

Rainer Aaltonen

# Muutoksen mittaaminen tulosevaluaation ongelmana

*Aaltonen, Rainer 1985. Muutoksen mittaaminen tulosevaluaation ongelmana. Aikuis-  
kasvatus 5, 3, 111–120. — Artikkelissa tarkastellaan muutoksen toteamisen, yleistä-  
misen ja analysoimisen ongelmia. Tutkimuksen sisäinen pätevyys on muutoksen totea-  
misen edellytys ja ulkoinen pätevyys on muutoksen yleistämisen edellytys. Näitä uh-  
kaavia tekijöitä eritellään. Muutoksen analysoiminen erotuspistemäärien avulla on nii-  
den lukuisista heikkouksista johtuen ongelmallista. Näitä heikkouksia eritellään ja  
vaihtoehtoja esitellään. Lopuksi tarkastellaan muutoksen eri lajeja ja niiden mittaami-  
seksi kehiteltyjä menetelmiä.*

Aloittakaamme kuvitteellisella esimerkillä. Erään suuren valtion viraston toiminnasta oli tullut paljon valituksia. Siellä asiointia moitittiin hitaaksi ja kankeaksi. Virkailijoita pidettiin töykeinä ja ylimielisinä, ja saadut palvelut koettiin vähämerkityksisinä. Viraston johto päätti parantaa laitoksen mainetta. Koska viraston maine suuren yleisön keskuudessa muodostuu ennen kaikkea sen vaikutelman perusteella, jonka ihmiset saavat asioidessaan tiskin takana olevan virkailijan kanssa, päätettiin järjestää heille koulutusta. Henkilöstöosaston koulutussuunnittelijat suunnittelivat kurssin ”Asiakas on aina oikeassa”, jonka tavoitteena oli lisätä virkailijoiden asiakaspalveluhenkisyttä. Kurssi koostui itseopiskelumateriaalin lisäksi kahdesta kahdenpäivän opiskelujaksosta valtion laitoksille tarkoitettussa koulutuskeskuksessa. Ennen koulutusohjelman varsinaista käyttöönottoa päätettiin sitä kokeilla, ettei turhaan koulutettaisi kaikki 2000 virkailijaa eri puolilla maata. Koulutukseen määrättiin 50 virkailijaa. Koulutuksen vaikuttavuutta päätettiin arvioida siten, että ennen ensimmäiselle kurssille tuloa heille tehtiin palveluhenkisyttä mittaava testi, jonka henkilöstöosaston tutkija oli soveltaen kääntänyt amerikkalaisesta tekstikäsikirjasta. Samanlainen testi toistettiin puolen vuoden kuluttua kun jälkimmäisen kurssin päättymisestä oli viikko kulunut. Tutkija löysi tilastollisesti merkitsevän eron (korreloivien ostosten t-testillä laskettuna) mittausten välille. Saattoiko hän nyt luottavaisesti suosittelaa viraston johdolle, että koulutusohjelma kannattaa ottaa käyttöön? — Ei.

Miksi ei? Jotta suunniteltujen toimenpiteiden muutosta aiheuttavasta vaikutuksesta voidaan varmistua ei vielä riitä että muutosta on tapahtunut vaan on varmistuttava myös siitä, että muutos on juuri po. toimenpiteiden aiheuttama. Yleisemmin sanoen on siis kysymys kausaalisuhteiden toteamisesta. Tämän edellytyksenä taas on, että muutoksen tai yhteisvaihdelun toteamisen ja oletetun syyn ja seurauksen ajallisen peräkkäisyyden lisäksi varmistutaan siitä, ettei mikään ulkopuolinen, kolmas tekijä ole aiheuttanut sitä. Esimerkkitapauksessamme sellainen olisi voinut olla vaikka samaan aikaan televisiossa lähetetty, suurta suosiota saavuttanut aikuiskasvatussarja ”...ja lähimmäisesi niinkuin itseäsi”.

Tämän artikkelin tarkoituksena onkin tarkastella niitä lukuisia muutosanalyysin ongelmia, joita käytännön tutkimustyössä tulee esiin, ja puhua huolellisesti kontrolloitujen kenttäkokeiden ja täsmällisyyteen pyrkivän mittaamisen puolesta, silloin kun tutkimuksen tavoitteena on muutoksen todentaminen ja siihen vaikuttavien syiden selvittäminen. Erityisesti tämä koskee tilanteita, joissa on kysymys tarkkaan suunniteltujen ja ennalta rajattujen toimenpiteiden kuten esim. tietyn koulutusohjelman tai menetelmän vaikutuksien arvioinnista l. tulosevaluaatiosta. Sen sijaan esim. kontekstin, panosten ja prosessien evaluaatio onnistuu mielestäni vähemmän formaalisilla lähestymistavoillakin.

# 1. Muutoksen toteamisen ongelmia

Muutoksen toteaminen on periaatteessa yksinkertainen tehtävä. Tutkimuksen alussa ja lopussa suoritetaan mittausta niissä seikoissa, joihin suunnitelluilla toimenpiteillä on tarkoitus vaikuttaa. Alku- ja loppumittauksen erotus (tai jos on useita tutkimusryhmiä, erotuksen erotus) ilmaisee toimenpiteiden vaikutuksen. Mutta näin on vain periaatteessa, edellyttäen että on käytössä täydellinen koeasetelma satunnaistettuine ryhmineen ja virheettömine mittareineen. Käytännössä on tilanne kuitenkin lähes päinvastainen. Alku- ja loppumittauksen vertailtavuutta häiritsevät lukuisat eri tekijät; vertailtavat ryhmät voivat olla jo alunperin erilaisia; mittaamistilanteisiin sisältyy häiriötekijöitä, ja mittarit ovat aina jossain määrin puutteellisia ja virheellisiä. Voiko siis muutoksen mittaamisesta tulla yhtään mitään?

Tutkimuksen logiikan l. niiden edellytysten tunteminen, jolloin syy- ja seuraussuhteita koskevia päätelmiä voidaan luotettavasti tehdä sekä näiden päätelmien luotettavuutta heikentävien tekijöiden tunteminen auttaa asianmukaisten tutkimusjärjestelyjen suunnittelussa. Sillä jos tutkimusasetelma ei tee mahdolliseksi luotettavia kausaalipäätelmiä, on aivan yhdenmukaista miten hyvin tutkimuksen muissa vaiheissa on onnistuttu — muutosta koskevat päätelmät jäävät viimekädessä arvauksen ja uskon varaan.

## Sisäinen pätevyys kausaalipäätelmien edellytyksenä

Tutkimuksen kausaalipäätelmien pätevyyttä uhkaa käytännön tutkimuksessa suuri joukko erilaisia tekijöitä, joihin tulisi tutkimusjärjestelyin voida vaikuttaa. Tarkastelemme seuraavassa näitä *tutkimuksen sisäistä pätevyyttä* (internal validity) l. sitä missä määrin yksittäisestä tutkimuksesta saadut tiedot pitävät paikkansa, vähentäviä tekijöitä Campbellin ja Stanleyn (1963) klassista esitystä hieman muunnellen (vrt. Cook & Campbell 1979; Borg & Call 1983):

1. *Historia* — tutkimusmittausten välillä voi kriteerimuuttujaan päästä vaikuttamaan muutkin kuin varsinaisesti vaikuttamaan tarkoitetut syytekijät. Jos esim. tutkimme työpaikkakoulutuksen vaikutusta työsuoritukseen on vaikea eristää koulutuksen omavaikutus jos

työpaikalla on samaan aikaan tapahtunut jokin muuta, jonka voidaan ajatella vaikuttavan työsuoritukseen. Laboratoriotutkimuksessa tätä kontrolloidaan eristämällä koehenkilöt ulkopuolisilta vaikutteilta, mutta nk. kenttäkokeissa ei näin juuri voida menetellä.

2. *Kypsyminen* — jos tutkimus kestää pitkään tapahtuu koehenkilöissä tutkimuksen kuluessa erilaisia kasvu- ja kehitysilmiöitä (myös vanhenemis- ja taantumislmiöitä), jotka mahdollisesti vaikuttavat kriteeriin ja häiritsevät koevariaabelin (toimenpiteiden) oma-vaikutuksen arviointia. Esim. lapsilla ja nuorilla fyysinen kasvu ja kypsyminen, aikuisilla vanhenemiseen liittyvät seikat.

3. *Testaus* — tutkimuksen alkumittaukset saattavat esim. harjaantumisen ja oppimisen kautta vaikuttaa loppumittauksen tulokseen.

4. *Mittaväline* — alku- ja loppumittausten välillä voi mittavälineessä (varsinkin silloin kun ihminen toimii mittavälineenä) tapahtua muutoksia, mitta-asteikko muuttuu jos havainnointi muuttuu omaa sisäistä havainnointiasteikkoaan, havainnointitilanteet voivat muuttua huomattavasti tai mittaväline on vaihdettu toiseen, jolloin vertailtavuus voi kärsiä.

5. *Tilastollinen regressio* — alkumittauksessa ääriarvoja (matalia tai korkeita) saaneet henkilöt saavat loppumittauksessa lähempänä keskiarvoa olevia pistemääriä. Ilmiö johtuu mittauksiin sisältyvästä satunnaisvirheestä l. reliabiliteetin puutteesta.

6. *Valikoituminen* — jos tutkimusryhmiä muodostettaessa joudutaan käyttämään erilaisia valintaperusteita (mikä käytännössä on yleistä) vaikuttaa se kriteerimuuttujaan; esim. vapaaehtoinen tutkimukseen osallistuminen tekee ryhmistä useimmiten erilaisia.

7. *Koekuoilleisuus* — tutkimusryhmien kokoonpano muuttuu ennen loppumittausta kun syystä tai toisesta henkilöitä poistuu niistä kesken tutkimuksen.

8. *Valikoitumisen yhdysvaikutukset* — useat edellä luetelluista seikoista voivat muodostaa valikoitumisen kanssa yhdysvaikutuksia, jotka entisestään vaikeuttavat koevariaabelin omavaikutuksen arviointia. Näitä ovat mm. valikoitumisen ja historian yhdysvaikutus, joka voi ilmetä silloin kun tutkimusryhmät ovat hyvin erilaisissa olosuhteissa kenttäkokeissa. Valikoitumisen ja kypsymisen yhdysvaikutus ilmenee silloin kun tutkimusryhmät kehittyvät (tai taantuvat) eri tahdissa. Valikoitumisen ja

mittavälineen yhdysvaikutus voi ilmetä esim. silloin kun ryhmät ovat mitta-asteikon eri kohdissa ja asteikko ei olekaan tasavälinen (ks. jäljempänä nk. katto- ja lattiavaikutus).

Borg ja Gall (1983, 637) lisäävät Campbellin ja Stanley'n luetteloon vielä kaksi sisäistä pätevyyttä uhkaavaa tekijäryhmää:

9. *Koevaikutuksen diffuusio* — vain koe-ryhmälle tarkoitettu informaatio siirtyy mitausten välillä myös vertailuryhmään. Esim. uutta opetusmenetelmää kokeiltaessa koeryhmän opettajat kertovat siitä vertailuryhmän opettajille, lainaavat ehkä uutta opetusmateriaalia heille jne., ja jotka sitten soveltavat sitä vertailuryhmän opetuksessa; tai työpaikkakoulutuksessa koeryhmälle opetettu uusi työmenetelmä ”vuotaa” sitä käytäntöön sovellettaessa kuin vertailuryhmään kuuluvat työtoverit näkevät ja omaksuvat sen omaan työhönsä.

10. *Koevaikutuksen vastustaminen* — jos tutkimuksen aikaansaamiseksi käytetään panoksia, jotka arvioidaan hyödyllisiksi ja haluttaviksi kuten esim. useissa kompensoivissa ohjelmissa on ollut (taloudellista apua, ylimääräistä opetusta tms.) tai jos panokset aiheuttavat uhkia (esim. työvauhdin kiristyminen, työpaikkojen vähentyminen) voivat tutkimukseen osallistujat (tutkimuksen toteuttajat, koe- ja vertailuryhmiin kuuluvat) toiminnallaan enemmän tai vähemmän tietoisesti estää olettamuksen mukaisen koevaikutuksen syntymisen. Esim. kompensoivissa ohjelmissa koevaikutuksen aiheuttava avustus ohjataan myös vertailuryhmään, koska katsotaan ”kyllä se heillekin kuuluu”, tai vertailuasema koetaan kilpailutilanteeksi, jossa tavallisemmin juuri vertailuryhmä pyrkii ylittämään ”tavallisen tasonsa” ja kompensoimaan tilanteen koeryhmään nähden (nk. John Henry -efekti). Kompensoivan kilpailun vastailmiönä voi olla myös ”vähemmälle jääneen” vertailuryhmän kaurainen suhtautuminen tutkimukseen, kieltäytyminen yhteistyöstä tai ”lossiksi lyöminen” niin ettei se yllä edes aikaisempaan normaaliin suoritukseen.

On lisäksi mahdollista, että edellä esiteltyt tekijät eivät suinkaan esiinny yksittäin vaan muodostavat keskenään joko kumulatiivisia tai vastakkaisia vaikutuksia ja tällä tavalla entisestään lisäävät kilpailevien selitysten mahdollisuutta.

Paras keino näiden johtopäätösten sisäistä pätevyyttä uhkaavien tekijöiden hallitsemiseksi ovat asianmukaiset tutkimusasetelmat. Tehokkaimpia näistä ovat nk. *varsinaiset kokeelliset asetelmat* (true experimental designs), joi-

den tunnuspiirteenä on satunnaistamalla muodostetut koe- ja vertailuryhmät. Koska satunnaistaminen käytännön koulutustutkimuksissa on harvoin mahdollista, voidaan hyviin tuloksiin päästä myös nk. *näennäiskokeellisilla asetelmilla* (quasi-experimental designs), joissa satunnaistaminen puuttuu ja joissa alku- ja loppumittauksissa voidaan käyttää eri henkilöitä tai joissa käytetään erilaisia aikasarjoja (ks. tarkemmin Campbell & Stanley 1963). Kaikkia tekijöitä ei edes varsinaisilla koasetelmilla voida kontrolloida, esimerkiksi kohta 10 on sellainen. Näitä vaikutuksia voidaan ehkäistä mm. sopivalla tiedottamisella. Sen sijaan esimerkiksi me mainittu ennen-jälkeen-asetelma ilman vertailuryhmää, joka Campbellin ja Stanley'n terminologian mukaan kuuluu nk. *esikokeellisiin asetelmiin* (pre-experimental designs), ei kontrolloi kohtia 6. ja 7. lukuunottamatta muita mainittuja tekijöitä. Näitä esikokeellisia asetelmia ei tule käyttää toimenpiteiden vaikutuksia arvioivissa tutkimuksissa, koska ne eivät mahdollista luotettavia kausaalipäätelmiä (ks. myös Cascio 1982, 298; Becker 1970).

## 2. Muutoksen yleistämisen ongelmia

Kuten esimerkkitapauksessamme, pyritään tutkimuksella yleensä tuloksiin, jotka olisivat laajemminkin kuin pelkkään tutkimusryhmään soveltuvia: halutaan esim. 50 henkilön otoksen perusteella tehdä 2000 henkilön yleistettävissä olevia päätelmiä. Koulutuksen järjestäjinä olemme kiinnostuneita sen lisäksi että suunniteltu muutos tapahtuu, myös siitä tapahtuuko samanlaisia muutoksia kun koulutukseen osallistuvat eri henkilöt ja vielä mahdollisesti eri organisaatioissa (Goldstein 1978).

### *Ulkoisen pätevyys yleistämisen edellytyksenä*

Tällöin on kysymys tutkimuksen *ulkoisesta pätevyydestä* (external validity) i. siitä missä määrin saadut tulokset ovat yleistettävissä tutkimuksen ulkopuolelle, erilaisiin ihmisiin, tilanteisiin, ajankohtiin jne.

Bracht ja Glass (1968) täsmentävät ulkoisen pätevyuden käsitettä jakamalla sen vielä yksilöitä koskevaan yleistettävyyteen ja tilanteita koskevaan yleistettävyyteen (population/ecological validity). Yksilöyleistämässä on niinkään kaksi tarkastelukulmaa: missä määrin tutkimuksen kohteena olleesta (yksilö) joukos-

ta voidaan tehdä yleistäviä päätelmiä laajempiin ihmisryhmiin, ja toisaalta missä määrin tutkimuksen kohteena olleiden yksilöiden henkilökohtaiset ominaisuudet, ikä, sukupuoli, koulutus, jne., vaikuttavat yhdessä koekäsittelyn (toimenpiteiden) kanssa siten, että saadaan yleistämiskelpoisia tuloksia. Tilanneyleistettävyydessä tarkastelun painopiste on muissa tutkimusjärjestelyissä 1. miten ne vaikuttavat yleistämiskelpoisten tulosten saamiseen. Brachtin ja Glassin esitystä hieman muunnellen tarkastelemme seuraavassa erityisesti tilanneyleistettävyyttä uhkaavia tekijöitä (vrt. Cook & Campbell 1979, 64—82; Borg & Call 1983, 640—643):

1. *Toimenpiteiden eksplikointi* — vaikutuksen aiheuttaneet toimenpiteet ja menettelytavat tulee voida kuvata niin tarkasti ja yksityiskohtaisesti, että ne kyetään toisissa olosuhteissa ja toisten henkilöiden toimesta toistamaan, muutoin toimenpiteistä aiheutuneilla vaikutuksillakaan ei ole yleistettävyyttä.

2. *Monikäsitteilyn häiriövaikutus* — laboratoriokeissa käytetään joskus tutkimusasetelmää, jossa kukin koehenkilö saa useita koekäsittelyjä. Jos sitten havaitaan, että viimeinen käsittely on saanut kriteerimuuttujassa aikaan suuremman muutoksen kuin muut käsittelyt, ei vielä voida kuitenkaan yleistää, että viimeinen käsittely on muita tehokkaampi, sillä tehokkuus voi riippua juuri edeltäneistä käsitteilyistä. Tästä voidaan varmistua tutkimusasetelmalla, jossa käytetään vain yhtä koekäsittelyä vuoron perään. Samanlainen häiriövaikutus voi syntyä myös kenttäkokeissa, jossa erilaisen koulutustaan ja -kokemuksen omaavat henkilöt (esim. taloudellisen, teknisen ja humanistisen) saavat samanlaisen johtamistaidollisen koulutuksen.

3. *Hawthorne-efekti* — tutkijan tutkimusryhmään kiinnittämä huomio eikä varsinainen koekäsittely saa aikaan sen, että tutkimukseen kuuluvien henkilöiden käyttäytyminen muuttuu (kutsutaan myös placebo-efektiksi). Koehenkilöt voivat yrittää arvailla mitä tutkija heiltä odottaa ja muuttaa käyttäytymistään sen mukaisesti. Joskus voi esiintyä myös erilaisia pelkoreaktioita kun koehenkilöt joutuvat esim. persoonallisuustestissä asiantuntijoiden arvioimaksi. Tutkimuksen kestäessä pidempään on kuitenkin oletettavaa, että em. efektit lakkaavat vaikuttamasta. Yleistettävyyttä ne kuitenkin heikentävät, koska ne ovat ominaisia juuri ko. tutkimukselle.

4. *Uutuusvaikutus* — koekäsittelyn vaikutus voi perustua eräänlaiseen uutuuden viehätys-

seen, jonka esim. uuden opetusmenetelmän kokeilu aiheuttaa. Käytön jatkuessa uutuus menettää arvonsa, eikä näin ollen tulos ole yleistettävissä pitempiaikasta käyttöä varten. Uutuuden viehätyskelle päinvastainen ilmiö tapahtuu silloin, kun uusi menetelmä ei aluksi näytäkään tehokkaalta, mutta käyttöä jatkettaessa osoittautuu sellaiseksi. Tällöin alustavien kokeilujen tuloksia ei voidakaan yleistää jatkuvan käytön tuloksiin.

5. *Kokeenjohtaja-efekti* — kokeenjohtajan ennakkokäsitykset koehenkilöistä voivat vaikuttaa kokeista saatuihin tuloksiin (kutsutaan myös Pygmalion- tai Rosenthal-efektiksi). Tällöin tuloksia ei voida yleistää tilanteisiin, joissa on ollut eri kokeenjohtaja.

6. *Mittaustapa* — kokeen alku- ja loppumittauksissa käytetty riippuvan muuttujan mittaustapa voi vaikuttaa tulosten yleistettävyyteen. Jos esim. mittaamme ohjelmoitua ja tavallista oppikirjaa opiskeluvälineinä monivallintatyypillisellä testillä ja saamme tulokseksi, että ohjelmoitu oppikirja on tehokkaampi, voi tulos johtua siitä, että tietty opiskelutapa suosii tiettyä mittaustapaa. Yleistämiskelpoisiin tuloksiin päästään vasta mittaustapoja vaihtelemalla.

7. *Testauskysymyksen yhdysvaikutukset* — myös itse mittaus voi toimia koevaikutuksen osana ja näin vaikuttaa tulokseen. Alkumittaus voi esim. koulutuksen ollessa kyseessä tehdä opiskelijat vastaanottavemmaksi opettavalle ainekselle, ja voidaan kysyä, mikä olisi ollut tulos ilman alkumittausta. Samalla tavoin jos loppumittauksia on ollut useita, voidaan ajatella tuloksen olevan riippuvainen niiden lukumäärästä. On myös mahdollista, että tulos olisi erilainen, esim. heikompi, jos loppumittausta ei lainkaan tehtäisi (silloin kun loppumittaus on oppimistilanne sinänsä ja lisää koekäsittelyn vaikutusta). Yleistämiskelpoisiin tuloksiin voidaan päästä lisäämällä tutkimusryhmiä em. vaihtoehtojen edellyttämällä tavalla.

8. *Testausajankohta* — loppumittaus suoritetaan tavallisesti välittömästi koekäsittelyn jälkeen. Tulokset saattavat kuitenkin muuttua huomattavasti, jos mittauksia vielä toistetaan myöhemmin. Vasta tällöin saamme tietää, mitä opitusta muistetaan tai onko opittua kyetty soveltamaan. Näin yleistämiskelpoisuus parane.

9. *Koekäsittelyn yhdysvaikutukset* — koekäsittelyn vaikutuksen yleistettävyyttä monimutkaistavat myös useat eri yhdysvaikutustekijät, valikoituminen, historia ja ympäristö. Onko saatu tulos yleistettävissä koskemaan eri sosi-

aalisia, taloudellisia, poliittisia, rodullisia, ikäisiä jne. ihmisryhmiä? Voidaanko tietynä ajankohtana saatu tulos yleistää koskemaan myös tulevaisuutta? Voiko esim. luokkahuoneessa hyväksi tutkittu menetelmä sopia tehdassaliin?

Tulosevaluatiotutkimuksessa ovat yleistettävyysongelmat ehkä vieläkin suurempia kuin sisäiset pätevyysongelmat. Teoreettisesti on kysymys induktiivisesta päättelystä, joka ei viime kädessä ole koskaan täysin varmaa, mutta se on kuitenkin ainoa tapa, millä empiiristä evidenssiä voidaan hankkia. Yleistettävyyden on pikemminkin ideaali kuin normi (Saariluoma 1984). Käytännössä yleistettävyyden varmistaminen pyritään huolehtimaan sopivien otantamenetelmien ja koeolosuhteiden sekä näitä varioivien replikaatiotutkimusten avulla.

Klassiseen laboratoriokokeeseen on kohdistettu kritiikkiä, jonka mukaan se käyttää keinotekoisia oppimistilanteita ja saa aikaan epäluonnollista ja yleistämisarvoltaan kyseenalaisia oppimiskäyttäytymistä. Koulutuksen tutkijan tulisi siirtyä oppimislaboratorioista käytännön tilanteisiin, missä oppimista tapahtuu ja sitä sovelletaan. Tutkimusjärjestely tulisi rakentaa siten, ettei keinotekoisesti rajoiteita luonnollisia oppimistilanteita ja inhimillisen subjektin aktiivista, tietoa järjestelevää ja prosessoivaa toimintaa vain reagoimiseksi kokeenjohtajan säätelemiin yksinkertaisiin ärsykeisiin (Snow 1974). Erityisesti aikuiskasvatuksen tutkimusta ajatellen kritiikkiin on helppo yhtyä, kunhan muistetaan säilyttää perinteisen kokeen logiikasta johtuvat edut myös kenttäkokeissa.

### 3. Muutoksen analysoimisen ongelmia

Palatkaamme vielä alun esimerkkiin. Sen lisäksi, että viraston tutkija oli kiinnostunut koulutuksen aiheuttamasta muutoksen määrästä, häntä kiinnosti myös millaista oli yksilöissä tapahtunut muutos: ketkä koulutettavista olivat muuttaneet käsityksiään enemmän ja ketkä vähemmän. Mielenkiintoiselta tuntui niinkään selvittää, mitkä koulutettavien taustaan, ikään, sukupuoleen, aikaisempaan koulutukseen, työkokemukseen jne., liittyvät seikat olivat yhteydessä muutokseen. Näitä tarkasteluja varten tuntui luonnolliselta laskea jokaisen koulutettavan kohdalta erikseen ensimmäisen ja toisen testauskerran erotus, ja käyt-

tää tätä erotuspistemäärää yksilöllistä muutosta kuvaavana muuttujana.

#### *Erotuspistemäärät muutoksen mittana*

Myös tätä menettelyä vastaan voidaan kohdistaa varteenotettavaa kritiikkiä. On nimittäin jossain määrin ironista, kuten Linn ja Slinde (1977) asian ilmaisevat, että tämä yksinkertainen ja usein käytetty menettely antaa tulokseksi mittaluvun, jolla on useita perusheikkouksia. Nämä heikkoudet johtuvat mittauksiin (aina) sisältyvästä virheestä, paikkaansapitämättömistä alkuoletuksista ja mitta-asteikoihin sisältyvistä käytännön ongelmista. Erittelemme näistä muutamia yleisimpiä seuraavassa (ks. O'Connor 1972; Linn & Slinde 1977; Johns 1981):

1. *Negatiivinen korrelaatio alkumittaukseen* — erotuspistemäärä ei ole riippumaton alkumittauksesta; tavallisesti korrelaatio on negatiivinen (joskus harvoin positiivinen), mikä vaikeuttaa kausaalisuhteiden tunnistamista ja vääristää todellisen muutoksen arviointia. Tämä ilmenee esim. siten, että alkumittauksessa alhaisia pistemääriä saaneet opiskelijat edistyvät erotuspistemäärien mukaan enemmän kuin korkeita alkupistemääriä saaneet. Nyt voidaan kysyä, onko opetus todella vaikuttanut siten, että heikompi tasoiset ovat hyötynyt siitä enemmän vai johtuuko tulos, ja missä määrin, erotuspistemäärien psykometrisistä ominaisuuksista (Borg & Call 1983, 720)?

2. *Erotuspistemäärien tekniset korrelaatiot muihin muuttujiin* — koska erotuspistemäärä on riippuvainen osatekijöistään, aiheutuu tästä tekninen korrelaatio erotuspistemäärän ja sen osatekijöiden kanssa korreloivien muiden muuttujien välille. Jos esim. aikaisempi koulumenestys korreloi positiivisesti alkumittaukseen, joka puolestaan korreloi negatiivisesti erotuspistemäärään, on tuloksena nolla- tai negatiivinen korrelaatio erotuspistemäärään, mikä on harhaanjohtava tulos. Tekniset korrelaatiot näyttävät siis vähentävän mahdollisuuksia löytää muutokseen yhteydessä olevia tekijöitä.

3. *Erotuspistemäärän alhainen reliabiliteetti* — alku- ja loppumittauksen välinen korrelaatio yhdessä mittausvirheen kanssa aiheuttaa sen, että erotuspistemäärän reliabiliteetti jää tavallisesti varsin alhaiseksi (ks. Linn & Slinde 1977, 123 taulukko 1.). Erotuksen reliabiliteettia voidaan parantaa lisäämällä alku- ja loppumittauksien reliabiliteettia tai pienentämällä niiden välistä korrelaatiota. Jälkimmäinen kei-

no on kuitenkin kyseenlainen, koska silloin voidaan epäillä, että eri mittauskerroilla ei olekaan mitattu samaa asiaa.

4. *Regressiovaikutus* — jo aikaisemmin mainittu regressiovaikutuksen heikentää erotuspistemääriä muutoksen mittana. Ilmiö johtuu siitä, että alkumittaukseen vaikuttavat satunnais-tekijät (esim. onnistunut arvaaminen) eivät samalla tavoin vaikuta enää loppumittauksessa (hyvä tuuri heikkenee ja huono paranee). Näin regressiovaikutus suosii alkukokeessa satunnaisesti heikosti menestyneitä. Erityisen ongelmalliseksi regressiovaikutus on osoittanut muutoksen analysoinnissa silloin, kun vertailtavat ryhmät ovat jo aluperin erilaisia, kuten erilaisten kompensoivien ohjelmien kyseessä ollen on tavallista (ks. esim. Campbell & Erlebacher 1975 ja Campbell & Boruch 1975).

5. *Mittausten vertailukelpoisuus* — jotta muutoksesta voidaan järkevästi puhua, on alku- ja loppumittauksien oltava vertailukelpoiset: on mitattava samaa asiaa samalla mittarilla (skaalalla) tai tiedettävä muunnosfunktio, jos käytetään eri mittareita. Asteikkojen on oltava tasavälisiä, jotta voidaan laskea erotuspistemäärät; esim. koulukokeissa usein käytetty vaikeutuva asteikko aiheuttaa negatiivisen korrelaation erotuspistemäärän ja alkumittauksen välille. Vertailukelpoisuuden periaatetta rikotaan käytännössä ehkä eniten.

6. *Katto- ja lattiavaikutus* — tämä ilmiö johtuu asteikon rajallisuudesta: alkumittauksessa korkeita pistemääriä saaneet eivät asteikon ylärajasta johtuen voi juuri parantaa ja alhaisia pistemääriä saaneet eivät voi huonontaa pistemääriään loppumittauksessa. Tämäkin ilmiö edistää negatiivisen korrelaation syntymistä erotuksen ja alkumittauksen välille. Mittaasteikon riittävyys tulee keskeiseksi kysymykseksi erityisesti tutkimuksissa, joissa alku- ja loppumittauksen välinen aika on pitkä ja joissa erilaiset kasvu- ja vanhenemisilmiöt vaikuttavat mitaustuloksiin.

Erotuspistemäärien edellä luetelluista heikkouksista johtuen useat tutkijat ovatkin suositelleet niiden käytöstä luopumista erityisesti yksilöitä koskevan muutoksen indikaattoreina (esim. Cronbach & Furby 1970; O'Connor 1972; Nunnally 1975; Linn & Slinde 1977; Johns 1981). Erotuspistemäärien käytölle on tosin löytynyt jatkuvasti puoltajiakin ja niitä koskeva tutkimustyö näyttää vireältä. Erityistapauksia, joissa erotuspistemäärien käyttö näyttäisi perustellulta löytyy jatkuvasti (ks. esim. Richards 1975; Labouvie 1980; Maxwell & Howard 1981; Zimmerman & Williams 1982; Glasnapp 1984; Zimmerman 1985).

## *Muut muutoksen mittaamismenetelmät*

Tavallista erotuspistemääriä parempia muutoksen mittoja on kehitelty runsaasti. Viidestä perustyyppistä on olemassa yli 50 erilaista variaatiota riippuen siitä millaisiin mittaustaso- ja teoriaoletuksiin ne perustuvat ja millaisiin taroituksiin ne parhaiten soveltuvat (Fortune & Hutson 1984). Näistä perustyypeistä nk. *ryhmävertailumallit* (esim. Campbell & Erlebacher 1975), *kasvuanalyysimallit* (esim. Bryk & Weisberg 1977) ja *rakenneyhtälömallit* (esim. Jöreskog 1979) ovat pisimmälle kehitettyjä, monimutkaisia tilastollisia analyyseja vaativia menetelmiä. Ne soveltuvat erityisesti ryhmien väliseen vertailuun, jossa ryhmien erityispiirteet, esim. lähtötaso- ja kehitysnopeudesta johtuvat erot, on otettu huomioon. Jätämme ne tässä esityksessä kuitenkin vain maininnalle ja tarkastelemme seuraavassa kahta muutoksen mittaamisen perusmallia, joissa näkyy samalla, miten erotuspistemäärien puutteita on yritetty korjata. Mallit on tarkoitettu erityisesti muutoksen korrelaatioiden etsimiseen ja eri tavoin muuttuneiden yksilöiden tunnistamiseen.

Alkumittauksesta riippumaton muutoksen mitta saadaan nk. *residuaalimuutosmalleilla*, joiden perusideana on poistaa loppumittauksesta se osa informaatiota, joka on suoraan ennustettavissa alkumittauksesta. Se mitä jää jäljelle on varsinainen muutos ja mittausrvirhe. Käytännössä tämä tapahtuu esim. osittaiskorrelaatiomenetelmän avulla vakioimalla alkumittaus tai laskemalla lineaarinen regressio loppumittaukseen alkumittaus selittävänä muuttujana, joka sitten vähennetään varsinaisesta loppumittauksen pistemäärästä. Näin saadaan erotus, joka on riippumaton alkumittauksesta, mutta sen reliabiliteetti ei ole juuri parempi tavallista erotuspistemääriä (ks. Linn & Slinde 1977, 125 taulukko 2.). Residuaalimallit välttävät myös ne ongelmat, joita syntyy kun alku- ja loppumittauksissa on käytetty erilaisia mittareita (Linn 1985, 669). (Residuaalimuutosmallien käytöstä tutkimuksessa ks. esim. Taipale 1984).

Mittausrvirheiden aiheuttamat ongelmat pyritään poistamaan nk. *todellisen muutoksen malleilla*. Klassisen testiteorian mukaan mitaustulos koostuu kahdesta toisistaan riippumattomasta osasta: todellisesta mittaluvusta ja mittausrvirheestä. Todellinen mittaluku voidaan saada vain ideaalioloissa. Tämän mukaisesti todellinen muutos on mittaluku, joka saadaan vain täysin virheettömän alku- ja loppumittauksen erotuksena. Todellista muutosta

voidaan kuitenkin estimoida esim. McNemarin (1958) ja Lordin (1963) esittämällä menetelmällä, joka ottaa huomioon alku- ja loppumittausmittausvirheestä johtuvan regressiovaikutuksen ja mittauskertojen välisen korrelaation. Gronbach ja Furby (1970) laajensivat menetelmää käyttämään hyväksi otoksesta saatavaa lisäinformaatiota (muutakin muuttujia kuin alku- ja loppumittauksista) ja ottamaan huomioon myös mahdollisen mittausvirheiden korrelaation, mikä on epäiltävissä erityisesti silloin kun käytetään samaa mittaria alku- ja loppumittauksissa. Eräänlainen residuaalimallin laajennus todellisen muutoksen mallien suuntaan on Tuckerin ym. (1966) esittämä *base-free* muutospistemäärä, joka on tavallista residuaalimuutospistemäärää parempi muutoksen mitta, sillä se ottaa huomioon mittausvirheiden reliabiliteetin. Menetelmän kehittäjät suosittelevat sitä erityisesti muutoksen korrelatiiviseen tutkimukseen. Perusluonteeltaan nämäkin menetelmät ovat lineaarisen regressioanalyysin kehitelmiä.

### *Alfa-, beta- ja gammamuutos*

Kun edellä on tarkasteltu muutoksen mittauksen ja analyysitekniikan psykometrisiä erityiskysymyksiä, katsomme vielä lopuksi mitä muutosanalyysin alueella on saatu aikaan, kun kysymysten painopistettä on siirretty muutoksen käsitteellisen erittelyn suuntaan.

Muutosta on usein pidetty yksidimensionaalisenä, määrällisenä ilmiönä. Muutoksen laadullinen tai rakenteellinen ulottuvuus on jäänyt vähemmälle huomiolle. Samoin on korostunut stabiilisuus mittareiden ominaisuutena ilmiöiden tosiasiallisen labiilisuuden kustannuksella (Nesselroade 1977). Muutoksen moniulotteisuutta lisää vielä se yhteiskunta- ja käyttäytymistieteille ominainen piirre, että muutoksen kohteena on ihminen ja tätä muutosta arvioi ihminen, usein jopa itse itseään. Kun nyt kysymme kumpi muuttuu, kohde vai mittari vai molemmat, ei vastausta ole aivan helppoa löytää.

Tähän suuntaan muutoksen käsitettä erittelmällä ovat Golembiewski, Billingsley ja Yeager kehittäneet lähestymistavan, joka soveltuu erityisesti kysely- ja haastattelumenetelmillä kerättyjen tutkimusaineistojen analyysiin ja joka on tarkoitettu erilaisten organisaation kehittämishojelmien (OD) evaluatioon. Tämä Douglas McGregor-palkinnon vuonna 1975 voittanut tutkimus on sittemmin aiheuttanut laajaa keskustelua ja runsaasti lisätutkimusta (ks. esim. Lindell ja Drexler 1979, Golembiewski & Billingsley 1980; Lindell & Drexler

1980; Armenakis ym. 1983 ja alla viitteissä oleva kirjallisuus). Sitä on pidetty yhteinä tärkeimmistä edistysasteista, joita evaluatioon metodologian alueella on viime vuosina otettu.

Golembiewski ym. (1976) jakavat muutoksen kolmeksi eri lajiksi:

*Alfamuuтокsella* tarkoitetaan muutosta, joka tapahtuu todellisuuden jollakin suhteellisen pysyvällä ulottuvuudella tai alueella, ja jota mitataan vakioasteikkoisella mittarilla. Kyseessä on muutos, joka perinteisesti on ymmärretty "todelliseksi" muutokseksi. Kirjoittajat käyttävät esimerkkinä lapsen kenkien ostoa, jossa alfamuutos l. lapsen jalan kasvu arvioidaan vertaamalla jalan kokoa kahtena eri ajankohtana tietty vakioiseen (kenkien numerointijärjestelmä) mittaan. *Betamuuтокsella* tarkoitetaan muutosta, joka tapahtuu em. alueella käytetyssä mittarissa tapahtuu asteikkomuutos. Jos betamuutos on tapahtunut, esim. aikaisempi kenkien numerointijärjestelmä on muuttunut, ei lapsen jalan kasvua voida tietää vertailemalla suoraan mittalukuja toisiinsa. *Gammamuuтокsella* tarkoitetaan muutosta, joka tapahtuu siirryttäessä todellisuuden joltakin ulottuvuudelta tai tilasta toiseen. Kirjoittajat käyttävät alfa- ja gammamuutoksen erosta esimerkkinä aineen olomuodoissa tapahtuvaa muutosta: alfamuutos on esim. veden lämpötilan kasvu ja gammamuutos veden muuttuminen nesteestä vesihöyryksi. Betamuutos esimerkiksi sovellettuna olisi esim. Celsius-lämpömittarin vaihtaminen Fahrenheit-mittariin.

OD-tutkimuksen alueella vietynä gammamuutos tarkoittaa sitä, että mitattavassa ilmiössä (käsitteessä) tapahtuu rakenteellinen muutos, mikä ilmenee merkityksen muutoksena käsitteen kielellisessä ilmaisussa, esim. sanomme että johtamistyyli muuttui OD-koulutuksen vaikutuksesta autoritaarisesta demokraattiseksi. Betamuutos merkitsee puolestaan sitä, että käsitteiden arvioinnissa käytetyissä (havainnoitsijoiden sisäisissä) asteikoissa tapahtuu muutosta, esim. alkumittauksen skaala-arvo 5 demokraattisen johtamistyylin ulottuvuudella arvioidaan 4:ksi loppumittauksessa vaikka ko. ulottuvuudella ei tosiasiaassa olisi tapahtunut tällaista muutosta. Alfamuutos on jäljelle jäänyt muutos kun mahdollinen gamma- ja betamuutos on joko vähennetty tai suljettu pois kokonaisuudesta.

Kaikki nämä muutoksen eri lajit ovat tavallisesti organisaation kehittämishojelmien tavoitteena. Alfamuutos edustaa perinteistä määrällisen muutoksen tavoitetta. Betamuutos voi edustaa esim. oikeamman ja realistisem-

man käsityksen syntymistä muutosohjelman tuloksena, kun saadan lisää tietoa arvioitavista seikoista. Gammamuutos on taas osoitus laadullisesta muutoksesta, arvioinnin kohteena oleva asia ymmärretään uudella ja ehkä syvälisemmällä tavalla.

Muutoksen mittaaminen tapahtuu siten, että ensin varmistetaan gammamuutoksesta, sillä alfa- ja betamuutosta ei voida arvioida jos gammamuutos on tapahtunut. Jos gammamuutosta ei havaita, arvioidaan seuraavaksi mahdollinen betamuutos, ja jos sitäkään ei havaita, vasta sitten voidaan luotettavasti arvioida alfamuutos (Randolph 1982; Van de Vliert ym. 1985).

Koska gammamuutoksessa on kysymys mitatun käsitteen merkityksen (rakenteen muutoksesta) voidaan muutos tunnistaa vertailemalla käsitteen rakennetta eri mittauskerroilla. Tähän soveltuu esim. jokin faktoreiden vertailuun kehitetty menetelmä. Golembiewski ym. (1976) käyttivät Almavaaran (1954) kehittämää transformaatioanalyysiä, Armenakis ym. (1977) kehittivät vertailumenetelmän itse, ja Schmitt (1982) käytti nk. konfirmatorista faktorianalyysia.

Betamuutoksen mittaamiseksi on niinkään kehitynyt useita erilaisia menetelmiä. Zmud ja Armenakis (1978; ks. myös Armenakis & Zmud 1979) käyttivät menetelyä, jossa arvioitsijoina toimivat henkilöt esittivät kullakin mittauskerrolla arvioitavasta ominaisuudesta tms. seikasta sekä senhetkisen, aktuaalisen käsityksensä, että ideaalisen käsityksensä, so. millainen ko. ominaisuuden heidän mielestään pitäisi olla. Vertaamalla aktuaalisen, ideaalisen ja erotuspistemäärien erotuksia saadaan alfa- ja betamuutos selville. Tässä menetelmässä tulevat kuitenkin jo edellä käsitellyt erotuspistemäärien ongelmat mukaan. Bedeian ym. (1980) käyttivät myös aktuaalista ja ideaalista arviota kuvaavia pistemääriä, mutta siten, että arvioitavien ominaisuuksien ideaalipistemääristä lasketaan kullekin arvioitsijalle betamuutoksen ilmaiseva regressioyhtälö, jolla korjataan arvioitsijan antama aktuaalinen arvio (ks. menetelmän kriitikkistä Terborg ym. 1982). Terborg ym. (1980) käyttivät menetelyä, jossa loppumittauksen yhteydessä arvioitsijoilta pyydetään vielä uudelleen jokaisen arvioitavan osion osalta arviota siitä millaisia he *nyt* arvioivat ko. osioiden olleen juuri ennen kuin muutoksen tähtäävät toimenpiteet, esim. koulutus, alkoivat. Tätä he kutsuivat *silloin* (then) mittaukseksi, ja sen tehtävänä on toimia retrospektiivisenä alkumittauksena. Betamuutos saadaan selville kun lasketaan varsinaisen al-

kumittauksen ja silloin-mittauksen keskiarvojen erotu, ja alfamuutos saadaan laskemalla silloin-mittauksen ja loppumittauksen keskiarvojen erotus (ks. menetelmän kriitikkistä Armenakis & Bedeian 1982)

Muutoksen erittely em. tapaan näyttää tarjoavan mielenkiintoisen mahdollisuuden muutoksen analysoimiseksi. Muutoksen mittaamisessa tavallisesti psykometrinä, mittareiden puutteellisista ominaisuuksista johtuvina ongelmina pidetyt seikat saavat ainakin osittain selityksensä myös muutoksen moniuloitteisesta luonteesta käsin.

#### 4. Lopuksi

Viime vuosina on sekä muualla että meillä kirjoitettu runsaasti erilaisista uusista evaluatiomenetelmistä ja lähetymistavoista. Perinteinen tyleriläinen paradigma on saanut rinnalleen ja haastajikseen lukuisia nk. kvalitatiivisen evaluatiion malleja. Niin paljon myönteistä kehitystä kuin ne ovatkin saaneet aikaan korostaessaan evaluatiion kokonaisvaltaista luonnetta ja vaatimusta kasvatus- ja koulutusprosessin aidosta ja autenttisesta kuvaamisesta, eivät ne kuitenkaan ole ainakaan tämän kirjoittajan mielestä kyenneet panemaan viralta perinteisen näkemyksen mukaista tutkimuksen logiikkaa, mihin olennaisena osana kuuluu juuri virheitä ja epävarmuutta aiheuttavien tekijöiden mahdollisimman huolellinen kontrollointi — siitäkkin huolimatta, että näitä tekijöitä näyttää ihmistä ja yhteiskuntaa tutkivien tieteiden alueella olevan huomattavasti runsaammin kuin olisi tarpeellista.

Pätevään tietoon ei ole olemassa mitään kunninkaan tietä. Epävarmuus- ja virhetekijöiden tunteminen auttaa tutkijaa suunnittelemaan mahdollisimman hyvät tutkimusjärjestelyt myös kenttäkokeissa ja arvioimaan niiden osuutta lopullisissa tutkimustuloksissa.

#### Kirjallisuus

- Armenakis, A.A. & Bedeian, A.G.: On the Measurement and Control of Beta Change: Reply to Therborg, Maxwell, and Howard. *Academy of Management Review*, 1982 vol. 7 no. 2, 296—299.
- Armenakis, A.A., Bedeian, A.G. & Pond, S.B., III.: Research Issues in OD Evaluation: Past, Present, and Future. *Academy of Management Review*, 1983 vol. 8 no. 2, 320—328.
- Armenakis, A.A. & Zmud, R.W.: Interpreting the Measurement of Change in Organizational Research. *Personnel Psychology*, 1979 vol. 32, 709—723.
- Armenakis, A.A., Feild, H.S. & Wilmoth, J.N.: An Algorithm for Assessing Factor Structure Cong-

- ruence. *Educational and Psychological Measurement*, 1977 vol. 37 no. 2, 213—214.
- Becker, S.W.: *The Parable of the Pill*. *Administrative Science Quarterly*, 1970 vol. 15 no. 1, 94—96.
- Bedeian, A.G., Armenakis, A.A. & Gibson, R.W.: *The Measurement and Control of Beta Change*. *Academy of Management Review*, 1980 vol. 5 no. 4, 561—566.
- Borg, W.R. & Gall, M.D.: *Educational Research. An Introduction*. Fourth Edition. Longman: New York, 1983.
- Bracht, G.H. & Glass, G.V.: *The External Validity of Comparative Experiments in Educational and Social Sciences*. *American Educational Research Journal*, 1968 vol. 5, 437—474.
- Bryk, A.S. & Weisberg, H.I.: *Use of the Nonequivalent Control Group Design When Subjects Are Growing*. *Psychological Bulletin*, 1977 vol. 84 no. 5, 950—962.
- Campbell, D.T. & Boruch, R.: *Making the Case for Randomized Assignment to Treatments by Considering the Alternatives: Six Ways in Which Quasi-Experimental Evaluations in Compensatory Education Tend to Underestimate Effects*. Teoksessa Bennet, C.A. & Lumsdaine, A.A. (Eds.): *Evaluation and Experiment. Some Critical Issues in Assessing Social Programs*. Academic Press, Inc.: New York, 1975.
- Campbell, D.T. & Erlebacher, A.: *How Regression Artifacts in Quasi-Experimental Evaluations Can Mistakenly Make Compensatory Education Look Harmful*. Teoksessa: Struening, E.L. & Guttentag, M. (Eds.): *Handbook of Evaluation Research*, vol. 1. Sage Publications: Beverly Hills, 1975.
- Campbell, D.T. & Stanley, J.C.: *Experimental and Quasi-Experimental Designs for Research on Teaching*. Teoksessa: Gage, N.L. (Ed.): *Handbook of Research on Teaching*. Rand McNally: Chicago, 1963.
- Cascio, W.F.: *Applied Psychology in Personnel Management*. Second Edition. Reston Publishing Company, Inc.: Reston, Virginia, 1982.
- Cook, T.D. & Campbell, D.T.: *Quasi-Experimentation. Design & Analysis Issues for Field Settings*. Rand McNally: Chicago, 1979.
- Cronbach, L.J. & Furby, L.: *How We Should Measure "Change" — Or Should We?* *Psychological Bulletin*, 1970 vol. 74 no. 1, 68—80.
- Fortune, J.C. & Hutson, B.A.: *Selecting Models for Measuring Change When True Experimental Conditions Do not Exist*. *Journal of Educational Research*, 1984 vol. 77 no. 4, 197—206.
- Glassnapp, D.R.: *Change Scores and Regression Suppressor Conditions*. *Educational and Psychological Measurement*, 1984 vol. 44, 851—867.
- Goldstein, I.L.: *The Pursuit of Validity in the Evaluation of Training Programs*. *Human Factors*, 1978 vol. 20 no. 2, 131—144.
- Golembiewski, R.T. & Billingsley, K.R.: *Measuring Change in OD Panel Designs: A Response to Critics*. *Academy of Management Review*, 1980 vol. 5 no. 1, 97—103.
- Golembiewski, R.T., Billingsley, K. & Yeager, S.: *Measuring Change and Persistence in Human Affairs: Types of Change Generated by OD Designs*. *The Journal of Applied Behavioral Science*, 1976 vol. 12 no. 2, 133—157.
- Johns, G.: *Difference Score Measures of Organizational Behavior Variables: A Critique*. *Organizational Behavior and Human Performance*, 1981 vol. 27, 443—463.
- Jöreskog, K.G.: *Statistical Estimation of Structural Models in Longitudinal-Developmental Investigations*. Teoksessa: Nesselroade, J.R. & Baltes, P.B. (Eds.): *Longitudinal Research in the Study of Behavior and Development*. Academic Press: New York, 1979.
- Labouvie, E.W.: *Measurement of Individual Differences in Intraindividual Changes*. *Psychological Bulletin*, 1980 vol. 88 no. 1, 54—59.
- Lindell, M.K. & Drexler, J.A., Jr.: *Equivocality of Factor Incongruence As an Indicator of Type of Change in OD Interventions*. *Academy of Management Review*, 1980 vol. 5 no. 1, 105—107.
- Lindell, M.K. & Drexler, J.A., Jr.: *Issues in Using Survey Methods For Measuring Organizational Change*. *Academy of Management Review*, 1979 vol. 4 no. 1, 13—19.
- Linn, R.L.: *Change Assessment in Students*. Teoksessa: Husén, T. & Postlethwaite, T.N. (Eds.): *The International Encyclopedia of Education Research and Studies*, vol. 2. Pergamon Press: Oxford, 1985.
- Linn, R.L. & Slinde, J.A.: *The Determination of the Significance of Change Between Pre- and Post-testing Periods*. *Review of Educational Research*, 1977 vol. 47 no. 1, 121—150.
- Lord, F.M.: *Elementary Models for Measuring Change*. Teoksessa: Harris, C.W. (Ed.): *Problems in Measuring Change*. University of Wisconsin Press: Madison, 1963.
- Maxwell, S.E. & Howard, G.S.: *Change Scores — Necessarily Anathema?* *Educational and Psychological Measurement*, 1981 vol. 41, 747—756.
- NcNemar, Q.: *On Growth Measurement*. *Educational and Psychological Measurement*, 1958 vol. 18, 47—55.
- Nesselroade, J.R.: *Issues in Studying Developmental Change in Adults from A Multivariate Perspective*. Teoksessa: Birren, J.E. & Schaie, K.W. (Eds.): *Handbook of the Psychology of Aging*. Van Nostrand Reinhold Company: New York, 1977.
- Nunnally, J.C.: *The Study of Change in Evaluation Research: Principles Concerning Measurement, Experimental Design, and Analysis*. Teoksessa: Struening, E.L. & Guttentag, M. (Eds.) *Handbook of Evaluation Research*, vol. 1. Sage Publications: Beverly Hills, 1975.
- O'Connor, E.F.: *Extending Classical Test Theory to The Measurement of Change*. *Review of Educational Research*, 1972 vol. 42 no. 1, 73—97.
- Randolph, W.A.: *Planned Organizational Change and Its Measurement*. *Personnel Psychology*,

- 1982 vol. 35, 117—139.
- Richards, J.M., Jr.: A Simulation Study of the Use of Change Measures to Compare Educational Programs. *American Educational Research Journal*, 1975 vol. 12 no. 3, 299—311.
- Saariluoma, P.: Ekologinen validiteetti — normi vai ideaali? *Psykologia*, 2/1984, 83—88.
- Schmitt, N.: The Use of Analysis of Covariance Structures to Assess Beta and Gamma Change. *Multivariate Behavioral Research*, 1982 vol. 17, 343—358.
- Snow, R.: Representative and Quasi-Representative Designs for Research on Teaching. *Review of Educational Research*, 1974 vol. 44 no. 3, 265—291.
- Taipale, E.: Kasvatustuostosten muutosten mittamisesta ja analysoinnista, osat I—V. Jyväskylän yliopiston kasvatustieteen laitoksen julkaisuja A 4 a ja b/1984. Jyväskylä 1984.
- Terborg, J.R., Howard, G.S. & Maxwell, S.E.: Evaluating Planned Organizational Change: A Method for Assessing Alpha, Beta, and Gamma Change. *Academy of Management Review*, 1980 vol. 5 no. 1, 109—121.
- Terborg, J.R., Maxwell, S.E. & Howard, G.S.: On the Measurement and Control of Beta Change: Problems with the Bedeian, Armenakis, and Gibson Technique. *Academy of Management Review*, 1982 vol. 7 no. 2, 292—295.
- Tucker, L.R., Damarin, F. & Messick, S.: A Base-free Measurement of Change. *Psychometrika*, 1966 vol. 73, 457—473.
- Van de Vliert, E., Huisman, S.E. & Stok, J.J.L.: The Criterion Approach to Unraveling Beta and Alpha Change. *Academy of Management Review*, 1985 vol. 10 no. 2, 269—275.
- Zimmerman, D.W. & Williams, R.H.: On the High Predictive Potential of Change and Growth Measures. *Educational and Psychological Measurement*, 1982 vol. 42, 961—968.
- Zimmerman, D.W., Andrews, D.A., Robinson, D. & Williams, R.H.: A Note on Non-parallelism of Pretest and Posttest Measures in Assessing Change. *The Journal of Experimental Education*, 1985 vol. 53 no. 4, 234—236.
- Zmud, R.W. & Armenakis, A.A.: Understanding the Measurement of Change. *Academy of Management Review*, 1978 vol. 3 no. 3, 661—669.