *The Review of Economics and Statistics*, 98(5), 848.
- Doecke, B. & Kostogriz, A. & Illesca, B. (2010) Seeing "Things" Differently: Recognition, Ethics, Praxis. *English Teaching: Practice and Critique*, 9(2), 81–98.
- Dowley, M. & Rice, S. (2022) Comparing Student Motivations for and Emotional Responses to National Standardised Tests and Internal School Tests: The Devil in the Detail. *Australian Journal of Education*, 66(1), 92–110.
- Dymoke, S. (2012) Opportunities or constraints? Where is the space for culturally responsive poetry teaching within high-stakes testing regimes at 16+ in Aotearoa New Zealand and England? *English Teaching-Practice and Critique*, 11(4), 19–35.
- Ehren, M. C. M. & Hatch, T. (2013) Responses of schools to accountability systems using multiple measures: The case of New York City elementary schools. *Educational Assessment, Evaluation and Accountability*, 25(4), 341–373.

- Eklof, H. & Nyroos, M. (2013) Pupil Perceptions of National Tests in Science: Perceived Importance, Invested Effort, and Test Anxiety. *European Journal of Psychology of Education*, 28(2), 497–510.
- Elo, J. & Nygren-Landgards, C. (2021) Teachers' perceptions of autonomy in the tensions between a subject focus and a cross-curricular school profile: A case study of a Finnish upper secondary school. *Journal of Educational Change*, 22(3), 423–445.
- Federicová, M. & Münich, D. (2017) The impact of high-stakes school admission exams on study achievements: Quasi-experimental evidence from Slovakia. *Journal of Population Economics*, 30(4), 1069–1092.
- Fischer, C. & Bol, L. & Pribesh, S. (2011) An Investigation of Higher-Order Thinking Skills in Smaller Learning Community Social Studies Classrooms. *American Secondary Education*, 39(2), 5–26.
- Friedrich, J. C. & Perrotta, K. A. (2022) Head above Water: A Study of K-12 Teachers' Perspectives on Emergency Remote Learning during the COVID-19 Pandemic. *Curriculum and Teaching Dialogue*, 24(1–2), 191–209.
- Fuchsman, Dillon & Sass, Tim & Zamarro, Gama (2023) Testing, Teacher Turnover, and the Distribution of Teachers across Grades and Schools. Society for Research on Educational Effectiveness. *Education Finance and Policy*, 18(4), 654-675.
- Furuta, J. (2022) The Rationalization of "Education for All": The Worldwide Rise of National Assessments, 1960-2011. *Comparative Education Review*, 66(2), 228–252.
- Gannon, S. (2019) Teaching Writing in the NAPLAN Era: The Experiences of Secondary English Teachers. *English in Australia*, 54(2), 43–56.
- Giambo, D. A. (2017) "I Will Study More. . . And Pray": Metacognition about High-Stakes Test Preparation among Culturally & Linguistically Diverse Students. *Multicultural Education*, 24(3–4), 26–34.
- Giouroukakis, V. & Honigsfeld, A. (2010) High-Stakes Testing and English Language Learners: Using Culturally and Linguistically Responsive Literacy Practices in the High School English Classroom. *TESOL Journal*, 1(4), 470–499.
- Göloğlu Demir, C. & Kaplan Keles, Ö. (2021) The Impact of High-Stakes Testing on the Teaching and Learning Processes of Mathematics. *Journal of Pedagogical Research*, 5(2), 119–137.
- Gonzalez, A. & Peters, M. L. & Orange, A. & Grigsby, B. (2017) The Influence of High-Stakes Testing on Teacher Self-Efficacy and Job-Related Stress. *Cambridge Journal of Education*, 47(4), 513–531.
- Goodacre, L. (2023) The Myth of Student Progress: Deconstructing Attainment Measures of Secondary English. *Changing English: Studies in Culture and Education*, 30(1), 3–12.
- Guill, K. & Lintorf, K. (2019) Private tutoring when stakes are high: Insights from the transition from primary to secondary school in German. *International Journal of Educational Development*, 65, 172–182.

- Gunnulfson, A. E. & Roe, A. (2018) Investigating Teachers' and School Principals' Enactments of National Testing Policies. *Journal of Educational Administration*, 56(3), 332–349.
- Hardy, I. & Lewis, S. (2018) Visibility, Invisibility, and Visualisation: The Danger of School Performance Data. *Pedagogy, Culture and Society*, 26(2), 233–248.
- Heissel, J. A. & Adam, E. K. & Doleac, J. L. & Figlio, D. N. & Meer, J. (2021) Testing, Stress, and Performance: How Students Respond Physiologically to High-Stakes Testing. *Education Finance and Policy*, 16(2), 183–208.
- Hoadley, U. & Muller, J. (2016) Visibility and Differentiation: Systemic Testing in a Developing Country Context. *Curriculum Journal*, 27(2), 272–290.
- Hofflinger, Alvaro & von Hippel, Paul T. (2020) Missing children: How Chilean schools evaded accountability by having low-performing students miss high-stakes tests. *Educational Assessment, Evaluation and Accountability*, 32(2), 127–152.
- Högberg, B. & Lindgren, J. & Johansson, K. & Strandh, M. & Petersen, S. (2021) Consequences of School Grading Systems on Adolescent Health: Evidence from a Swedish School Reform. *Journal of Education Policy*, 36(1), 84–106.
- Högberg, B. & Lindgren, J. (2023) From a Crisis of Results to a Crisis of Wellbeing—Education Reform and the Declining Sense of School Belonging in Sweden. *Comparative Education*, 59(1), 18–37.
- Holm, L. (2017) Constructions of the Literacy Competence Levels of Multilingual Students. *Language and Education*, 31(5), 449–462.
- Holmes, S. D. & Meadows, M. & Stockford, I. & He, Q. (2018) Investigating the Comparability of Examination Difficulty Using Comparative Judgement and Rasch Modelling. *International Journal of Testing*, 18(4), 366–391.
- Hong, M. & Kang, N.-H. (2010) South Korean and the US Secondary School Science Teachers' Conceptions of Creativity and Teaching for Creativity. *International Journal of Science and Mathematics Education*, 8(5), 821–843.
- Huddleston, A. P. (2014) Achievement at Whose Expense? A Literature Review of Test-Based Grade Retention Policies in U.S. Schools. *Education Policy Analysis Archives*, 22(18), 1–34.
- Huss, J. A. & Eastep, S. (2011) A Tri-State Study: Is the Middle School Movement Thriving...or Barely Surviving? *RMLE Online: Research in Middle Level Education*, 34(9), 1–13.
- Ioannidou, O. & Finch, K. & Erduran, S. (2022) Secondary Teachers' Views about Teaching and Assessing the Diversity of Scientific Methods in Practical Science. *Journal of Education for Teaching: International Research and Pedagogy*, 48(5), 592–608.
- Journell, W. (2010) The Influence of High-Stakes Testing on High School Teachers' Willingness to Incorporate Current Political Events into the Curriculum. *High School Journal*, 93(3), 111–125.

- Kiomrs, R. & Abdolmehdi, R. & Rashidi, N. (2011) On the Interaction of Test Washback and Teacher Assessment Literacy: The Case of Iranian EFL Secondary School Teachers. *English Language Teaching*, 4(1), 156–161.
- Klein, J. (2017) How Schools Cope with the Double Challenge of Excellence in High-Stakes Risk Tests and Investment in Education. *Assessment in Education: Principles, Policy & Practice*, 24(4), 474–488.
- Krause, U. & Beneker, T. & Tartwijk, J. van (2022) Higher Order Thinking by Setting and Debriefing Tasks in Dutch Geography Lessons. *European Journal of Investigation in Health Psychology and Education*, 12(1), 11–27.
- Lambert, K. & Wright, P. R. & Currie, J. & Pascoe, R. (2015) Data-Driven Performativity: Neoliberalism's Impact on Drama Education in Western Australian Secondary Schools. *Review of Education, Pedagogy & Cultural Studies*, 37(5), 460–475.
- Larsson, M. & Olin-Scheller, C. (2020) Adaptation and resistance: Washback effects of the national test on upper secondary Swedish teaching. *Curriculum Journal*, 31(4), 687–703.
- Leckie, G. & Goldstein, H. (2019) The importance of adjusting for pupil background in school value-added models: A study of Progress 8 and school accountability in England. *British Educational Research Journal*, 45(3), 518–537.
- Leite, C. & Fernandes, P. & Figueiredo, C. (2018) Challenges of Curricular Contextualisation: Teachers' Perspectives. *Australian Educational Researcher*, 45(4), 435–453.
- Leite, C. & Fernandes, P. & Figueiredo, C. (2020) National curriculum vs curricular contextualisation: Teachers' perspectives. *Educational Studies*, 46(3), 259–272.
- Leo, A. (2023) Agency in Constrained Spaces: How English Learners and their Teachers Negotiate the Power of High-stakes Exams. *TESOL Quarterly*, 57(2), 351–374.
- Lewis, S. & Hardy, I. (2015) Funding, Reputation and Targets: The Discursive Logics of High-Stakes Testing. *Cambridge Journal of Education*, 45(2), 245–264.
- Li, G. Y. & Zhu, X. & Cheong, C. M. (2020) Secondary Teachers' Conceptions of Integrated Writing Skills: Are Teachers' Conceptions Aligned with the Curriculum Objectives. *Asia Pacific Education Review*, 21(3), 379–391.
- Li, H. & Xiong, Y. (2018) The Relationship between Test Preparation and State Test Performance: Evidence from the Measure of Effective Teaching (MET) Project. *Education Policy Analysis Archives*, 26(64), 1–23.
- Lidar, M. & Karlberg, M. & Almqvist, J. & Östman, L. & Lundqvist, E. (2018) Teaching Traditions in Science Teachers' Practices and the Introduction of National Testing. *Scandinavian Journal of Educational Research*, 62(5), 754–768.

- Lilja, A. & Osbeck, C. (2020) Understanding, Acting, Verbalizing and Persevering—Swedish Teachers' Perspectives on Important Ethical Competences for Students. *Journal of Moral Education*, 49(4), 512–528.
- Lu, Q. & Liu, E. (2012) Alignment between High School Biology Curriculum Standard and the Standardised Tests of Four Provinces in China. *Journal of Biological Education*, 46(3), 149–164.
- Lundqvist, E. & Lidar, M. (2021) Functional coordination between present teaching and policy reform in Swedish science education. *Education Inquiry*, 12(2), 163–182.
- MacQuarrie, S. & Lyon, F. & Dawson, S. (2021) Implementation of Assessment Reform: Insights from Minority Language Provision. *Language and Education*, 35(4), 316–336.
- Malsbary, C. B. (2016) The Refusal: Teachers Making Policy in NYC. *International Journal of Qualitative Studies in Education (QSE)*, 29(10), 1326–1338.
- Mammadov, S. (2019) Being Gifted in Turkey: Educational and Social Experiences of High-Ability Students. *Gifted Education International*, 35(3), 216–236.
- Man-Chit Chak (2024) Washback in Language Learning Strategies under High Stakes Language Testing—A Study of the Hong Kong Secondary System. *rEFlections*, 31(1), 1–24.
- Marinho, P. & Leite, C. & Fernandes, P. (2017) Mathematics Summative Assessment Practices in Schools at Opposite Ends of Performance Rankings in Portugal. *Research in Mathematics Education*, 19(2), 184–198.
- Marks, G. N. (2015) School Sector Differences in Student Achievement in Australian Primary and Secondary Schools: A Longitudinal Analysis. *Journal of School Choice*, 9(2), 219–238.
- Marks, G. N. (2021) Should Value-Added School Effects Models Include Student- and School-Level Covariates? Evidence from Australian Population Assessment Data. *British Educational Research Journal*, 47(1), 181–204.
- Martin, S. M. (2012) Education as a Spectral Technology: Corporate Culture at Work in Ontario's Schools. *Journal of Pedagogy*, 3(2), 263–278.
- May, L. & Brenner, B. (2016) The Role of the Arts in School Reform. *Arts Education Policy Review*, 117(4), 223–229.
- Mccarthy, F. (2024) Testing times? Exploring how pupils reacted to 2020 Covid-19 GCSE and A level exam cancellation. *Teachers and Teaching*, 30(7–8), 1056–1069.
- Mccoy, S. & Byrne, D. (2024) Shadow Education Uptake in Ireland: Inequalities and Wellbeing in a High-Stakes Context. *British Journal of Educational Studies*, 72(6), 693–719.
- Meeter, M. (2023) Predicting Retention in Higher Education from high-stakes Exams or School GPA. *Educational Assessment*, 28(1), 1–10.

- Michaelides, M. P. (2014) Validity Considerations Ensuing from Examinees' Perceptions about High-Stakes National Examinations in Cyprus. *Assessment in Education: Principles, Policy & Practice*, 21(4), 427–441.
- Misco, T. & Patterson, N. & Doppen, F. (2011) Policy in the Way of Practice: How Assessment Legislation Is Affecting Social Studies Curriculum and Instruction in Ohio. *International Journal of Education Policy and Leadership*, 6(7), 1–13.
- Moore-Anderson, C. (2022) Measuring and Fostering Biological Thinking beyond Short-Answer Questions. *School Science Review*, 103(385), 21–27.
- Mora, R. (2011) "School Is So Boring": High-Stakes Testing and Boredom at an Urban Middle School. *Penn GSE Perspectives on Urban Education*, 9(1), 1–9.
- Morgan, H. (2016) Relying on High-Stakes Standardized Tests to Evaluate Schools and Teachers: A Bad Idea. *Clearing House: A Journal of Educational Strategies, Issues and Ideas*, 89(2), 67–72.
- Murchan, D. (2018) Introducing school-based assessment as part of junior cycle reform in Ireland: A bridge too far? *Educational Assessment Evaluation and Accountability*, 30(2), 97–131.
- Murphy, C. & Lundy, L. & Emerson, L. & Kerr, K. (2013) Children's perceptions of primary science assessment in England and Wales. *British Educational Research Journal*, 39(3), 585–606.
- Musoleno, R. R. & White, G. P. (2010) Influences of High-Stakes Testing on Middle School Mission and Practice. *RMLE Online: Research in Middle Level Education*, 34(3), 1–10.
- Nehring, J. H. & Szczesiul, S. (2015) Redefining high performance in Northern Ireland: Deeper learning and twenty-first century skills meet high stakes accountability. *Journal of Educational Change*, 16(3), 327–348.
- Neves, T. & Pereira, M. J. & Nata, G. (2012) One-Dimensional School Rankings: A Non-Neutral Device that Conceals and Naturalises Inequalities. *International Journal on School Disaffection*, 9(1), 7–22.
- Njiru, S. M. & Karuku, S. (2021) Effects of Performance Ranking on Students' Voice and Agency in the Mathematics Classroom. *International Online Journal of Primary Education*, 10(1), 137–147.
- Noboa, J. (2013) Teaching History on the Border: Teachers Voice their Views. *International Journal of Qualitative Studies in Education (QSE)*, 26(3), 324–345.
- Nortvedt, G. A. (2018) Policy Impact of PISA on Mathematics Education: The Case of Norway. *European Journal of Psychology of Education*, 33(3), 427–444.
- Oh, J. & Graber, K. C. (2017) National Curriculum for Physical Education in the United States. *Quest*, 69(2), 220–235.
- Omur, Y. E. & Bavli, B. (2020) High Stakes Testing and Transformation in the Occupational Identity of Teachers. *Journal of Economy Culture and Society*, 117–137.

- O'Sullivan, K.-A. & Goodwyn, A. (2020) Contested Territories: English Teachers in Australia and England Remaining Resilient and Creative in Constraining Times. *English in Education*, 54(3), 224–238.
- Park, R. J. (2022) Hot Temperature and High-Stakes Performance. *The Journal of Human Resources*, 57(2), 400.
- Puspitasari, M. (2024) Social and Cultural Factors Beyond Washback of the National Examination: Exploring the Impact of High-Stakes Testing in Indonesian Secondary Schools from Teachers' Experiences and Sociocultural Factors. *Qualitative Report*, 29(6).
- Putwain, D. & Remedios, R. (2014) The Scare Tactic: Do Fear Appeals Predict Motivation and Exam Scores? *School Psychology Quarterly*, 29(4), 503–516.
- Putwain, D. W. & Remedios, R. & Symes, W. (2015) Experiencing fear appeals as a challenge or a threat influences attainment value and academic self-efficacy. *Learning and Instruction*, 40, 21–28.
- Ramsay-Jordan, N. (2020) Preparation and the Real World of Education: How Prospective Teachers Grapple with Using Culturally Responsive Teaching Practices in the Age of Standardized Testing. *International Journal of Educational Reform*, 29(1), 3–24.
- Rapplee, J. & Komatsu, H. (2018) Stereotypes as Anglo-American Exam Ritual? Comparisons of Students' Exam Anxiety in East Asia, America, Australia, and the United Kingdom. *Oxford Review of Education*, 44(6), 730–754.
- Razan Mohammed Alqahtani & Alhamami, M. (2024) English as a Foreign Language Teaching Approaches in Saudi K-12 Education: Teacher-Centered or Student-Centered. *Journal of Education and Learning (EduLearn)*, 18(3), 817–824.
- Redman, E. (2013) Opportunities and Challenges for Integrating Sustainability Education into K-12 Schools: Case Study Phoenix, AZ. *Journal of Teacher Education for Sustainability*, 15(2), 5–24.
- Redon Pantoja, S. (2020) The Teacher's Voice: Citizenship and Neoliberalism—The Case of Chile. *Journal for Critical Education Policy Studies*, 18(2), 1–29.
- Rivas, A. & Sanchez, B. (2022) Race to the Classroom: The Governance Turn in Latin American Education. The Emerging Era of Accountability, Control and Prescribed Curriculum. *Compare: A Journal of Comparative and International Education*, 52(2), 250–268.
- Roda, A. (2017) "More [Time] Is Better or Less Is More?" Neoliberal Influences on Teaching and Learning Time. *Journal of Education Policy*, 32(3), 303–321.
- Saeki, E. & Pendergast, L. & Segool, N. K. & von der Embse, N. P. (2015) Potential Psychosocial and Instructional Consequences of the Common Core State Standards: Implications for Research and Practice. *Contemporary School Psychology*, 19(2), 89–97.
- Sanzo, K. L. & Sherman, W. H. & Clayton, J. (2011) Leadership Practices of Successful Middle School Principals. *Journal of Educational Administration*, 49(1), 31–45.

- Schoenbach, R. & Greenleaf, C. L. & Hale, G. (2010) Framework Fuels the Need to Read: Strategies Boost Literacy of Students in Content-Area Classes. *Journal of Staff Development*, 31(5), 38–42.
- Scott, E. P. (2012) Short-Term Gain at Long-Term Cost? How Resit Policy Can Affect Student Learning. *Assessment in Education: Principles, Policy & Practice*, 19(4), 431–449.
- Serrano Corkin, D. & Coleman, S. L. & Ekmekci, A. (2019) Navigating the Challenges of Student-Centered Mathematics Teaching in an Urban Context. *Urban Review: Issues and Ideas in Public Education*, 51(3), 370–403.
- Sieverding, M. & Krafft, C. & Elbadawy, A. (2019) An Exploration of the Drivers of Private Tutoring in Egypt. *Comparative Education Review*, 63(4), 562–590.
- Smyth, E. & Banks, J. (2012) High stakes testing and student perspectives on teaching and learning in the Republic of Ireland. *Educational Assessment Evaluation and Accountability*, 24(4), 283–306.
- Solanki, V. & Evans, B. R. (2020) A Comparative Study of Higher-Order Thinking and Teaching in the United States and United Kingdom. *Curriculum and Teaching*, 35(2), 45–63.
- Somers, M.-A. & Zhu, P. & Wong, E. (2011) Whether and How to Use State Tests to Measure Student Achievement in a Multi-State Randomized Experiment: An Empirical Assessment Based on Four Recent Evaluations. NCEE 2012-4015.
- Soto-Pena, M. & Gonzalez, V. (2024) Tensions using the "master's tools": A case study analysis of two middle school dual language bilingual education programs. *Bilingual Research Journal*.
- Spina, N. (2017) Governing by Numbers: Local Effects on Students' Experiences of Writing. *English in Education*, 51(1), 14–26.
- Sporre, K. (2019) Assessing Ethics Education through National Tests—An Advantage or Not? *Journal of Curriculum Studies*, 51(2), 262–278.
- Strandler, O. (2017) Constraints and Meaning-Making: Dealing with the Multifacetedness of Social Studies in Audited Teaching Practices. *Journal of Social Science Education*, 16(1), 56–67.
- Stroup, W. M. & Hills, T. & Carmona, G. (2011) Computing the Average Square: An Agent-Based Introduction to Aspects of Current Psychometric Practice. *Technology, Knowledge and Learning*, 16(3), 199–220.
- Thompson, G. & Cook, I. (2014) Manipulating the Data: Teaching and NAPLAN in the Control Society. *Discourse: Studies in the Cultural Politics of Education*, 35(1), 129–142.
- Thompson, G. L. & Allen, T. G. (2012) Four Effects of the High-Stakes Testing Movement on African American K-12 Students. *Journal of Negro Education*, 81(3), 218–227.
- Town, J. R. & Espinosa, A. M. (2015) Racing toward Algebra and Slope. *Mathematics Teaching in the Middle School*, 21(3), 169–174.

- Troncoso, P. & Pampaka, M. & Olsen, W. (2016) Beyond Traditional School Value-Added Models: A Multilevel Analysis of Complex School Effects in Chile. *School Effectiveness and School Improvement*, 27(3), 293–314.
- Tveit, Sverre (2018) Ambitious and ambiguous: Shifting purposes of national testing in the legitimization of assessment policies in Norway and Sweden (2000-2017). *Assessment in Education-Principles Policy & Practice*, 25(3), 327–350.
- van Hover, S. & Hicks, D. & Sayeski, K. (2012) A Case Study of Co-Teaching in an Inclusive Secondary High-Stakes World History I Classroom. *Theory and Research in Social Education*, 40(3), 260–291.
- Vetter, A. & Myers, J. & Reynolds, J. & Stumb, A. & Barrier, C. (2017) The Daybook Defense: How Reflection Fosters the Identity Work of Readers and Writers. *Journal of Adolescent & Adult Literacy*, 61(1), 37–44.
- Vosganoff, D. & Paatsch, L. E. & Toe, D. M. (2011) The Mathematical and Science Skills of Students Who Are Deaf or Hard of Hearing Educated in Inclusive Settings. *Deafness and Education International*, 13(2), 70–88.
- Wahleithner, J. M. (2018) Five Portraits of Teachers' Experiences Teaching Writing: Negotiating Knowledge, Student Need, and Policy. *Teachers College Record*, 120(1).
- Wang, L. C. (2016) The effect of high-stakes testing on suicidal ideation of teenagers with reference-dependent preferences. *Journal of Population Economics*, 29(2), 345–364.
- Wang, T. (2021) Urban Schooling and Social Integration of Ethnic Migrant Students in China. *Education and Urban Society*, 53(6), 708–733.
- White, P. & Anderson, J. (2012) Pressure to Perform: Reviewing the Use of Data through Professional Learning Conversations. *Mathematics Teacher Education and Development*, 14(1), 60–77.
- Wiberg, M. & Rolfsman, E. (2019) The Association between Science Achievement Measures in Schools and TIMSS Science Achievements in Sweden. *International Journal of Science Education*, 41(16), 2218–2232.
- Wiberg, M. (2019) The Relationship between TIMSS Mathematics Achievements, Grades, and National Test Scores. *Education Inquiry*, 10(4), 328–343.
- Wilkinson, H. M. & Putwain, D. W. & Mallaburn, A. (2020) How Do Teachers Communicate to Students about Forthcoming GCSE Exams? An Observational Study. *Psychology of Education Review*, 44(2), 61–67.
- Winter, C. (2017) Curriculum Policy Reform in an Era of Technical Accountability: "Fixing" Curriculum, Teachers and Students in English Schools. *Journal of Curriculum Studies*, 49(1), 55–74.

- Wright, K. L. & Hodges, T. S. & Dismuke, S. & Boedeker, P. (2020) Writing Motivation and Middle School: An Examination of Changes in Students' Motivation for Writing. *Literacy Research and Instruction*, 59(2), 148–168.
- Wuthrich, V. M. & Belcher, J. & Kilby, C. & Jagiello, T. & Lowe, C. (2021) Tracking stress, depression, and anxiety across the final year of secondary school: A longitudinal study. *Journal of School Psychology*, 88, 18–30.
- Wuthrich, V. M. & Jagiello, T. & Azzi, V. (2020) Academic Stress in the Final Years of School: A Systematic Literature Review. *Child Psychiatry & Human Development*, 51(6), 986–1015.
- Xiao, Y. & Cai, Y. & Ge, Q. & Yang, Y. (2023) The Potential of Using Formative Assessment to Enhance Academic Achievement in the Confucian-Heritage Culture: A Comparison between Hong Kong and Shanghai. *Asia-Pacific Education Researcher*, 32(6), 867–876.
- Xiao, Y. (2017) Formative Assessment in a Test-Dominated Context: How Test Practice Can Become More Productive. *Language Assessment Quarterly*, 14(4), 295–311.
- Yazicilar, Ü. & Bümen, N. T. (2019) Crossing over the Brick Wall: Adapting the Curriculum as a Way Out. *Issues in Educational Research*, 29(2), 583–609.
- Yedigöz-Kara, Z. & Bümen, N. T. (2022) The Search for an Effective Curricular Change Adoption in Foreign Language Education: A Meta-Synthesis. *International Journal of Curriculum and Instructional Studies*, 12(2), 337–366.
- Yell, M. L. & Katsiyannis, A. & Collins, J. C. & Losinski, M. (2012) Exit Exams, High-Stakes Testing, and Students with Disabilities: A Persistent Challenge. *Intervention in School and Clinic*, 48(1), 60–64.
- Yin, X. & Buck, G. A. (2015) There Is Another Choice: An Exploration of Integrating Formative Assessment in a Chinese High school Chemistry Classroom through Collaborative Action Research. *Cultural Studies of Science Education*, 10(3), 719–752.
- Yung, K. W.-H. (2021) Shadow Education as a Form of Oppression: Conceptualizing Experiences and Reflections of Secondary Students in Hong Kong. *Asia Pacific Journal of Education*, 41(1), 115–129.
- Zavacky, Francesca & Michael, Shannon L. (2017) Keeping Recess in Schools. *Journal of Physical Education, Recreation & Dance*, 88(5), 46–53.
- Zhang, Fengjuan & Liu, Yongbing (2014) A Study of Secondary School English Teachers' Beliefs in the Context of Curriculum Reform in China. *Language Teaching Research*, 18(2), 187–204.
- Zhang, Jing (2021) Parents' Involvement in High-Stakes Language Assessment: A Review of Test Impact on Parent Behavior. *English Language Teaching*, 14(12), 55–69.
- Zohar, Anat & Alboher Agmon, Vered (2018) Raising Test Scores vs. Teaching Higher Order Thinking (HOT): Senior Science Teachers' Views on How Several Concurrent Policies Affect Classroom Practices. *Research in Science & Technological Education*, 36(2), 243–260.

Özerk, Kamil & Whitehead, David (2012) The Impact of National Standards Assessment in New Zealand, and National Testing Protocols in Norway on Indigenous Schooling. *International Electronic Journal of Elementary Education*, 4(3), 545–561.



Labore eli Työn ja talouden tutkimus

LABORE (ent. Palkansaajien tutkimuslaitos) on vuonna 1971 perustettu itsenäinen tutkimuslaitos, jossa keskitytään yhteiskunnallisesti merkittävään ja tieteen kansainväliset laatukriteerit täyttävään soveltavaan taloustieteelliseen tutkimukseen. Tutkimuksen painopistealueisiin kuuluvat työn taloustiede, julkistaloustiede sekä makrotaloustiede ja toimialan taloustiede. Lisäksi teemme suhdanne-ennusteita ja toimialakatsauksia sekä julkaisemme Talous & Yhteiskunta -lehteä ja podcasteja.

Vahvuuksiamme ovat tutkijoiden korkea tieteellinen osaaminen sekä tiivis yhteistyö kotimaisten ja ulkomaisten yliopistojen ja tutkimuslaitosten kanssa. Tutkijoillamme on tärkeä asiantuntijarooli eri yhteyksissä ja he osallistuvat aktiivisesti yhteiskunnalliseen keskusteluun.

Työn ja talouden tutkimus LABORE

Arkadiankatu 7 (Economicum)

00100 Helsinki

Puh. +358 40 940 1940

labore.fi

ISBN 978-952-209-241-0 (verkkojulkaisu)

ISSN 2984-2158 (verkkojulkaisu)