

Reliabiliteettitutkimus: Havainnoijien välinen yhteneväisyys observoinnissa

Helsingin yliopisto
Kasvatustieteellinen tiedekunta
Opettajankoulutuslaitos
Varhaiskasvatus
Pro gradu -tutkimus
Huhtikuu 2018
Tuomas Sarkkinen

Ohjaaja: Nina Sajaniemi

Tiedekunta – Fakultet – Faculty Kasvatustieteellinen tiedekunta	Laitos – Institution – Department Opettajankoulutuslaitos
Tekijä – Författare – Author Tuomas Sarkkinen	
Työn nimi – Arbetets titel – Title Reliabiliteettitutkimus: Havainnoijien välinen yhteneväisyys observoinnissa	
Oppiaine – Läroämne – Subject Varhaiskasvatus	
Työn ohjaaja(t) – Arbetets handledare – Supervisor Nina Sajaniemi	Vuosi – År – Year 2018
Tiivistelmä – Abstrakt – Abstract <p><i>Tavoitteet.</i> Kehittävä palaute varhaiskasvatuksessa on 12 kunnan varhaiskasvatuksen kehittämishanke, jossa aineistoa kerätään lasten observoinnin sekä kasvattajien ja esimiehen itsearvioinnin kautta. Tämän tutkimuksen tavoitteena oli tarkastella hankkeen observointien yhteneväisyyttä eli samasta tilanteesta tehtyjen merkintöjen samankaltaisuutta. Havainnoijien välinen yhteneväisyys on yleinen tapa arvioida havaintoon pohjautuvan aineistonkeruumenetelmän reliabiliteettia ja validiteettia.</p> <p><i>Menetelmät.</i> Tämän tutkimuksen aineistona käytettiin hankkeen kautta kerättyä pariobservointidataa. Aineisto koostui 9 parin observoinneista, joita oli yhteensä 411 kappaletta. Aineisto koodattiin yhteneväisyysanalyysia varten kvantitatiiviseksi aineistoksi, ja sen analysointiin käytettiin kvantitatiivisia menetelmiä. Yhteneväisyys selvitettiin laskemalla jokaiselle observointimittarille ja observoijaparille Cohenin kappan -kerroin. Cohenin kappan laskee havainnoinnin yhteneväisyyden asteen ja ottaa sattuman vaikutuksen huomioon yhtälössä. Observointimerkintöjen jakautumista mittareiden kategorioihin tutkittiin ristiintaulukoinnin kautta. Ristiintaulukoinnin avulla oli mahdollista tarkastella, miten yhteneväisesti ja tasapainoisesti observoijat käyttävät mittareiden kategorioita.</p> <p><i>Tulokset ja johtopäätökset.</i> Landisin ja Kochin (1977) määrittelyn perusteella yhteneväisyys oli yhdellä mittarilla lähes täydellinen, kahdella huomattava, kolmella keskinkertainen ja kolmella kohtalainen. Merkintöjen eriäväisyys oli joillakin mittareilla suurta, ja mittareiden sekä parien yhteneväisyydessä oli selviä eroja. Tulosten ja lähdekirjallisuuden perusteella observointimenetelmän luotettavuutta tulisi pyrkiä jatkossa kehittämään. Kehittämisessä tulisi painottaa koulutuksen lisäämistä, observoinnin sisältöjen arviointia sekä mittareiden kalibroimista. Cohenin kappan käyttö tässä tutkimuksessa ei ollut ongelmatonta, ja tulevissa tutkimuksissa tulisi painottaa usean eri menetelmän käyttöä tasapainoisen tuloksen takaamiseksi. Tutkimuksen loppuosassa on esitetty erilaisia tapoja laajentaa hankkeen observointimenetelmän yhteneväisyyden tutkimusta, ja pohdittu yleistettävyysteorian sekä argumenttipohjaisen validiteettimallin mahdollisuuksia havainnointipohjaisten arviointimenetelmien reliabiliteetin ja validiteetin tutkimuksessa.</p>	
Avainsanat – Nyckelord – Keywords Varhaiskasvatus, havainnointi, observointi, havainnoijien välinen yhteneväisyys, reliabiliteetti, kappan	
Säilytyspaikka – Förvaringsställe – Where deposited Helsingin yliopiston kirjasto – Helda / E-thesis (opinnäytteet) ethesis.helsinki.fi	



Tiedekunta – Fakultet – Faculty Faculty of Educational Sciences	Laitos – Institution – Department Teacher Education
Tekijä – Författare – Author Tuomas Sarkkinen	
Työn nimi – Arbetets titel – Title A reliability study: Inter-rater reliability in structured observation	
Oppiaine – Läroämne – Subject Early Childhood Education	
Työn ohjaaja(t) – Arbetets handledare – Supervisor Nina Sajaniemi	Vuosi – År – Year 2018
Tiivistelmä – Abstrakt – Abstract <p><i>Objectives.</i> <i>Progressive Feedback in Early Childhood Education</i> is a project for development of early childhood education in 12 Finnish municipalities. The data in the project is collected through child observation and staff and manager self-evaluation. The objective of this study was to assess the inter-rater agreement of the project's observation method. Inter-rater agreement is a common way of assessing the reliability and validity of a method that is based on observational data collection.</p> <p><i>Methods.</i> This study used paired observation data that was collected through the project. The data consisted of observations conducted by 9 different pairs of observers, generating 411 paired observations in total. The data was coded as quantitative data and quantitative methods were used for the reliability analysis. Cohen's kappa was used to calculate the level of agreement for the rating system scales and for all the paired observers using the rating system. Cohen's kappa calculates the level of agreement among raters and accounts for agreement based on chance. Cross-tabulation was used to examine the distribution of agreement and disagreement of the observational data on the rating scales.</p> <p><i>Results.</i> Based on labels assigned by Landis and Koch (1977) the level of agreement for one rating scale was almost perfect, for two rating scales substantial, for three rating scales moderate and for three rating scales fair. The disagreement was notable on certain scales and there were clear differences between scales and raters. Based on the results and sources the reliability of the observational method should be improved in the future. The improvement should focus on increasing the training of the observers, evaluating the objectives of the observation and calibrating the observation rating scales. Using Cohen's kappa was not problem-free, so in future research multiple methods should be used to ensure balanced results. Further studies include ways to follow up on the reliability of the project's observation method and the possibilities of generalizability theory and argument based validity as part of the research on reliability and validity of observation based quality assessment methods.</p>	
Avainsanat – Nyckelord – Keywords Early childhood education, observation, inter-rater agreement, reliability, kappa	
Säilytyspaikka – Förvaringsställe – Where deposited Helsingin yliopiston kirjasto – Helda / E-thesis (opinnäytteet) ethesis.helsinki.fi	

Sisällysluettelo

1	JOHDANTO.....	1
2	KEHITTÄVÄ PALAUTE VARHAISKASVATUKSESSA -HANKE.....	4
	2.1 Hankkeen historia ja kasvatustieteellinen teoriatausta.....	4
	2.2 Observointi käytännössä.....	8
	2.3 Observointimittarit	9
3	RELIABILITEETTI JA VALIDITEETTI	20
	3.1 Reliabiliteetti yleisesti	20
	3.2 Reliabiliteetti kvantitatiivisessa tutkimuksessa	21
	3.3 Cohenin kappaa	23
	3.4 Validiteetti	26
4	AIKAISEMPI TUTKIMUS	27
5	TUTKIMUSKYSYMYKSET JA TUTKIMUKSEN TOTEUTUS	31
	5.1 Tutkimuksen ote	32
	5.2 Tutkimuksen kohderyhmä	33
	5.3 Tutkimusaineiston keruu ja analysointimenetelmät	35
6	TULOKSET	37
	6.1 Observoinnin yhteneväisyys.....	37
	6.1.1 Parien kappaa kaikilla mittareilla	39
	6.1.2 Parien kappaa keskiarvo	42
	6.1.3 Mittarien kappaa keskiarvo	43
	6.2 Merkintöjen jakautuminen.....	45
	6.2.1 A. Yleinen toiminta	47
	6.2.2 B. Lapsen oma toiminta.....	49
	6.2.3 C. Lapsen pääasiallinen huomion kohde	52
	6.2.4 D. Lähin sosiaalinen lapsikontakti.....	53
	6.2.5 E. Lapsen fyysinen aktiivisuus	55
	6.2.6 F. Lapsen sitoutuneisuus.....	56
	6.2.7 G. Lapsen emootio.....	58
	6.2.8 H. Sosiaalinen orientaatio.....	59
	6.2.9 I. Lähimmän aikuisen toiminta.....	61
7	TULOSTEN POHDINTA	62
	7.1 Kappaa-tulosten pohdinta.....	62

7.2	Ristiintaulukoinnin pohdinta	65
7.2.1	A. Yleinen toiminta	65
7.2.2	B. Lapsen oma toiminta	65
7.2.3	C. Lapsen pääasiallinen huomion kohde	66
7.2.4	D. Lähin sosiaalinen lapsikontakti	67
7.2.5	E. Lapsen fyysinen aktiivisuus	68
7.2.6	F. Lapsen sitoutuneisuus	68
7.2.7	G. Lapsen emootio	70
7.2.8	H. Sosiaalinen orientaatio	73
7.2.9	I. Lähimmän aikuisen toiminta	74
7.2.10	Ristiintaulukoinnin pohdinnan yhteenveto	75
8	KAPPAN KÄYTTÖÖN LIITTYVÄT ONGELMAT	76
8.1	Merkintöjen esiintyvyys ja reunafrekvenssit	76
8.2	Kappan eri versioiden vaikutus mataliin tuloksiin	79
8.3	Otoskoon vaikutus kappaan	80
9	JOHTOPÄÄTÖKSET	82
10	EHDOTUKSIA HAVAINNOINTIPOHJAISTEN MENETELMIEN LUOTETTAVUUDEN TUTKIMUKSEEN	85
10.1	Hankkeen observoinnin laajempi tutkiminen	85
10.2	Yleistettävyysteoria	87
10.3	Argumenttipohjainen validiteetti	90
10.4	Yhteenveto	94
	LÄHTEET	96
	LIITTEET	106
	KUVIOT	
	Kuvio 1. Sulauttamisen ja mukauttamisen käsitteet	5
	Kuvio 2. Sopeutumisen käsite	6
	Kuvio 3. Vaikuttamisen käsite	7
	Kuvio 4. A. Yleinen tilanne	10
	Kuvio 5. B. Lapsen oma toiminta	11
	Kuvio 6. C. Lapsen pääasiallinen huomionkohde	12
	Kuvio 7. D. Lähin sosiaalinen lapsikontakti	12
	Kuvio 8. E. Lapsen fyysinen aktiivisuus	13
	Kuvio 9. F. Lapsen sitoutuneisuus	13

Kuvio 10. G. Emootio	14
Kuvio 11. H. Sosiaalinen orientaatio	15
Kuvio 12. Hahmotuksen agenssin nelikenttä muokattuna	16
Kuvio 13. Lapsen hahmotuksen agenssin nelikenttä	17
Kuvio 14. I. Lähimmän aikuisen toiminta	18
Kuvio 15. Klassinen testiteoria, varianssin määrä ja reliabiliteetti	22
Kuvio 16. Mittausvirheen osat	89
Kuvio 17. Argumenttipohjainen validiteetti tutkimuksessa	92

TAULUKOT

Taulukko 1. Kappa ja yhteneväisyyden aste	37
Taulukko 2. Kappa pareittain kaikilla mittareilla	40
Taulukko 3. Parien kappan keskiarvo	42
Taulukko 4. Mittarien kappan keskiarvo	43
Taulukko 5. Esimerkki observointien ristiintaulukoinnista	46
Taulukko 6. Ristiintaulukointi A. Yleinen toiminta	48
Taulukko 7. Ristiintaulukointi B. Lapsen oma toiminta	50
Taulukko 8. Ristiintaulukointi C. Lapsen pääasiallinen huomion kohde	52
Taulukko 9. Ristiintaulukointi D. Lähin sosiaalinen lapsikontakti	54
Taulukko 10. Ristiintaulukointi E. Lapsen fyysinen aktiivisuus	55
Taulukko 11. Ristiintaulukointi F. Lapsen sitoutuneisuus	57
Taulukko 12. Ristiintaulukointi G. Lapsen emootio	58
Taulukko 13. Ristiintaulukointi H. Sosiaalinen orientaatio	60
Taulukko 14. Ristiintaulukointi I. Lähimmän aikuisen toiminta	61
Taulukko 15. Luotettavan datan määrä aineistossa kappan perusteella	64
Taulukko 16. Reunafrekvenssien vaikutus kappan kokoon	78
Taulukko 17. Kappan eri versioiden vaikutus alhaiseen kertoimeen	79

KAAVAT

Kaava 1. Cohenin kappa -kertoimen kaava	24
Kaava 2. Cohenin kappa -kertoimen kaava yksinkertaistettuna	24

KUVAT

Kuva 1. FACS-koodausasteikon osa	71
--	----

1 Johdanto

Kehittävä palaute varhaiskasvatuksessa -hanke on Helsingin yliopiston ja 12 kunnan yhteistyössä toteuttama tutkimukseen perustuva varhaiskasvatuksen kehittämisen projekti, jonka vastuullinen tutkija on dosentti Jyrki Reunamo. Hanke on jatkoa vuonna 2009 alkaneelle *Orientaatioprojekti*-hankkeelle. *Kehittävä palaute varhaiskasvatuksessa* -hankkeen tavoitteena on selvittää, miten lasten näkemykset, taidot ja kaverisuhteet muotoutuvat sekä miten aikuisen toiminta ja oppimisympäristö ovat yhteydessä lasten toimintaan. Hankkeen tavoitteina on luoda toimivat mittarit varhaiskasvatustoiminnan syvälliseen arviointiin sekä tutkimustulosten perusteella laadittujen kehitystehtävien avulla luoda malleja varhaiskasvatustoiminnan edistämiseksi. Hankkeen aineistonkeruumenetelminä ovat varhaiskasvatustiimien suorittama oppimisympäristön arviointi, päiväkodinjohtajien suorittama arviointi sekä lapsihavainnointi eli observointi. (Reunamo, 2018.)

Havainnoinnista puhutaan hankkeen teksteissä observointina ja havainnoijista observoijina, joten johdonmukaisuuden vuoksi myös tässä tutkimuksessa käytetään jatkossa havainnointi-sanasta sanaa observointi. Johnson ja Christensen (2004) määrittelevät observoinnin toimintana, jossa havainnoidaan toisten ihmisten käyttäytymismalleja, ja jossa tavoitteena on hankkia tietoa valitusta mielenkiinnonkohteesta. Observointi voidaan määritellä tässä tapauksessa vielä tarkemmin kvantitatiiviseksi strukturoiduksi observoinniksi, sillä aineistonkeräykseen liittyvät asiat kuten havainnoinnin kohde, havainnoitavat asiat, havainnointiaika sekä havainnoinnin toteuttaminen ja kirjaaminen ovat standardisoitu systemaattisen datan keräämisen mahdollistamiseksi. (Johnson & Christensen, 2004, s. 186–187.)

Tämä tutkimus keskittyy hankkeen observoinnissa käytettäviin mittareihin ja niiden merkintöihin. Hankkeen observointia toteutettiin 12 kunnassa syksyllä 2017 ja keväällä 2018. Olin itse syksyllä 2017 observoija neljässä helsinkiläisessä päiväkodissa. Keväällä 2017 osallistuin koulutuksiin, joissa tutustuttiin hankkeeseen ja observoinnin sisältöihin. Näissä tilanteissa nousi ensimmäistä kertaa esille havaintojen tekemisen tarkkuus; joidenkin observoijien mielestä tietyt mittarit olivat vaikeampia havainnoida yhteneväisesti kuin toiset. Omat huomioni olivat samansuuntaisia, ja erityisesti lapsen sitoutuneisuuden ja

sosiaalisen orientaation kohdalla selkeän havainnon määrittely oli vaikeaa. Samoihin aikoihin pohdin, kuinka hyödyllistä olisi, jos hankkeessa tehtyjen havaintojen luotettavuutta voisi konkreettisesti tarkastella ja arvioida. Kevään aikana pidetyt koulutuspäivät sekä observoinnin harjoittelu ja siihen liittyvät ajatukset olivat taas ajankohtaisia kesällä 2017, kun sain tietää päässeeni varhaiskasvatuksen maisteriopintoihin Helsingin yliopistoon. Tutkimusaihetta pohtiessani ajankohtainen hanke, mielenkiintoinen aihe sekä selkeä tavoite tekivät tutkimusaiheen valinnan lopulta helpoksi.

Tämän tutkimuksen aiheena on observointien yhteneväisyys eli se, kuinka samalla tavalla observoitavat tilanteet merkitään. Tutkimuksen tieteellinen ja yhteiskunnallinen merkitys perustuvat observoinnin luotettavuuden tarkasteluun. Tieteellisesti tarkasteltuna tämän tutkimuksen tulokset antavat näkökulman observointien luotettavuuteen ja sen kehittämiseen. Tutkimuksen yhteiskunnallinen merkitys liittyy hankkeen mittakaavaan: *Kehittävä palaute varhaiskasvatuksessa* on mittava hanke, johon osallistuu kaksitoista Suomen kuntaa. Kansainvälisesti hankkeessa ovat mukana Taiwan, Singapore, Hongkong ja Turkki (Reunamo, 2018). Hankeen rooli varhaiskasvatuksen kehittämisessä voi siis olla tulevana vuosina merkittävä; on tärkeää, että hanketta ja sen sisältöjä tutkitaan.

Tutkimuksen aloitusvaiheessa oli huojentavaa lukea Vallin (2015, s. 142) lause: ”Reliabiliteetti on yksi vaikeimmista ja abstrakteimmista käsitteistä, joihin aloitteleva tutkija tutkimustaan tehdessään törmää”. Lause piti melko hyvin paikkansa, sillä mittauksen reliabiliteetti todella on laaja käsite, joka kattaa useita luotettavuuden ulottuvuuksia. Tämä tutkimus ei käsittele observoinnin reliabiliteettia kokonaisuutena, ja tutkimus tulisivikin nähdä katsauksena yhteen reliabiliteetin osa-alueeseen eli havainnoijien väliseen yhteneväisyyteen. Iso osa aiheeseen liittyvästä tutkimuksesta on tehty muussa kuin varhaiskasvatuksen kontekstissa, ja tässä tutkimuksessa viitataan lähteisiin, jotka sijoittuvat sosiologiaan, psykologiaan ja lääketieteeseen. Pisteistä ja testeistä huolimatta tekstien keskeinen sisältö liittyy tutkittavaan aiheeseen eli havainnoinnin luotettavuuteen.

LoCasale-Crouch ym. (2007) huomauttavat luokkahuonetilanteiden arvioinnin olevan monimutkainen tehtävä. Suoran havainnoinnin lisääntyminen arvioinnin työkaluna kasvatuksessa ja koulutuksessa on kuitenkin luonut tarpeen luotettaville mittareille ja menetelmille. (LoCasale-Crouch ym., 2007, s. 5.) Koulun tavoin myös varhaiskasvatuksen

arviointiin on olemassa lukuisia mittareita, ja Sheridan (2007, s. 198) toteaa sopivan mittarin valitsemisen arviointia varten olevan ensiarvoisen tärkeää.

Rentzoun (2017) mukaan varhaiskasvatusta arvioivat mittarit ovat usein yhdistelmä metodologisia kuvauksia, epistemologisia oletuksia sekä havainnoitsijan ja havainnoitavan suhteeseen liittyviä suosituksia. Jotkut mittareista ovat tarkoitukseensa sopivia siinä mielessä, että niiden tavoitteena on todella tavoittaa ja kartoittaa laadukasta pedagogiikka ja sitä kautta kehittää varhaiskasvatuksen laatua. Samalla Rentzou kuitenkin huomauttaa joidenkin mittareiden olevan vain asianomaisten viranomaisten laatimia yksinkertaisia rasti ruutuun -listoja, joiden kautta voidaan todeta tietyt minivaatimukset saavutetuiksi. (Rentzou, 2017, s. 4.) Mielestäni näiden mittareiden tarkastelu ja arvioiminen ovat ensiarvoisen tärkeitä aiheita myös suomalaisen varhaiskasvatuksen kontekstissa: mitä mitataan, miksi mitataan, kuinka mitataan? Tämä tutkimus tarkastelee sitä, *kuinka luotettavasti* mitataan.

Yhteneväisten merkintöjen lisäksi tässä tutkimuksessa tarkastellaan observointiin liittyviä keskeisiä kasvatustieteellisiä teorioita ja malleja. Mielestäni tämän tutkimuksen kysymykset observoinnin yhteneväisyydestä vaativat rinnalleen ymmärryksen niiden kontekstista: mistä lähtökohdista observointia toteutetaan ja miksi se on päätynyt hankkeen aineistonkeruumenetelmäksi. *Kehittävä palaute varhaiskasvatuksessa* -hanke on laaja kokonaisuus, joka sisältää useita tutkimuksenaiheita. Observoinnin arviointi on jatkuva prosessi, jossa tämän tutkimuksen sisällöt toivottavasti osoittautuvat hyödyllisiksi. Tämä tutkimus on rajattu merkintöjen yhteneväisyyteen, mutta observoinnista riittäisi ainesta useaan tutkimukseen ja julkaisuun.

2 Kehittävä palaute varhaiskasvatuksessa -hanke

Tässä osiossa tutustutaan *Kehittävä palaute varhaiskasvatuksessa* -hankkeen historiaan ja siihen liittyviin kasvatustieteellisiin malleihin. Tämän jälkeen käydään läpi observoinnin toteuttaminen käytännössä. Lopuksi tarkastellaan observoinnin yhdeksää mittaria ja sitä, miten niiden sisällöt ovat muotoutuneet. Jatkossa hankkeeseen tullaan viittaamaan *Kehittävä palaute* -nimellä.

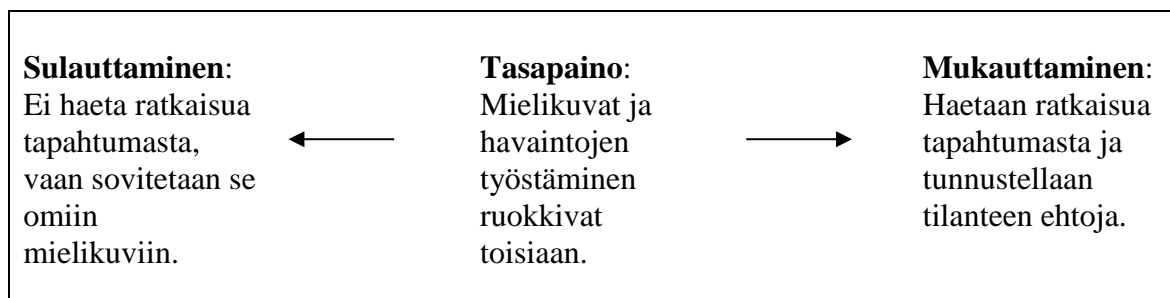
2.1 Hankkeen historia ja kasvatustieteellinen teoriatausta

Kehittävä palaute -hankkeen observoinnin teoreettinen viitekehys perustuu Jyrki Reunamon orientaatioteoriamallin. Reunamon (2007) mukaan päiväkodin tapahtumat koostuvat pääasiassa lasten tuottamasta aineksesta. Lapset eivät tuota tätä ainesta vain reagoimalla tapahtumiin, vaan heidän omat perusteensa ja näkemyksensä ohjaavat heidän osallistumistaan. (Reunamo, 2007, s. 11.) Mitä pienempi lapsi on, sitä enemmän hän toimii välittömästi käsillä olevien asioiden kanssa. Hänen todellisuutensa muuttuu hänen toimintansa seurauksena samalla, kun hän itse muuttuu ympäristönsä kanssa vuorovaikutuksessa. Lapselle ei ole vielä kehittynyt abstraktien skeemojen verkkoa, ja hänen toiminnallisissa skeemoissaan on jotain, mitä aikuisen on vaikea tavoittaa. (Reunamo, 1998, s. 10.) Reunamon (2007, s. 19) näkemyksen mukaan lapset mukauttavat näkemyksiään paremmin aikuisen maailmaan sopivaksi, mutta myös aikuiset suhtautuvat lapsiin eri tavalla sen mukaan, miten lapset näkevät tilanteet. Koska kaikilla tilanteeseen osallistuvilla on vaikutusta lopputulokseen, täytyy huomioida kunkin osallistujan tuottamat sisällöt (Reunamo, 2007, s. 28).

Orientaatioteoriamallin keskiössä on Reunamon (2007) mukaan muutoksen käsite: toiminnan aiheuttama muutos tapahtuu joko tarkastelijan näkemyksen muuttumisena tai vaihtoehtoisesti tarkastelijan näkemys muuttaa ympäristöä. Toisessa päässä mielikuvia ja havaintoja muokataan, toisessa päässä niitä sovelletaan tai omaksutaan. Mallin ajatuksena on, että lapset ja aikuiset mukauttavat omia näkemyksiään toiminnassa saadakseen paremman yhteyden toisiinsa. He luovat yhteisen kartan, jolla suunnistaa. Samalla he tuottavat polkua yhdessä. Tämä polku eli lapsen ja aikuisen tuottama toiminta, sekä se,

miten sitä tarkastellaan, korostuu mallin oppimisen käsityksessä. (Reunamo, 2007, s. 18–20.)

Monet orientaatioteoriamallin keskeisistä käsitteistä on jäsenneilty nelikenttiin havainnollistamaan erilaisten näkemysten toiminnallisia vaikutuksia. Reunamon (2007, s. 6) mukaan nelikentän tarkoituksena ei ole kuvata erilaisten ihmisen toimintatapojen tyyppjä vaan tarkastella ideoiden ja ympäristön molemminpuolista vaikutusta. Orientaatioteoriamallissa on kaksi jatkumoa: akkommodaatio–assimilaatio (mukauttaminen–sulauttaminen) ja adaptaatio–agenssi (sopeuttaminen–vaikuttaminen).

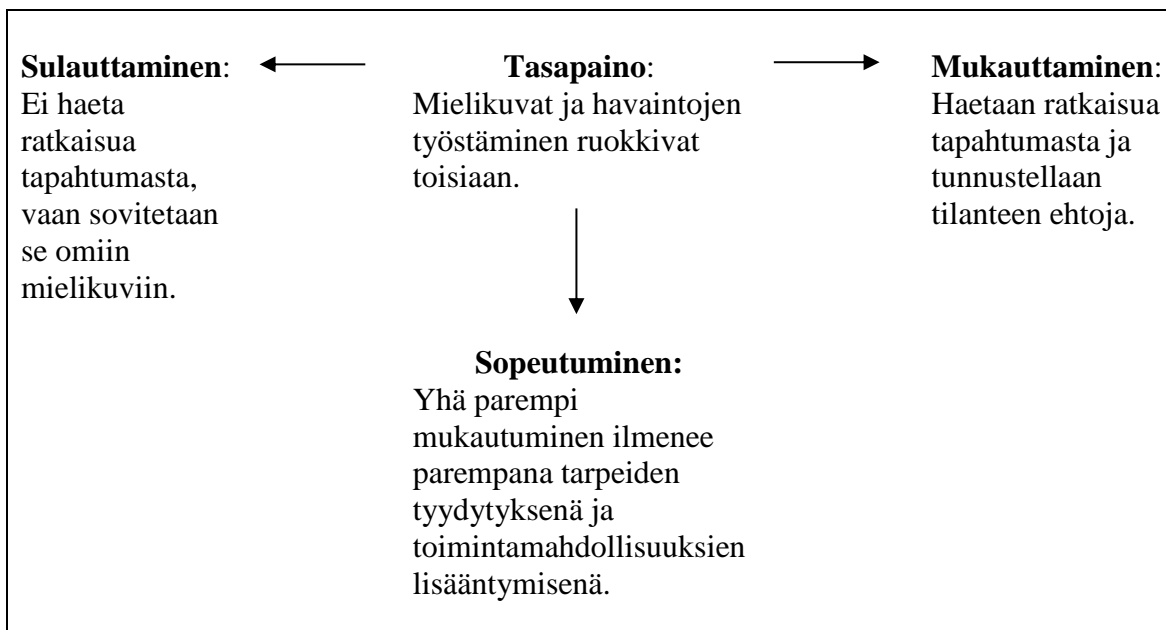


Kuvio 1. Sulauttamisen ja mukauttamisen käsitteet (Reunamo, 2007, s. 22)

Yllä olevassa *kuviossa 1* on tiivistettynä assimilaation eli sulauttamisen ja akkommodaation eli mukauttamisen jatkumoa orientaatioteoriamallin näkemysten perusteella. Muun muassa Jean Piaget on käsitellyt assimilaation ja akkommodaation käsitteitä geneettisen epistemologian kontekstissa. Piaget'n teoriassa assimilaatiolla eli sulauttamisella tarkoitetaan Gruberin ja Vonèchen (1995, s. 216) mukaan ulkoisen maailman muuttamista olennaiseksi osaksi minua. Piaget on itse määritellyt assimilaation tapahtumana, jossa uusi kohde sulautetaan jo olemassa olevaan malliin tai skeemaan (Piaget & Inhelder, 1977, s. 15).

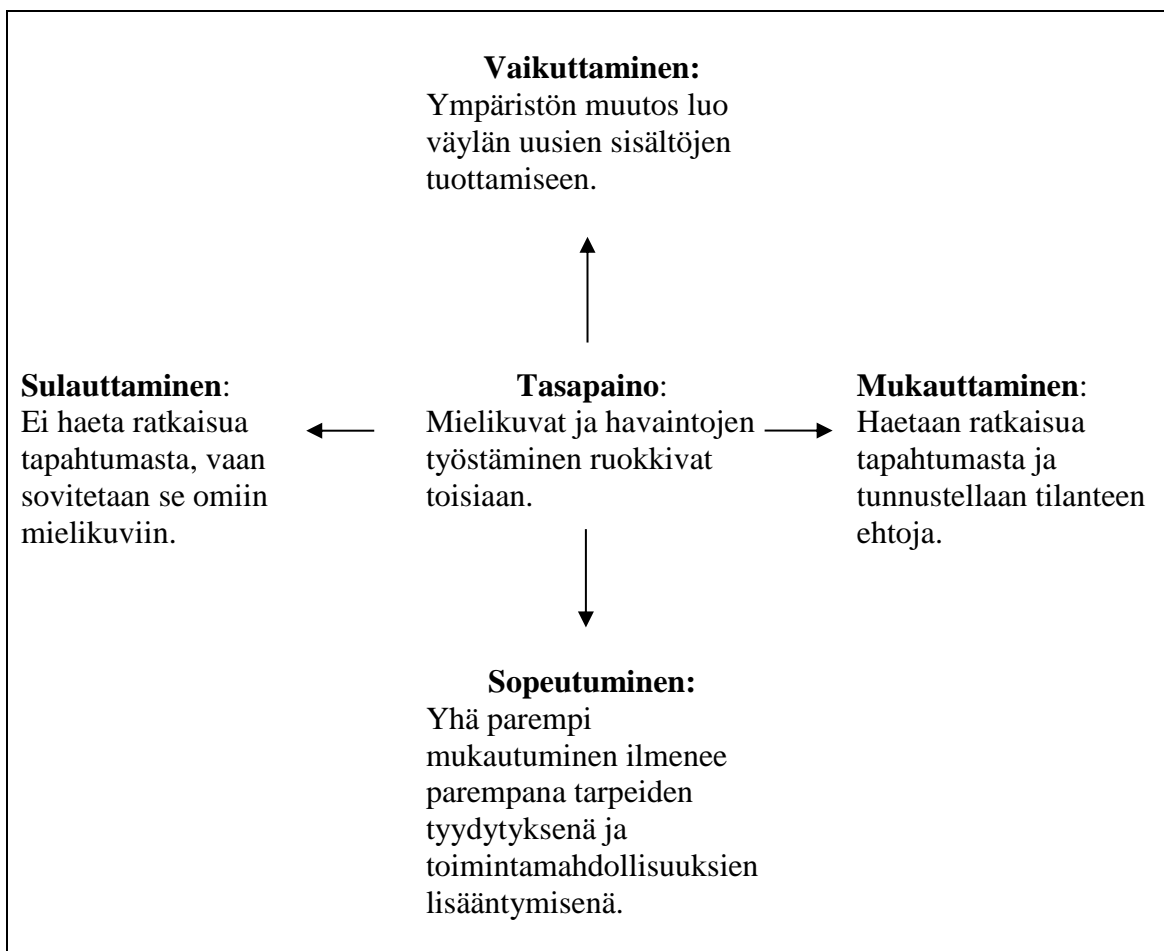
Kitchenerin (1986, s. 54) mukaan akkommodaation eli mukauttamisen voi nähdä prosessina, jossa subjekti ja sen rakenteet muokkautuvat vastaamaan paremmin ulkoista objektia: kuten biologinen organismi myös yksilö ja sen kognitiiviset prosessit mukautuvat sopimaan ympäristöönsä. Piaget'n (1976, s. 18) määritelmän mukaan akkommodaatio on ulkoisesta objektista johtuva muutos skeemassa, ja Reunamo (2007, s. 20) näkee

mukauttamisen havaitsemisena ja valmiutena muuttaa näkemyksiä. Orientaatioteoriamallin kolmas keskeinen käsite on adaptaatio eli sopeutuminen.



Kuvio 2. Sopeutumisen käsite (Reunamo, 2007, s. 25)

Yllä olevassa *kuviossa 2* sopeutuminen on lisätty orientaatioteoriamallin näkemykseen muutoksesta. Kitchener (1986) näkee adaptaation eli sopeutumisen yksilön ja ympäristön välisenä tasapainona, joka muuttuu ympäristön muuttuessa. Hänen mukaansa sopeutuminen on yhteistyötä sulauttamisen ja mukauttamisen välillä, jolloin sopeutuminen voidaan nähdä sisäisten mielikuvien käytön ja ympäristön havainnoinnin vuoropuheluna. (Kitchener, 1986, s. 54–55.) Reunamon (2007, s. 25) mukaan adaptaatio lisää mahdollisuuksia käyttää olemassa olevaa ympäristöä hyödyksi. Mallin neljäntenä ja viimeisenä ulottuvuutena on agenssi eli vaikuttaminen.



Kuvio 3. Vaikuttamisen käsite (Reunamo, 2007, s. 27)

Yllä olevassa *kuviossa 3* vaikuttaminen viimeistelee orientaatioteoriamallin näkemyksen muutoksen neljästä keskeisestä käsitteestä. Reunamo (2007, s. 26–27) näkee agenssin johonkin vaikuttavana toimintana, jonka kautta luodaan väyliä uusien sisältöjen tuottamiseen. Caiman ja Lundegard (2014) toteavat agenssiin liittyvän useita eri määritelmiä. Filosofissa ja sosiologiassa se voidaan nähdä sinä kapasiteettina, jonka kautta ihminen voi vaikuttaa omaan elämäänsä. (Caiman & Lundegard, 2014, s. 437.)

Kuviossa 3 neljä käsitettä muodostavat orientaatioteoriamallin keskeisen näkemyksen muutoksesta. Orietaatioteoriamallin näkemyksen mukaan lapsen kasvu on siis mukautumista ympäristöön, omien mielikuvien soveltamista, sopeutumista sekä toimintaympäristöön vaikuttamisen harjaantumista. Lasten näkemysten osuutta varhaiskasvatuksen sisältöihin ei siis saada selville vain tarkastelemalla tapahtuvaa

toimintaa, vaan ne on kytkettävä yhteen esiin nousevien tapahtumien kanssa. (Reunamo, 2007, s. 11–12.)

Oriantaatioteoriamalli on melko monimutkainen kokonaisuus, jossa teorian siirtymistä käytännön toimintaan on välillä haastavaa hahmottaa. Mallin käyttöön liittyvää vahvistavaa tai kritisovaa kolmannen osapuolen tutkimusta ei myöskään ole helppo löytää, mikä tekee mallin toimivuuden arvioinnista vaikeaa. Yllä esitetyn mallin ulottuvuuksiin perustuen *Kehittävä palaute* -hankkeen lapsihavainnoinnissa keskitytään siis lapsen toiminnan havainnointiin sekä sen ympäristön havainnointiin, jossa lapsi toimii.

2.2 Observointi käytännössä

Jokainen observoija toteutti observointia 2017 syksyllä 2–5 eri päiväkodissa kunnasta riippuen. Jokaisessa päiväkodissa havainnoitiin 1–5 lasta, neljä tuntia päivässä kahtena eri päivänä. Lapset olivat valikoituneet satunnaisesti niistä lapsista, joiden vanhemmilta oli saatu lupa osallistua observointiin. Kaikki observoijat käyttivät havaintojen merkitsemiseen tablettia ja sähköistä lomaketta (liite 1).

Lomakkeeseen merkittiin observoijan oma koodi, päiväkodin koodi sekä observoitavan lapsen numero. Lapsia observoitiin kronologisessa numerojärjestyksessä 1–5. Yhden observoinnin kesto oli 5 minuuttia, josta 2 minuuttia käytettiin tilanteen hahmottamiseen ja observoijan asemoitumiseen, 1 minuutti lapsen havainnointiin ja viimeiset 2 minuuttia havaintojen merkitsemiseen tabletille. Tämän jälkeen lomake lähetettiin suojattuun pilvipalveluun. Yhden observoinnin suoritettuaan observoija siirtyi listassa seuraavan lapsen kohdalle, ja listan viimeisen lapsen observoituaan hän siirtyi taas observoimaan listan ensimmäistä lasta. Kyseisellä systemaattisen otannan menetelmällä yhdelle observoijalle kertyi 4 tunnin observoinnin aikana noin 50 observointia. (Reunamo, 2017, s. 11.)

Observoitujen lasten ryhmässä työskennelleet kasvattajat sekä observointipäiväkodin johtajat täyttivät syksyllä 2017 omat sähköiset lomakkeensa, joissa he vastasivat erilaisiin kysymyksiin työstä ja työympäristöstä. Kaikkiin aineistoihin merkitty päiväkotikoodi mahdollisti aineistojen yhdistämisen kunnittain ja alueittain. Aineistot tallentuivat

pilvipalveluun, johon vain kuntien yhdyshenkilöillä oli pääsy. Pilvipalvelussa aineistoa pystyi yhdistelemään, vertaamaan ja analysoimaan. Yksittäisten lasten ja päiväkotien tunnistaminen oli estetty tietoturvasyistä, ja jokaisella kunnalla oli pääsy vain oman datansa analysointiin. (Reunamo, 2017, s. 12.)

2.3 Observointimittarit

Seuraavaksi tarkastellaan *Kehittävä palaute* -hankkeen observointimittareiden rakentumista. Mittari-sana voi kuulostaa hieman erikoiselta puhuttaessa lasten arkisen toiminnan havainnoinnista. Käyttäytymistieteissä sanaa mittari käytetään kuitenkin melko laajasti, ja termillä kuvataan kaikkia niitä välineitä, joilla alkuperäinen, tutkimuksessa käsiteltävä ja tulkittava tieto hankitaan (Karma, 1983, s. 54). Mittarit perustuvat aikaisemmin esiteltyyn orientaatioteoriamalliin sekä fyysiseen aktiivisuuteen (Puhl, Greaves, Hoyt, & Baranowski, 1990), sitoutuneisuuteen (Laevers, 1994) ja tunnetiloihin (Ekman, 1992a) liittyviin malleihin. Jokainen malli käydään läpi siltä osin kuin on tarpeellista siihen liittyvän mittarin hahmottamisen kannalta. *Kehittävä palaute* -hankkeen mittarit ja niiden sisällöt ovat tarkentuneet vuosien aikana, ja niitä on muokattu hankkeen eri vaiheissa (ks. Reunamo 2007, s. 167–169).

Observoinnissa käytettiin yhteensä yhdeksää eri mittaria. Näiden mittareiden kautta jokaisen observoinnin kohdalla tehtiin merkintä yleisestä tilanteesta, lapsen omasta toiminnasta, lapsen pääasiallisesta huomion kohteesta, lähimmästä sosiaalisesta lapsikontaktista, lapsen fyysisestä aktiivisuudesta, lapsen sitoutuneisuudesta, emootiosta, sosiaalisesta orientaatiosta ja lähimmän aikuisen toiminnasta. Osa mittareista keskittyy orientaatioteoriamallia mukaillen ympäristöön, jossa lapsen näkemykset vaikuttavat, ja osa keskittyy lapsen näkemyksen muutokseen kyseisessä ympäristössä. Mittarit on kuvattu alempana. Jokaisen mittarin kohdalla on kuvio, joka sisältää mittarin kategoriat niin kuin ne ovat observointilomakkeessa ja kuvion alla on mittarin sisältöjä avaava teksti.

- A. Yleinen tilanne: Lapsen toiminnan yleinen kehys (mitä lapsen pitää/kuuluu tehdä, sisältää siirtymät tilanteiden välissä)
1. Suora kasvatustoiminta sisällä (enimmäkseen aikuisen aloittamaa toimintaa, suunniteltua toimintaa: opetusta, ohjeistusta, ryhmätoimintaa, kirjan lukua, esityksiä, ryhmäkeskustelua).
 2. Tuettu leikki sisällä (lapsilähtöistä, aikuinen havainnoi, osallistuu ja on valmis tarvittaessa tukemaan leikkiprosesseja).
 3. Vapaa leikki sisällä (itsenäisempi leikki yksin tai kavereiden kanssa, kasvattaja ei osallistu prosessiin).
 4. Ulkotoiminta (pihalla, puistossa, retkellä).
 5. Perushoito (pukeminen, riisuminen, WC, hygienia, voi sisältää odotusta ja valmistautumista).
 6. Ruokailu (aamupala, lounas, ruuan odotus, järjestelyt ennen/jälkeen syömisen).
 7. Lepohetki.

Kuvio 4. A. Yleinen tilanne (liite 1)

Mittariin *A. Yleinen tilanne (kuvio 4)* merkitään tilanteen toiminnan kehys. Mittarin kategoriat perustuvat aikaisempaan tutkimukseen (ks. Reunamo, 2000), jossa strukturoidun havainnoinnin kautta luokiteltiin neljän helsinkiläisen päiväkodin tapahtumia. Nämä havainnot luokittelemalla saatiin yleiskuva päiväkodin toiminnan jakautumisesta erilaisiin luokkiin. (Reunamo, 2007, s. 30.)

B. Lapsen oma toiminta (mitä lapsi tekee)

1. Roolileikki tai mielikuvaleikki (lelulla tai itsellä on rooli).
2. Sääntöleikki (esim. pallopelejä, lautapelejä, elektroninen peli, pysyvät säännöt, kilpailu).
3. Esineleikki ja kokeilu materiaaleilla ja leluilla (esim. hiekkalaatikolla, paperilla).
4. Fyysinen leikki tai aktiviteetti (juoksu, keinuminen, hyppiminen, riehuminen, tanssi, kiipeily, harjoitus).
5. Lukeminen, opettaja lukee tai lapsi lukee/selaa kirjaa itse.
6. Tehtävä (esim. kotitehtävä, kynä-paperityöt, pukemisen harjoittelu, lorun opettelu, taidon harjoitus, työ).
7. Toiminta ilman kiintopistettä tai kunnon kontaktia muihin (esim. kuljeskelu, etsiminen, odottelu).
8. Yhdessäolo muiden kanssa (esim. hengailee/kävelee kavereiden kanssa, juttelee kiipeilytelineellä).
9. Kielletty toiminta (esim. ohjeiden rikkominen, kiusaaminen, häirintä).
10. Lapsi toimii yleisen kehyksen sisällä, mutta ei sovi yllä oleviin luokkiin B1–B9 (esim. lapsi syö ruokailussa, pukeutuu ulos lähettäessä).
11. Muu toiminta (ei sovi muihin luokkiin, esim. sekavuus, paljon vaihdoksia, ei rakennetta).

Kuvio 5. B. Lapsen oma toiminta (liite 1)

Mittariin *B. Lapsen oma toiminta* (kuvio 5) merkitään se toiminta, mitä lapsi tekee.

Mittarin käytön tavoitteena on tarkastella sitä, mitä muuta lapset tekevät päiväohjelmaan kirjattun toiminnan lisäksi. Esimerkiksi ruokailussa lapset voivat ruokailun lisäksi jutella keskenään tai hakea aikuisen huomiota, tuottaen toimintaan omaa ainesta. (Reunamo, 2007, s. 32–33.)

C. Lapsen pääasiallinen huomion kohde

1. Ei-sosiaalinen kohde (esim. lelu, hiekka, auto, palikat, vesi, oma itse).
2. Aikuinen (esim. seuraa aikuisen kertomusta, keskustelee, huomio voi sisältää esim. opetusmateriaalia).
3. Toinen lapsi (mukana voi olla myös leluja ja muita ei-sosiaalisia kohteita esim. lapsen kädessä).
4. Ryhmä lapsia (2 tai useampia lapsia, mukana voi olla myös leluja ym.).
5. Kokonaistilanne (tilanteessa on niin paljon elementtejä, ettei pääasiallista kohdetta voi nimetä, esim. lapsia, aikuisia, materiaaleja, usein dynaaminen tilanne).

Kuvio 6. C. Lapsen pääasiallinen huomionkohde (liite 1)

Mittarin *C. Lapsen pääasiallinen huomionkohde (kuvio 6)* tavoitteena on tarkastella, miten lapsen huomio suuntautuu suhteessa päiväkodin toimintaan. Reunamo (2007) viittaa aikaisempiin aineistoihin todetessaan esimerkiksi sisäleikissä lasten suuntaavan usein huomion toisiinsa, kun toisaalta ruokailussa sekä perushoitotilanteissa lapset antavat huomion harhailla kohteesta toiseen. Aikuiset ovat mukana kokonaistilanteissa, mutta lapsen huomio kohdistuu harvoin pelkästään aikuiseen. (Reunamo, 2007, s. 35.)

D. Lähin sosiaalinen lapsikontakti

1, 2, 3, 4, 5, 6

Jos ei kontaktia tutkittaviin lapsiin, valitse 6.

Kuvio 7. D. Lähin sosiaalinen lapsikontakti (liite 1)

D. Lähin sosiaalinen lapsikontakti -mittarille (*kuvio 7*) tehdään merkintä, jos lapsella on observoinnin aikana kontakti toiseen hankkeen observointiin osallistuvaan lapseen.

Mittarin vaihtoehdoksi merkitään se numero, joka kontaktilapselle on annettu observoinnin alussa (1–5) tai vaihtoehto 6, jos lapsella ei ole kontaktia muihin observoitaviin lapsiin.

- E. Lapsen fyysinen aktiivisuus (painotetaan viimeistä 10 sekuntia)
1. Liikkumaton (esim. nukkuminen, seisominen, istuu vaunuissa ja seisominen).
 2. Paikallaan tapahtuva raajojen/kehon liike (esim. nouseminen ylös, kevyehkön esineen liikuttaminen).
 3. Hitaat ja kevyet liikkeet (esim. käveleminen, kevyt keinuminen).
 4. Kohtalaisesti kuormittavat liikkeet (esim. käveleminen ylämäkeen, vähintään 2 hyppelyä, kiipeily telineessä, käsillä roikkuminen ja vauhdinotto jaloilla).
 5. Nopeat ja kuormittavat liikkeet (esim. juokseminen, portaiden nousu, vähintään kolme hyppelyä, kiipeilytelineessä roikkuminen jalat liikkeessä).

Kuvio 8. E. Lapsen fyysinen aktiivisuus (liite 1)

E. Lapsen fyysinen aktiivisuus -mittarilla (kuvio 8) havainnoidaan lapsen liikkumista. Mittari pohjautuu CARS (Children's Activity Rating Scale) -havainnointimenetelmän asteikkoon. Puhlin ym. (1990) kehittämä CARS luokittelee lapsen fyysisen aktiivisuuden muun muassa arvioimalla energiankulutusta ja sykettä sekä näihin yhteydessä olevia kehonliikkeitä. Alkuperäinen CARS-asteikko on: 1. paikoillaan oleva ja liikkumaton, 2. paikoillaan oleva ja liikkuva, 3. hidas liikkuminen, 4. kohtalainen liikkuminen ja 5. nopea liikkuminen. (Puhl ym., 1990, s. 26–27.)

- F. Lapsen sitoutuneisuus
1. Yksinkertainen, kaavamainen, toistuva, passiivinen toiminta, ei energiaa, ei älyllistä haastetta.
 2. Usein keskeytyvä toiminta ja toimintaan kiinnittyminen.
 3. Enimmäkseen jatkuva toiminta, huomio hajoaa helposti, vähän energiaa.
 4. Jatkuva toiminta jossa intensiivisiä hetkiä, tarkkaavaisuus ei helposti hajoa.
 5. Kestävä intensiivinen toiminta, keskittyminen, luovuus, energia ja sitkeys.

Kuvio 9. F. Lapsen sitoutuneisuus (liite 1)

F. Lapsen sitoutuneisuus -mittari (kuvio 9) mukailee Ferre Laeversin kehittämää LIS-YC (Leuven Involvement Scale for Young Children) -menetelmää. LIS-YC-menetelmä sisältää kaksi osaa: sitoutuneisuuden tunnusmerkit ja asteikon eri tasot. Hautamäen ja Laeversin

(1997) mukaan lapsen sitoutuneisuutta voi tarkastella tiettyjen tunnusmerkkien kuten keskittymisen, energian, sinnikkyyden ja tarkkuuden avulla. Näiden tunnusmerkkien pohjalta lapsen sitoutuneisuus voidaan määritellä seuraavalla asteikolla: 1. ei toimintaa, 2. toistuvasti keskeytyvä toiminta, 3. jossain määrin jatkuva toiminta, 4. intensiivisiä hetkiä sisältävä toiminta ja 5. pitkäkestoinen intensiivinen toiminta. (Hautamäki & Laevers, 1997, s. 6–10.)

G. Emootio

1. Viha.
2. Ärtymys.
3. Pelko, ahdistus.
4. Suru, masennus.
5. Ilo, riemu.
6. Onnellisuus, tyytyväisyys.
7. Hämmästys, valppaus, uteliaisuus, innostus.
8. Neutraali, tyyni, rauhallinen.
9. Muu tunne.

Kuvio 10. G. Emootio (liite 1)

G. Emootio -mittarin (*kuvio 10*) tavoitteena on saada havainto lapsen senhetkisestä tunnetilasta. Mittarin kategoriat pohjautuvat Paul Ekmanin (1992a) näkemykseen niin sanotuista perustunteista: viha, pelko, ilo, suru ja inho. Perustunteilla tarkoitetaan Ekmanin mukaan tunteita, joiden ominaispiirteet esiintyvät universaalisti kaikissa kulttuureissa. (Ekman, 1992a, s. 175–176.) Myöhemmin Ekman (1999, s. 55) on tarkentanut näkemystään ja lisännyt listaan muun muassa nolouden, innokkuuden, helpotuksen, nautinnon ja häpeän. Mittarin kategoriat eivät kuitenkaan suoraan mukaile kumpaakaan näistä määritelmistä, ja osa kategorioista pitää sisällään useita tunnetiloja. Selkeää lähdettä juuri *kuvion 10* kategorioille ei siis ole.

Ekmanin yli 50 vuoden tutkimus ihmisen tunnetiloista on perustunut vahvasti kasvoniilmeiden ja -liikkeiden tutkimukseen sekä tunnetilojen nimeämiseen valokuvien perusteella (ks. Ekman, 1989; 1992b; Ekman & Friesen, 1976). Ekmanin tutkimus liittyy

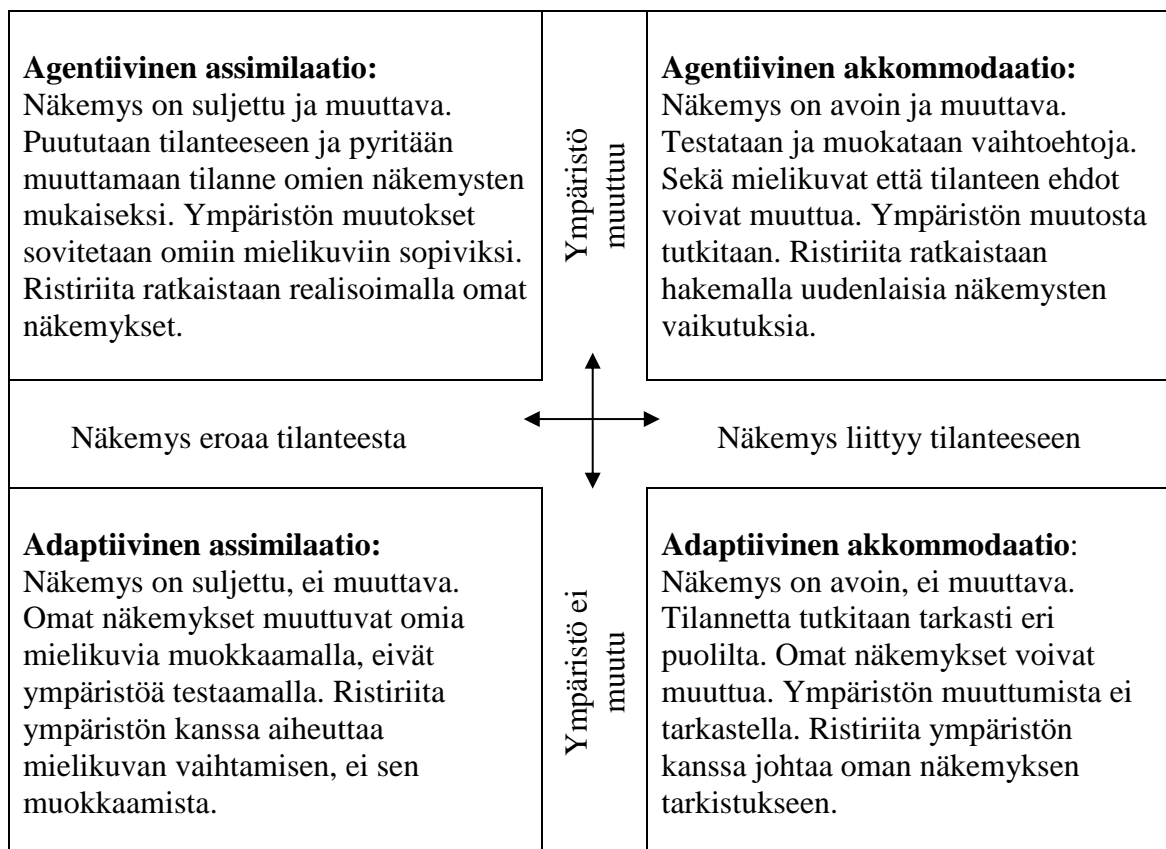
siis vahvasti mittarin aiheeseen eli tunnetiloihin, mutta hänen käyttämänsä menetelmät niiden havaitsemiseen ovat kovin erilaiset verrattuna siihen, miten lapsen tunnetiloja havainnoidaan *Kehittävä palaute* -hankkeessa. Ekmanin tutkimus tuntuu painottuvan siihen, miten eri tunteet ilmenevät fyysisesti ja siihen, mitkä tekijät erottavat yhden tunteen toisesta. Hän ei kuitenkaan väitä, että nämä erot olisivat selvästi havainnoitavissa arkisissa tilanteissa. Tähän aiheeseen ja tunteiden havainnoimisen haastavuuteen palataan tulosten pohdinta -osiossa.

H. Sosiaalinen orientaatio

1. Mukautuva ja avoin, hyväksyy ja huomioi.
2. Osallistuva, vuorovaikutteinen, avoin, tekee yhteistyötä.
3. Dominoiva, itsepintainen, ehkä itsekeskeinen tai määräilevä.
4. Vetäytyy sosiaalisesta tilanteesta, ei-sosiaalinen, ei-vuorovaikutteinen.
5. Ei voida määrittää.

Kuvio 11. H. Sosiaalinen orientaatio (liite 1)

H. Sosiaalinen orientaatio -mittarin (*kuvio 11*) tavoitteena on havainnoida lapsen näkemystä ja orientoitumista sen hetkiseen tilanteeseen. Termi sosiaalinen orientaatio ja sen määritelmä liittyvät aikaisemmin esitettyyn orientaatioteoriamalliin. *Kuviossa 3* esiteltiin horisontaalinen jatkumo, joka kuvasi hahmottamisen ideoita. Kokemus vaatii olemassa olevien rakenteiden mukauttamista (akkommodaatio) tai kokemuksen jo olemassa oleviin rakenteisiin sulauttamista (assimilaatio). Vertikaalisessa jatkumossa keskeistä on muutos joko henkilössä (adaptaatio) tai ympäristössä (agenssi). Reunamon (2007, s. 37) mukaan yhdistämällä nämä kaksi jatkumoa luodaan nelikenttä eri tavoista suhtautua muutokseen. Eri tavat suhtautua muutokseen on esitelty alla *kuviossa 12*.



Kuvio 12. Hahmotuksen agenssin nelikenttä muokattuna (Reunamo, 2007, s. 69, 87)

Kuviossa 12 on orientaatioteoriamallin neljä erilaista tapaa suhtautua muutokseen.

Kuviossa on yhdistetty kaksi eri nelikenttää (ks. Reunamo, 2007, s. 69, 87) joiden sisällöt ovat yhteensopivia. Reunamon (2007, s. 43) mukaan orientaatioteoriamallissa adaptiivinen akkommodaatio eli sopeutuva mukauttaminen on tilanne, jossa tilannetta tarkastellaan toisen silmin ja muiden näkemykset hyväksytään. Adaptiivinen assimilaatio eli sopeutuva sulauttaminen vuorostaan tarkoittaa tilannetta, jossa omia näkemyksiä sovelletaan siten, ettei ympäristö muutu. Näkemys on erillinen, sillä ihminen suhtautuu asiaan mielikuvansa mukaisella tavalla eikä sulautta ympäristön ehtoja omaan mielikuviinsa. (Reunamo, 2007, s. 47.) Agenttiivinen assimilaatio eli vaikuttava sulauttaminen on tilanne, jossa ympäristö muuttuu tietyn näkemyksen mukaiseksi. Näkemys on omaehtoinen, sillä tilanteessa käytetään omaa mielikuvaa ja samalla muutetaan tilanteen ehtoja. (Reunamo, 2007, s. 54.) Agenttiivinen akkommodaatio eli vaikuttava mukauttaminen on tilanne, jossa henkilö muokkaa omia näkemyksiään saadun palautteen perusteella. Näkemys on osallistuva, sillä ympäristö nähdään avoimena ja muutettavana asiana, johon voi vaikuttaa. (Reunamo, 2007, s. 61.)

Orientaatioprojektin aikana vuonna 2007 kerättiin lapsihaastatteluaineisto, jossa lapsilta kysyttiin, miten he itse näkivät toimintansa eri tilanteissa. Lasten antamat vastaukset jaoteltiin *kuvion 12* mukaisesti neljään eri kategoriaan. (Reunamo, 2007, s. 39.) Tätä kautta luotu lapsen hahmotuksen agenssin nelikenttä on kuvattu alla *kuviossa 13*.



Kuvio 13. Lapsen hahmotuksen agenssin nelikenttä (Reunamo, 2007, s. 69)

Kuviossa 13 on orientaatioteoriamallin kuvaus lapsen näkemyksen vaikutustavoista eli neljästä tavasta suhtautua tilanteeseen. Sopeutuvissa näkemyksissä lapsi ottaa tilanteessa toisen näkökulman huomioon, ja mielikuvat ovat ei-muuttavia (Reunamo, 2007, s. 43). Lapsen näkemys on erillinen eli vetäytyvä, kun näkemys ei liity suoraan hänen kohtaamaansa tilanteeseen vaan hänen omiin ratkaisuihinsa: lapsi joko jättää tilanteen avaamatta tai ei pyri muuttamaan sitä (Reunamo, 2007, s. 50). Omaehtoinen eli dominoiva

näkemyks tarkoittaa ympäristön muuttamista jonkin näkemyksen mukaiseksi. Lapsi ei anna tilanteen vaikuttaa omaan mielikuvaansa: joko tilanne muuttuu tai lapsi jättää tilanteen ominaisuudet huomioimatta. (Reunamo, 2007, s. 54–56.) Osallistuvassa näkemyksessä lapsi ottaa tilanteessa muiden näkökulman huomioon, mutta hän ei pelkästään sopeudu tilanteeseen vaan pyrkii aktiivisesti myös muokkaamaan sitä. Lapsi suhtautuu tilanteisiin avoimesti ja osallistuu tapahtumien tuottamiseen yhdessä muiden kanssa. (Reunamo, 2007, s. 65.) Nämä neljä näkemystä (mukautuva, osallistuva, dominoiva, vetäytyvä) muodostavat kategoriat mittarille *H. Sosiaalinen orientaatio*, joka on esitetty *kuviossa 11*. Reunamon (2007) mukaan mittarin kautta saatu tieto lapsen tavasta tarkastella asioita voidaan yhdistää tietoon toiminnan jakautumisesta. Tämä mahdollistaa lasten näkemysten ja toiminnan välisen yhteyden tarkastelun. (Reunamo, 2007, s. 39.)

I. Lähimmän aikuisen toiminta

1. Vastaanottaa: havainnoi, kuuntelee, on avoin, herkistyy, hyväksyy, huomio (syventyy lapsen kokemukseen).
2. Osallistuu: vuorovaikutteinen, avoin, tekee yhteistyötä, rikastaa, kokeilee (syventää lapsen kokemusta).
3. Tavoittelee: opettaa, motivoi, ohjaa, harjoittaa (syventää ennalta määrätyn tavoitteen suuntaa).
4. Pinnallinen: ei-vuorovaikutteinen, valvoo, päättää, rutiini, ei huomioi (ei syvennä kokemusta).
5. Negatiivinen: arvostelee, mitätöi, syyttää, ivailee, uhkailee (jyrää lapsen kokemuksen).
6. Ei aikuisen huomiota.

Kuvio 14. I. Lähimmän aikuisen toiminta (liite 1)

I. Lähimmän aikuisen toiminta -mittarin (*kuvio 14*) luomisessa on käytetty samaa *kuvion 12* hahmotuksen agenssin nelikenttää, jota on käytetty jo pohjana *H. Sosiaalinen orientaatio* -mittarissa (*kuvio 11*). Reunamon (2007, s. 107) mukaan mielikuvien ja toiminnan vaikutusten keskinäiset muutokset saavat mielikuvat joko liittymään tilanteisiin tai eroamaan niistä. Nelikentän kaksi jatkumoa ovat tilanteen avoimuus ja ympäristön muuttuminen, ja nelikentän perusteella Reunamo (2007, s. 109) ehdottaa neljää erilaista

näkökulmaa varhaiskasvattajan suhtautumiseen tilanteessa: harmonia, mahdollisuudet, tavoitteet ja kaaos. Tässä tutkimuksessa ei paneuduta tarkemmin näiden näkökulmien merkityksiin varhaiskasvatuksen arjessa vaan keskitytään niiden rooliin mittarin rakentumisessa.

Mittarin *I. Lähimmän aikuisen toiminta* -kategorioita (*kuvio 14*) ei löydy suoraan lähdekirjallisuudesta. Kategoriat saadaan kuitenkin muodostettua yhdistämällä hahmotuksen agenssin nelikenttä *kuviossa 12* yllä mainittuihin kasvattajan tapoihin suhtautua tilanteeseen. Harmonisesti tilannetta katsova aikuinen voidaan nähdä vastaanottavana. Mahdollisuutena tilannetta tarkasteleva aikuinen voidaan nähdä osallistuvana, ja tilannetta tavoiteltavana tarkasteleva aikuinen voidaan nähdä omaehtoisena. (Reunamo, 2007, s. 112–113.) Kaaos sijoitetaan lähdekirjassa samaan kohtaan nelikentässä kuin erillinen näkemys, mutta selkeää yhteyttä pinnallinen- tai negatiivinen aikuisen toiminta -kategorioille ja agenssin nelikentälle ei lähteissä ole.

Orientaatioteoriamallin yhteys observointimittareiden rakentumiseen on haastava kokonaisuus, jossa yhteydet teorian ja käytännön välillä ovat välillä vaikeasti hahmotettavat. Sitoutuneisuus ja tunnetilat ovat tärkeitä asioita, ja näiden yhdistäminen osaksi varhaiskasvatuksen arviointia on mielenkiintoinen näkökulma. Abstraktien asioiden havainnoiminen tarkasti toisessa ihmisessä on kuitenkin äärimmäisen haastavaa, varsinkin kun havainnoinnin kohteena on pieni lapsi. Vielä haastavampaa on, jos tavoitteena on tuottaa luotettavaa tietoa useasta eri lähteestä. Tämän tutkimuksen tavoitteena on selvittää, miten tämä haastavuus näkyy käytännössä.

Pohdinnan arvoista on myös se, kuinka paljon valmiiksi määritellyt mittarit ja kategoriat ohjaavat havainnointia. Lapsen sosiaalisen orientaation tarkastelu nelikentän kautta tarjoaa kyllä neljä selkeää kategoriaa, mutta samalla se määrittelee observoijan tarkastelemaan monimutkaista aihetta näiden neljän vaihtoehdon kautta. Vaihtoehtojen väliset erot on melko selvästi kuvailtu teoriassa, mutta niiden havainnoiminen käytännössä voi olla haastavaa. Yhteisestä ohjeistuksesta huolimatta jokainen näkee, orientaatioteoriamallia mukailleen, todellisuuden lopulta omalla tavallaan. Tämä voi olla ongelmallista tilanteessa, jossa pyritään tuottamaan luotettavaa tietoa yleisluontoista tarkastelua varten. Näihin teemoihin palataan tulosten pohdinta -osiossa.

3 Reliabiliteetti ja validiteetti

Tutkimusta tehtäessä on tavoitteena valita sellaiset mittauksen välineet, jotka mittaavat tutkittavia muuttujia parhaalla ja tarkimmalla mahdollisella tavalla. Näiden mittareiden valintaan ja arviointiin liittyvät käsitteet reliabiliteetti ja validiteetti, jotka ovat tärkeimmät psykometriset ominaisuudet mittarin arvioinnin prosessissa.

3.1 Reliabiliteetti yleisesti

Johnsonin ja Christensenin (2004, s. 132) mukaan reliabiliteetti viittaa tuloksen johdonmukaisuuteen, ja validiteetti viittaa tuloksesta tehtävien päätelmien tai tulkintojen tarkkuuteen. Valli (2015, s. 142) määrittelee reliabiliteetin lyhyesti tutkimuksen luotettavuutena. Hirsjärvi, Remes ja Sajavaara (2007, s. 226) puolestaan toteavat reliabiliteetin tarkoittavan mittauksen toistettavuutta eli mittauksen kykyä antaa ei-sattumanvaraisia tuloksia.

Vaikka reliabiliteetti ja validiteetti ovat toisiinsa liittyviä käsitteitä, ne eivät aina kulje käsi kädessä. Anttila (2006) toteaa, että tutkimusmenetelmä voi olla reliaabeli eli se voi mitata luotettavasti mitattavaa asiaa olematta silti validi, jolloin tulos ei vastaa siihen mitä sen tulisi mitata. Mutta jos tutkimuksen reliabiliteetti on alhainen, se pätee yleensä myös tutkimuksen alhaiseen validiteettiin. (Anttila, 2006, s. 518.) Reliabiliteetin ja validiteetin merkityksen voi Anttilan (2006) mukaan havainnollistaa käyttämällä vertausta tikkataulusta. Tikkojen osuessa umpimähkään ympäri taulua ovat sekä reliabiliteetti että validiteetti heikkoja. Tikkojen osuessa samaan kohtaan taulua on reliabiliteetti vahva, mutta se ei vielä kerro onko sama tulos validi. Vasta kun tikat osuvat taulussa samaan ja toivottuun kohtaan, voidaan tuloksen nähdä olevan sekä tarkka että luotettava. (Anttila, 2006, s. 512.)

Reliabiliteetin mittaaminen on erilaista laadullisessa ja määrällisessä tutkimuksessa, ja tutkimuksen luotettavuuteen ja yleistettävyyteen suhtaudutaan molemmissa hieman eri tavoin. Kvalitatiivisessa tutkimuksessa reliabiliteetilla tarkoitetaan usein aineiston käsittelyn ja analyysin luotettavuutta. (Anttila, 2006, s. 517.) Uusitalon (1991) mukaan reliabiliteetti tarkoittaa kvantitatiivisessa tutkimuksessa mittauksen toistettavuutta ja ei-

sattumanvaraisuutta. Sen sijaan kvalitatiivisen tutkimuksen yhteydessä reliabiliteetista tulisi puhua väljemmin analyysin arvioitavuutena ja toistettavuutena. (Uusitalo, 1991, s. 82–84.)

Toistettavuutta voi Uusitalon (1991) mukaan heikentää tutkimuksen aikana monet asiat. Satunnaisvirheitä voi syntyä esimerkiksi, jos vastaaja muistaa vastatessaan jonkin asian väärin tai ymmärtää asian eri tavalla kuin tutkija, haastattelija merkitsee vastauksen lomakkeeseen väärin tai tutkija tekee virheitä tallentaessaan. Tavoitteeksi tulisi asettaa mahdollisimman reliabeli mittausta, mutta puutteellinen reliabiliteetti ei välttämättä ole tuhoisa asia. Tärkeintä on pyrkiä arvioimaan reliabiliteettia, jolloin on mahdollista myös arvioida sen vaikutuksia. (Uusitalo, 1991, s. 84.) Siihen, mikä on ”riittävä” reliabiliteetti, ei ole yksiselitteistä vastausta, ja joillakin alueilla luotettavaan mittaukseen pääseminen onkin Karman (1983, s. 55) mukaan tavattoman hankalaa. Tässä tutkimuksessa ei paneuduta syvemmin reliabiliteetin arviointiin kvalitatiivisesta näkökulmasta, koska aineistosta johtuen kiinnostuksen kohteena on reliabiliteetti nimenomaan kvantitatiivisen tutkimuksen kontekstissa.

3.2 Reliabiliteetti kvantitatiivisessa tutkimuksessa

Kvantitatiivisessa tutkimuksessa on eri tapoja reliabiliteetin laskemiseen ja arviointiin. Nimensä mukaisesti toistomittauksessa tehdään ensimmäinen mittausta, joka toistetaan uudestaan jonkin ajan kuluttua, minkä jälkeen tuloksia verrataan. Toistomittauksen ongelmana on se, että mitattava ilmiö voi muuttua kahden mittauksen välissä; esimerkiksi tässä tutkimuksessa uudelleen havainnoiminen ei olisi tuottanut samoja tuloksia. Mittarin sisäisen konsistenssin mittausta taas perustuu mittarin puolittamiseen kahteen osaan. Mikäli mittarin kaikki osiot todella mittaavat samaa asiaa, ei ole väliä, miten puolitus tehdään, jolloin kahden puoliskon välinen korrelaatio on reliabiliteetin mitta. Tähän ajatukseen perustuu esimerkiksi Cronbachin alfa (Cronbach, 1951), jonka voidaan tulkita antavan kaikkien mahdollisten mittarista muodostettavien puolitusten keskiarvon. (Metsämuuronen, 2011, s. 76–77.)

Cronbachin alfan käyttöä reliabiliteetin mittaamisen menetelmänä on kritisoitu (ks. Green & Yang, 2009; Cortina, 1993) ja jopa hän itse on tehnyt niin (ks. Cronbach & Shavelson,

2004, s. 403). Davenport, Davidson, Liou ja Love (2015, s. 4) toteavat, että Cronbachin alfaa erehdytään usein tarkastelemaan korrelaation sijaan yhteneväisyyden mittarina. Tämän tutkimuksen kannalta paras vaihtoehto reliabiliteetin laskemiseen on rinnakkaismittaus, jossa sama asia mitataan samaan aikaan kahdella mittarilla (Metsämuuronen, 2011, s. 75).

Havainnoitsijoiden välinen yhteneväisyys viittaa havainnoijien luotettavuuteen eli johdonmukaisuuteen mittareita käytettäessä. Korkea havainnoitsijoiden välinen yhteneväisyys liittyy tutkimuksen objektiivisuuteen ja vähentää virheitä arvioinnin yhteneväisyydessä. (Johnson & Christensen, 2004, s. 504.) Klassisessa testiteoriassa psykometrisen mittarin kautta tehty havainnoitu tulos koostuu mittauksesta ilman virhettä sekä mittauksen sisältämän virheen määrästä (Hallgren, 2012, s. 23).

Klassinen testiteoria	Varianssin määrä	Reliabiliteetti
$X = T + E$	$Var(X) = Var(T) + Var(E)$	$\frac{Var(T)}{Var(X)} = \frac{Var(X) - Var(E)}{Var(X)} = \frac{Var(T)}{Var(T) + Var(E)}$

Kuvio 15. Klassinen testiteoria, varianssin määrä & reliabiliteetti (Hallgren, 2012, s. 23–24)

Yllä *kuviossa 15* on kolme eri yhtälöä. Klassisen testiteorian yhtälössä havainnoitu tulos X on yhtä kuin mittaustulos ilman virhettä T yhdistettynä virheen määrään E . Mittausvirhe E pitää sisällään ne syyt, joiden takia havainnoinnin todellista arvoa X ei pystytä suoraan määrittämään. Syynä voi olla esimerkiksi epätarkkuus, epämääräisyys, huonosti suunniteltu havainnointiasteikko, mittarin epätasaisuus saman asian uudelleenmittaamisessa tai mittarin epätasaisuus kahden eri koodaajan käytössä. Klassinen testiteoria -yhtälöllä on vastaava Varianssin määrä -yhtälö, jossa havainnoidun tuloksen varianssi $Var(X)$ on yhtä kuin ilman virhettä mitatun tuloksen varianssi $Var(T)$ ja mittausvirheen varianssi $Var(E)$. (Hallgren, 2012, s. 23.)

Havainnoitsijoiden välisen yhteneväisyysanalyysin tavoitteena on selvittää, miten suuri osa havainnoidun datan varianssista johtuu tulosten todellisista arvoista sen jälkeen, kun mittausvirheitä aiheuttavat tekijät on poistettu yhtälöstä. Koska todellisia arvoja T tai mittausvirheitä E ei voida suoraan arvioida, havainnoitsijoiden välistä yhteneväisyyttä ei

voida myöskään suoraan laskea. Sen sijaan todelliset arvot voidaan arvioida määrittämällä kovarianssi eri havainnoijien samasta kohteesta tehdyissä havainnoissa X . Silloin oletetaan, että jaettu varianssi havainnoissa on lähellä $Var(T)$ arvoja, ja ei-jaettu varianssi havainnoissa on lähellä $Var(E)$ arvoja. Tämä mahdollistaa reliabiliteetin arvioinnin kuvion 15 reliabiliteetti-yhtälön mukaisesti. (Hallgren, 2012, s. 24.)

Arvioijien välisen yhteneväisyyden arviointiin ja vahvistamiseen on kehitetty useita tekniikoita. Esimerkiksi usein käytetty mutta alkeellinen menetelmä on yhteneväisten havaintojen osuuden laskeminen koko aineistosta. Tapa on helppo, mutta yhteneväisyydessä ei tällöin ole huomioitu lainkaan sattuman vaikutusta. Toinen yleinen tapa, khiin neliö -testi, ei mittaa havaintojen yhteneväisyyttä vaan niiden yhteyden vahvuutta Cronbachin alfan tavoin. (Banerjee, Capozzoli, McSweeney, & Debajyoti, 1999, s. 5.) Kumpikaan näistä menetelmistä ei siis ole riittävä tätä tutkimusta varten. Cohenin kappa sen sijaan mahdollistaa yhteneväisyyden laskemisen ottaen samalla huomioon sattuman merkityksen yhtälössä (McHugh, 2012, s. 276).

3.3 Cohenin kappa

Cohenin kappa on yhteneväisyyden mitta, jossa kaksi henkilöä arvioi samaa luokitteluasteikollista muuttujaa saaden arvon välillä -1-1 perustuen siihen, luokittelevatko kaksi henkilöä arviointikohteen vastaavanlaisesti toisistaan riippumatta (Metsämuuronen, 2011, s. 377). Kappa-kertoimella on useita laskennallisia muotoja (riippuen arvioinnin muodosta ja arvioijien määrästä) ja oikean version valitseminen on tulosten tarkkuuden kannalta tärkeää. Perusmuodossaan painottamattoman kappan (Cohen, 1960) avulla voidaan tarkastella kahden arvioijan yhteneväisyyttä. Yhtälön muunnelmien kautta on mahdollista laskea yhteneväisyys muun muassa useamman kuin kahden arvioijan välille (Fleiss, 1971), aineistolle jossa sama tieto merkitään useampaan kohtaan samalla mittarilla (Mezzich, Kraemer, Worthington, & Goffman, 1981) tai kun halutaan painottaa mittarin tietyn osion yhteneväisyyden tärkeyttä (Cohen, 1968). Tämän tutkimuksen tapauksessa, kun tavoitteena on kartoittaa kahden henkilön välistä yhteneväisyyttä ja merkinnät ovat nominaaliasteikollisia, on kappan painottamaton perusversio riittävä (ks. Brennan & Hays, 1992, s. 155; Bartko & Carpenter, 1976, s. 310). Jatkossa kappa-sanalla tarkoitetaan painottamatonta Cohenin kappa, ellei toisin mainita.

Kappan eri versiot mittaavat merkintöjen yhteneväisyyden astetta ottaen huomioon sattuman vaikutuksen yhtälössä. Yhteneväisyyden aste saadaan ristiintaulukoimalla kahden havainnoijan merkinnät, ja sattumanvaraisen yhteneväisyyden merkitys määritellään merkintöjen reunafrekvenssien avulla. (Hallgren, 2012, s. 26.)

$$K = \frac{N \sum_i x_{ii} - \sum_i x_{i+} x_{+i}}{N^2 - \sum_i x_{i+} x_{+i}}$$

Kaava 1. Cohenin kappa -kertoimen kaava (Brennan & Hays, 1992, s. 155)

Kaavassa 1 N on kokonaismerkintöjen määrä ja x_{ii} on kertojen määrä, jolloin molemmat arvioijat tekivät merkinnän mittarin kategoriaan i . x_{i+} on ensimmäisen arvioijan kaikkien kategoriassa i olevien merkintöjen lukumäärää ja x_{+i} on toisen arvioijan kaikkien kategoriassa i olevien merkintöjen lukumäärää. Tällöin $\sum_i x_{i+} x_{+i}$ on tulos, joka koostuu molempien arvioijien tekemistä merkinnöistä kyseiseen kategoriaan. Toisin sanoen kappa on yhteneväisyyden aste x_{ii} , jossa on otettu huomioon sattuman kautta syntyvä yhteneväisyys. (Brennan & Hays, 1992, s. 155.) Alla *kaavassa 2* sama yhtälö on vielä yksinkertaisemmassa muodossa.

$$\kappa = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)}$$

Kaava 2. Cohenin kappa -kertoimen kaava yksinkertaistettuna (McHugh, 2012, s. 280)

Kaavassa 2 lauseke $\Pr(a)$ on kahden mittaajan yhteneväisten merkintöjen määrä ja lauseke $\Pr(e)$ on satunnaisuudesta johtuvan yhteneväisyyden määrä (McHugh, 2012, s. 280). Kappan ongelmana voidaan nähdä se, miten kerätyn datan lukumäärä ja jakautuminen voivat tuottaa matalan arvon, vaikka arvioijien välillä oleva yhteneväisyys olisi vahva. Datamäärän ollessa pieni ja sen keskittyessä tietylle osalle mittariasteikkoa voi

yksittäisen muuttujan merkitys kappan koolle olla merkittävä. Toisaalta tilanne, jossa arvioijat ovat täysin samaa mieltä, on harvinainen, ja se kertoo usein joko havaintojen homogeenisuudesta tai mittarin validiteetin puutteesta. (Brennan & Hays, 1992, s. 157.) Kappan epäjohdonmukaisuus tuli esiin myös tässä tutkimuksessa, ja asiaa käsitellään myöhemmin tulososiossa ja pohdinnassa.

Aineiston yhteneväisyyttä tutkittaessa kappaa yksinään ei ole aina riittävä menetelmä, jolloin sen rinnalla tulisi käyttää toista aineiston yhteneväisyyden tarkastelun menetelmää. Tämän tutkimuksen tapauksessa kappan lisäksi aineistoa on tarkasteltu ristiintaulukoinnin kautta. Ristiintaulukointi on aineiston kuvailemiseen tarkoitettu menetelmä, jonka kautta voidaan helposti havainnollistaa kahden muuttujan välistä suhdetta sijoittamalla ensimmäisen muuttujan arvot taulukon sarakkeisiin ja toisen muuttujan arvot riveihin (Cohen, Manion, & Morrisson, 2011, s. 623). Muuttujien arvot luovat yhdessä kontingenssitaulukon, josta voidaan nähdä, kuinka monta ensimmäisen muuttujan havaintoa kuuluu samaan luokkaan toisen muuttujan havaintojen kanssa (Nummenmaa, 2004, s. 293). Byman (2018) on luennolla luonnehtinut ristiintaulukointia kvantitatiiviseksi sisällönanalyysiksi ja Metsämuurosen (2011, s. 357) mukaan ristiintaulukointi voi auttaa löytämään mielenkiintoiset yhteydet muuttujien välillä.

Tässä tutkimuksessa kappaa siis kertoo havaintojen yhteneväisyyden asteen, mutta ristiintaulukoinnin kautta voidaan tarkastella, miten merkinnät jakautuvat eri mittareilla ja niiden kategorioilla. Tätä kautta on mahdollista saada monipuolisempi näkökulma tutkimusaiheen tarkasteluun. Painottamaton kappaa tuotti osassa aineistosta erikoisia tuloksia, joten niiden kohdalla kokeiltiin jo mainittujen menetelmien lisäksi myös kappan muita versioita sekä prosentuaalista yhteneväisyyttä. Osalle mittareista laskettiin myös korrelaatiokertoimet, jotka mahdollistivat tulosten vertailun mittariin liittyvien lähdeaineiston tulosten kanssa.

3.4 Validiteetti

Hyvä reliabiliteetti ei yksinään ole tae siitä, että mittari on hyvä. Reliabiliteetti kertoo vain sen, että mittari mittaa luotettavasti jotakin. Validiteetin käsite puolestaan tarkoittaa mittarin kykyä mitata juuri sitä, mitä sen on tarkoituskin mitata (Hirsjärvi, Remes, & Sajavaara, 2009, s. 231). Periaatteessa validiteetin voi määritellä yksinkertaisesti niin, että mittarin validiteetti on mittaustulosten ja mitattavaksi tarkoitettun ominaisuuden todellisen määrän välinen korrelaatio. Asia kuitenkin harvoin on näin yksinkertainen, ja validiteetin mittaaminen voi olla yksi tutkimuksen ongelmallisimmista kohdista. (Karma, 1983, s. 56.)

Validiteetti on laaja käsite, johon liittyy useita eri näkökulmia (ks. Cohen ym., 2011, s. 179; Mislevy, Moss, & Gee, 2009, s. 67; Mason, 2002, s. 188). Johnsonin ja Christensenin (2004) näkemyksen mukaan validiteettia voidaan kvantitatiivisessa tutkimuksessa tarkastella muun muassa tilastollisesta, sisäisestä, ulkoisesta ja käsitteellisestä näkökulmasta. Tilastollisen päättelyn validiteetilla tarkoitetaan asetelmaa, jossa kahden muuttujan välinen yhteys ja sen vahvuus ovat tilastollisesti laskettavissa. Sisäisellä validiteetilla viitataan kahden muuttujan väliseen kausaaliteettiin eli syy-seuraussuhteeseen. Ulkoinen validiteetti viittaa tutkimuksen tulosten yleistettävyyteen tutkimusasetelman ulkopuolella. (Johnson & Christensen, 2004, s. 257–258.)

Käsitevaliditeetti on Cronbachin ja Meelin (1955) tunnetuksi tekemä malli, jossa yksinkertaistettuna validiteettia arvioidaan sen perusteella, kuinka hyvin tutkimuksen käsite operationalisoidaan tutkimuksessa (Metsämuuronen, 2011, s. 75). Käsitevaliditeetti nähdään ehtona muiden validiteettimallien toteutumiselle, ja sen merkitys psykologiselle tutkimukselle on ollut merkittävä (Smith, 2005, s. 396). Kanen (2013, s. 7) mukaan käsitevaliditeetin asema validiteetin arvioinnin standardina on yleisesti tunnustettu. Toisaalta esimerkiksi Pellegrini, Symons ja Hoch (2004, s. 74) toteavat, että etnografiassa tutkimuksen validiteetti voidaan määritellä sen perusteella, kuinka hyvin useat eri datalähteet vahvistavat saman tulkinnan. Argumenttipohjainen validiteettimalli puolestaan tarkastelee validiteettia erilaisten tulkinta- ja validiusargumenttien kautta, ja Kanen (1992, s. 534) mukaan malli ei ole yhtä teoriakeskeinen kuin käsitevaliditeetin malli. Argumenttipohjaista validiteettimallia tarkastellaan enemmän jatkotutkimusosiossa.

Tutkimuksen validiteetti on siis mutkikas käsite, joka sisältää huomattavasti epistemologista ja ontologista pohdintaa. Aihetta olisi mielenkiintoista tutkia enemmän omana kokonaisuutenaan, mutta tämän tutkimuksen kohdalla on tyydyttävä yksinkertaisempaan ratkaisuun.

Tässä tutkimuksessa on päädytty käyttämään validiteetin määritelmänä käsitevaliditeettia sen yleisesti hyväksytyyn asemaan takia, ja koska tämän tutkimuksen tavoite, teorian operationalisoinnin arviointi, on juuri käsitevaliditeetille ominainen näkökulma. Anttila (2006) toteaa, että käsitevaliditeetti kertoo siitä, onko teoreettinen lähestymistapa ja käytetyt käsitteet tutkimuksessa oikeita. Käsitevaliditeetti on erityisen tärkeä näkökulma silloin, kun tarkastellaan tutkimushypoteeseja, joihin on tietyn teorian pohjalta laadittu erilaisia mittareita. (Anttila, 2006 s. 515.) Tässä tapauksessa siis observoinnin käsitevaliditeetin voidaan nähdä perustuvan siihen, miten hyvin mittarit mittaavat niitä asioita, joita niiden tulisi mitata. Ennen kuin voidaan kuitenkaan puhua validiteetista eli tuloksen oikeudesta, on kuitenkin ensin oltava samaa mieltä siitä, mitä mitataan. Kuten aikaisemmin todettiin, ovat validiteetti ja reliabiliteetti toisiinsa liittyviä käsitteitä.

Tämän tutkimuksen aineisto antaa mahdollisuuden tarkastella yhdestä näkökulmasta käsitteiden operationalisointia ja observoinnin validiteettia. Kokonaisvaltaisen käsitevaliditeetin määrittelyyn vaaditaan muutakin kuin observoinnin yhteneväisyyden tarkastelua, eikä tämä tutkimus yksin vastaa kysymykseen hankkeen observoinnin validiteetista tai reliabiliteetista. Siitä huolimatta tämän tutkimuksen syyt ja tavoitteet voidaan nähdä osana hankkeen observoinnin validiteetin tarkastelua Cronbachin ja Meelin (1955) käsitevaliditeetin kontekstissa.

4 Aikaisempi tutkimus

Kehittävä palaute -hankkeen observoinnin yhteneväisyydestä ei ole aikaisempaa julkaistua tutkimustietoa, joten tämä tutkimus voidaan nähdä aihetta kartoittavana. Observointia on käytetty menetelmänä myös *Kehittävää palautetta* edeltäneessä *Orientaatioprojektissa*, mutta tätä tutkimusta varten luetuissa julkaisuissa (ks. Reunamo, 2016; Arvola, Lastikka, & Reunamo, 2017) ei oltu tarkasteltu observoinnin kvantitatiivista yhteneväisyyttä. Lapsen

fyysistä aktiivisuutta ja sitoutuneisuutta tarkastelevat mittarit pohjautuvat kuitenkin CARS- ja LIS-YC -menetelmiin, joiden yhteneväisyydestä löytyy aikaisempaa tutkimustietoa.

Kaksi CARS-menetelmää käyttänyttä tutkimusta ilmoittaa merkintöjen yhteneväisyyden prosentuaalisena yhteneväisyytenä. Kaksi havainnoijaa merkitsi itsenäisesti menetelmän mukaisesti lapsen fyysisen aktiivisuuden asteen, jolloin yhteneväisyyden aste oli yhteneväisten havaintojen määrä kaikkien havaintojen kokonaismäärästä. Tällä menetelmällä ilmoitettu prosentuaalinen yhteneväisyys oli korkea 84,1 % ± 10.1 % (Puhl ym., 1990, s. 29). Toisessa CARS-menetelmää käyttäneessä artikkelissa ilmoitettu prosentuaalinen yhteneväisyys oli myös korkea 88 % (Noland, Danner, Dewalt, McFadden, & Kotchen, 1990, s. 148). Laskemistapa eroaa tässä tutkimuksessa käytettävästä Cohenin kappasta, joten tuloksia ei voi verrata suoraan keskenään. Tarkastelen prosentuaalisen yhteneväisyyden ilmoittamistavan ongelmallisuutta myöhemmin tässä kappaleessa.

Kehittävä palaute -hankkeen sitoutuneisuuden observoinnissa käytetty mittari mukailee Ferre Laeversin LIS-YC-menetelmää, ja hankkeen koulutuksessa viitataan Laeversin ohjeistukseen sitoutuneisuuden havainnoinnista. LIS-YC-menetelmän havaintojen reliabiliteetista puhuttaessa Hautamäki ja Laevers (1997, s. 18) viittaavat tutkimukseen, jossa arvioijien yhteneväisyys LIS-YC-menetelmää käytettäessä oli ilmoitettu muodossa Spearman 0.90. Valitettavasti tutkimusta, johon he viittaavat, ei löytynyt yliopiston tietokannasta. Siten ei saatu tarkempaa selvitystä siihen, miten luku on hankittu. Voidaan kuitenkin olettaa Hautamäen ja Laeversin tarkoittavan Spearmanin Rho -kerrointa, joka Fahertyn (2008, s. 184) mukaan kertoo kahden muuttujan välisen korrelaation voimakkuuden. Kyseessä on siis korrelaatiokerroin, ei yhteneväisyyskerroin.

Ebbeck ym. (2012, s. 612) ilmoittavat tutkimuksessaan LIS-YC-menetelmän havainnoijien välisen prosentuaalisen yhteneväisyyden olleen 90 %. Luku on korkea, mutta ei suoraan vertailukelpoinen kappan kanssa. Myös Kalliala (2011, s. 253) ilmoittaa menetelmän luotettavuuden olleen korkea ”inter-rater correlation 0.93”, jolloin kyseessä on jälleen korrelaatiokerroin eikä yhteneväisyyskerroin. Knoors ja Renting (2000, s. 271) ovat ilmoittaneet LIS-YC-menetelmän yhteneväisyyden vertailukelpoisena kappakertoimena 0.71. Näihin lukuihin palataan tulosten pohdinta -osiossa, jossa niitä verrataan tämän tutkimuksen tuloksiin.

Reliabiliteetin ja validiteetin käsitteiden psykometrasta mittaamista koskevaa tutkimusta on tehty paljon (ks. Banerjee ym., 1999; DeMonbrun, Finelli, & Shekhar, 2015), ja on syytä nostaa esiin muutama tämän tutkimuksen kannalta mielenkiintoinen aihe. Bartko ja Carpenter (1976) käsittelevät artikkelissaan useita reliabiliteetin käsitteeseen liittyviä yleisiä ongelmia psykologisen tutkimuksen kontekstissa. He esittävät yhtenä keskeisenä ongelmana erilaisten tapojen kirjon reliabiliteetin mittaamisessa ja tulosten ilmoittamisessa, kuten yllä mainittujen CARS- ja LIS-YC -menetelmien kohdalla kävi jo ilmi. Monet näistä tavoista (yksinkertainen datan kuvailu ilman siihen liittyvää tilastollista analyysia, yhtenevien havaintojen prosentuaalinen osuus kaikista merkinnöistä, khiin neliö) eivät ota huomioon havainnoijien lukumäärää, havainnoinnissa käytettyjen mittareiden luonnetta tai sattuman vaikutusta merkinnöissä. (Bartko & Carpenter, 1976, s. 308–309.) Tämän tutkimuksen tyyppisessä aineistossa (kaksi havainnoijaa ja nominaaliasteikolliset muuttujat) Bartko ja Carpenter (1976, s. 310) kuitenkin toteavat kappan olevan sopiva menetelmä reliabiliteetin ja yhteneväisyyden mittaamiseen.

Ishimine ja Tayler (2014) huomauttavat, että varhaiskasvatuksen laatua prosessien kautta mittaavien menetelmien tutkimus on ollut vähäistä verrattuna varhaiskasvatuksen laatua rakenteiden ja pedagogisten sisältöjen kautta määrittelevään tutkimukseen. Heidän mukaansa näiden menetelmien laatua ja sopivuutta tutkittaessa tulisi erityisesti kiinnittää huomiota siihen, soveltuvatko menetelmät kansainvälisesti varhaiskasvatuksen ja sen eri muotojen arviointiin. Samalla he toteavat, että menetelmien vahvuuksista ja heikkouksista on sekä teoreettisesti että psykometrisesti edelleen niukasti selkeää, yhtenevää tietoa, mikä tekee vertailusta vaikeaa. (Ishimine & Tayler, 2014, s. 272–273.)

Menetelmien kirjo ja eriävät tavat ilmoittaa tuloksia tekevät aiheen aikaisemman tutkimuksen tarkastelusta haastavaa. Toisaalta vaikka menetelmien keinot ja sisällöt eroavat toisistaan, antavat ilmoitetut reliabiliteettikertoimet yleiskuvan siitä luotettavuudentasosta, jonka havainnointipohjaiset laadunarviointimenetelmät pyrkivät saavuttamaan. Ishimine ja Tayler (2014) arvioivat artikkelissaan yhtätoista varhaiskasvatuksen laadunarviointimenetelmää, jotka käyttävät havainnointia pääasiallisena aineistokeruumenetelmänä. Kaikkien kohdalla on ilmoitettu reliabiliteetti yhden tai useamman eri kertoimen kautta arvojen ollessa yleisesti korkeita välillä 0.80–0.95. Vain yhden menetelmän kohdalla on ilmoitettu painottamaton kappan kolmelle eri osa-alueelle ollen 0.65–0.95. (Ishimine & Tayler, 2014, s. 275–277.)

Fenech, Sweller ja Harrison (2010, s. 283) puolestaan tutkivat kahden (ECERS-R ja ITERS) havainnointipohjaisen laadunarviointimenetelmän soveltuvuutta australialaisen varhaiskasvatuksen laadunarvioinnissa ja sen erojen kuvaamisessa. Molemmissa menetelmissä arvioitiin varhaiskasvatuksen eri osa-alueita (esimerkiksi fyysiset tilat, aktiviteetit ja kanssakäyminen) laatueroasteikollisilla muuttujilla, ja parihavainnointien reliabiliteetiksi ilmoitetaan korrelaatiokertoimet 0.96 ja 0.97 (Fenech ym., 2010, 286–287). Kolmannessa tutkimuksessa Tang ym. (2017) arvioivat luokanopettajien opetuksen laatua ECCOM-menetelmällä Suomessa ja Virossa. Reliabiliteetti kahdessa eri ryhmässä ilmoitetaan korrelaatiokertoimina hyvä 0.60 ja erinomainen 0.86. (Tang ym., 2017, s. 282.)

Lähteiden perusteella näyttäisi siis siltä, että monet varhaiskasvatuksen laatua havainnoinnin kautta arvioivista menetelmistä saavuttavat korkeita yhteneväisyyskertoimia. Tosin tätä tutkimusta varten luetuissa lähteissä vain harva ilmoittaa yhteneväisyyden kapp-kertoimena. Nostan esiin muutamia huomiota liittyen lähteiden menetelmien samankaltaisuuksiin. Ensinnäkin menetelmistä vain harva näyttäisi keskittyvän yksittäisen lapsen havainnointiin ja suurin osa havainnoi koko lapsiryhmää tai luokkahuonetta kokonaisuutena (ks. Ishimine & Tayler, 2014, s. 280). Toiseksi monet menetelmistä ovat olleet olemassa ja käytössä jo monia vuosia (ks. Fenech ym., 2010 s. 286), ja todennäköisesti niiden reliabiliteettia on tarkkailtu ja kehitetty pitkään.

Kolmanneksi monet menetelmistä vaikuttavat laajoilta kokonaisuuksilta (ks. Ishimine & Tayler, 2014, s. 275–277), joiden ymmärtämiseen vaaditaan käyttäjältä paljon työtä. Tämän tuntuvat huomanneen myös menetelmien kehittäjät, sillä monen menetelmän kohdalla havainnoijien koulutus pitää sisällään paljon vaivaa yhteneväisen havainnoinnin saavuttamiseksi. Esimerkiksi laajalti käytössä olevan CLASS-menetelmän tekijät vaativat havainnoijilta tiettyä yhteneväisyyden tasoa havainnoissa ennen kuin he saavat luvan käyttää menetelmää (Ishimine & Tayler, 2014, s. 281). Viimeisenä huomiona voi todeta havainnoinnin mittakaavojen eron. Artikkelien menetelmistä suuressa osassa ellei kaikissa havainnointia tekee pienempi määrä havainnoijia kuin *Kehittävä palaute* -hankkeessa, ja myös havainnoitavia lapsia on vähemmän.

5 Tutkimuskysymykset ja tutkimuksen toteutus

Teoriaosiossa olen esitellyt *Kehittävä palaute* -hankkeen sekä observoinnin osana hankkeen aineistonkeruumenetelmiä. Olen avannut kasvatustieteellisiä malleja hankkeen ja mittareiden taustalla sekä kuvaillut, miten observoijat käytännössä toteuttavat observointia. Olen tarkastellut reliabiliteetin käsitettä sekä Cohenin kappan ja ristiintaulukoinnin sopivuutta tämän tutkimuksen menetelmiksi. Lopuksi olen selvittänyt aikaisempien tutkimusten tapoja ilmoittaa havainnoinnin yhteneväisyys sekä kartoittanut muiden samantyyppisten menetelmien ilmoitettuja yhteneväisyyskertoimia.

Havainnoinnin yhteneväisyys liittyy mittarin luotettavuuteen, joka taas on keskeinen osa minkä tahansa menetelmän reliabiliteetin ja validiteetin muodostumista. Aikaisemman tutkimuksen perusteella havainnointipohjaisten menetelmien luotettavuuskertoimet ovat korkeita mutta vaihtelevasti ilmoitettuja, ja monet näistä havainnointimenetelmistä ovat tarkasti testattuja. Näyttäisi siltä, että luotettavan havainnointiin perustuvan menetelmän rakentaminen on monimutkainen prosessi. Observointi on keskeinen osa *Kehittävä palaute* -hankkeen aineistonkeruumenetelmiä, ja aineisto on poikkeuksellisen laaja. Observoinnin tavoitteena on saada luotettavaa tietoa lasten yksilöllisistä prosesseista ja liittää tämä tieto osaksi isompaa aineistoa. Tämän tiedon hankinnassa käytetään useaa mittaria ja useita havainnoijia. Observoinnin luotettavuudesta ei ole aikaisempaa tutkimustietoa, joten aiheen tutkiminen on tärkeää sekä tieteellisesti että yhteiskunnallisesti. Tämän tutkimuksen tavoitteena oli tarkastella *Kehittävä palaute* -hankkeen observointien luotettavuutta havainnoijien välisen yhteneväisyyden kautta. Tutkimuskysymykset olivat:

1. Miten yhteneväisesti *Kehittävä palaute varhaiskasvatuksessa* -hankkeen observoijat käyttävät observointimittareita?
2. Millaisia eroja yhteneväisyyksissä on parien välillä?
3. Millaisia eroja yhteneväisyyksissä on mittareiden välillä?
4. Miten merkinnät jakautuvat observointimittareiden kategorioihin?

Vastaukset tutkimuskysymyksiin saatiin hankkeen sähköisestä aineistosta, ja pariobservoinnin kautta luotu data sekä kvantitatiiviset tutkimusmenetelmät mahdollistivat observoinnin yhteneväisyyden tarkastelun.

5.1 Tutkimuksen ote

Tutkimuksen aihe syntyi tutkijan omista pohdinnoista liittyen siihen, miten luotettavasti ihminen havainnoi toista ihmistä. Tutkimuskysymykset muodostuivat näiden pohdintojen tuloksena, ja vastaukset kysymyksiin saatiin analysoimalla tutkimusaineistoa. Tutkimus toteutettiin käyttämällä kvantitatiivisia menetelmiä, ja aineistona oli strukturoidun havainnoinnin merkintöjä.

Kvantitatiiviset menetelmät perustuvat mittaamiseen ja mittayksikkö voi vaihdella ilmiön luonteen mukaan. Tavoitteena on mittayksiköiden kautta saada tarkastelun kohteena oleva ilmiö numeeriseen hallintaan. (Erätuuli, Leino, & Yli-Luoma, 1994, s. 10–11.) Tämän tutkimuksen ilmiö, havainnoijien välinen yhteneväisyys, liittyy usein tutkimusasetelmiin, joissa dataa kerätään koulutettujen tai ei-koulutettujen havainnoijien tekemistä arvioista. Yhteneväisyyskertoimen laskemisen kautta on mahdollista saada kvantifioitu tulos kahden havainnoijan yhteneväisyyden asteesta. (Hallgren, 2012, s. 23.)

Tämän tutkimuksen aihealue eli havainnointi on ollut historiallisesti merkittävä osa kvalitatiivista tutkimusperinnettä, sillä se on ollut keskeinen menetelmä antropologeille ja sosiologeille kulttuurintutkimuksen saralla (Silverman, 2001, s. 12). Havainnointi sopii erityisesti toiminnan ja käyttäytymisen kuvaamiseen, ja se on joustavuutensa takia erityisen sovelias menetelmä käytettäväksi dynaamisissa tilanteissa (Uusitalo, 1991, s. 89). Havainnoinnin kautta on mahdollista saada välitöntä ja suoraa tietoa sekä päästä luonnollisiin ympäristöihin. Tässä mielessä voidaan sanoa havainnoinnin olevan todellisen elämän ja maailman tutkimista, sillä se välttää monien muiden menetelmien rasitteena olevan keinotekoisuuden. (Hirsjärvi ym., 2009, s. 213.) Ulkopuolisessa havainnoinnissa havainnoidaan systemaattisella ja jäsenneyllä tavalla etukäteen valittuja piirteitä suhteellisen lyhytaikaisessa tilanteessa (Eskola, 1967, s. 142). *Kehittävä palaute* -hankkeen tapauksessa havainnoinnin strukturointi, koko ja koodaus ovat vieneet sen niin lähelle

määrällistä dataa, että sen analysoinnissa on päädytty käyttämään kvantitatiivisia menetelmiä.

Muuttujien suhteita voidaan tutkia tehokkaasti kokeellisessa tutkimuksessa, jossa tutkija pystyy kontrolloimaan kokeen muuttujia. Ei-kokeellisessa tutkimuksessa tutkijalla ei ole mahdollisuutta vaikuttaa muuttujiin, mikä samalla rajoittaa merkittävästi syy-seuraussuhteen tutkimista. (Johnson & Christensen, 2004, s. 39–40.) Tässä tutkimuksessa on ei-kokeellinen asetelma, sillä sekä tutkittava aihe (lapsen toiminta) että tutkimusmenetelmä (observointi) ovat täynnä muuttujia. Aineisto on annettu tutkijalle, eikä tutkija ole voinut kontrolloida muuttujia tai valvoa aineistonkeräystä.

Tutkimuksen luonne voi Hirsjärven ym. (2007) näkemyksen mukaan olla kartoittava, selittävä, kuvaileva tai ennustava. Kartoittava tutkimus pyrkii löytämään uusia näkökulmia tutkittaviin ilmiöihin, ja selittävä tutkimus tutkii ilmiöiden välisiä suhteita. Kuvailevan tutkimuksen tavoitteena on saada tarkasti dokumentoitua tietoa tutkittavasta ilmiöstä, kun taas ennustava tutkimus pyrkii ennakoimaan, mitä tutkittavasta ilmiöstä seuraa. (Hirsjärvi ym., 2007, 134–135.) Tämän tutkimuksen luonne on kartoittava, sillä tavoitteena on hankkia uutta tietoa siitä, kuinka yhteneväisiä hankkeen observoinnit ovat. Tiedolle ei ole aiempaa vertailukohdetta, joten tämän tutkimuksen tulos voi toimia pohjana laajemmalle jatkotutkimukselle.

5.2 Tutkimuksen kohderyhmä

Kehittävä palaute -hankkeeseen osallistui vuonna 2017 yhteensä 12 kuntaa. Yhteensä kunnissa oli observoijia noin 150. Kaikki *Kehittävä palaute* -observoijat osallistuivat keväällä 2017 kahteen koulutuspäivään. Ensimmäisen koulutuspäivän sisältönä oli hankkeeseen tutustuminen sekä observointien tekemisen ja luokittelun opetteleminen yhdessä. Toisella kerralla keskityttiin observointia ennakoiviin tehtäviin (yhteydenotto ryhmään, lomakkeiden lähettäminen) sekä jatkettiin haastaviksi koettujen mittareiden merkitsemisen harjoittelua. Näiden kahden koulutuspäivän lisäksi observoijat harjoittelivat keväällä 2017 kahtena päivänä parin kanssa observointia jommankumman varhaiskasvatyüksikössä. Hankkeen aikana observoijilla ja koordinaattoreilla oli oma nettisivu, johon observoijat pystyivät lähettämään kysymyksiä ja huomioita observoinnin

aikana esiin nousseista asioista. Lokakuussa 2017 pidettiin yksi koulutuspäivä, jossa käytiin läpi siihen asti tehtyjä observointeja ja esiin nousseita kysymyksiä.

Tämän tutkimuksen aineisto koostui pariobservoinneista, jotka toteutettiin osana hankkeen aineistonkeruuta 11.11.2017–4.12.2017. Tällä aikavälillä pariobservointia toteutti yhdeksän paria seitsemänä päivänä. Aineisto noudatti *Kehittävä palaute* -hankkeen aineistonhankintamenetelmiä, joten yksilöiviä tietoja observoijista (ikä, sukupuoli, työkokemus), päiväkodeista tai lapsista ei kerätty. 18 observoijaa eli 9 paria tekivät yhteensä 822 observointia, joista yhdistettynä muodostui 411 pariobservointia.

Parien valintaprosessin tätä tutkimusta varten suoritti hankkeen observoinneista vastuussa ollut tutkija. Aineiston parit oli arvottu satunnaisesti hankkeessa mukana olleista kymmenestä isoimmasta kunnasta. Kaksi pienintä kuntaa jäi pois arvonnasta, koska niissä oli liian vähän observoijia. Satunnaisesti rakennetusta kuntakohtaisesta observoijalistasta arvottiin kaksi numeroa. Jos kunnassa oli esimerkiksi 14 observoijaa, arvottiin kaksi numeroa 14 numeron joukosta. Pienempi arvottu numero oli ensimmäinen observoija ja hänen ensimmäinen observointikohteensa oli pariobservoinnin päiväkotia. Toinen numero kertoi, kuinka mones observoija listalta tulisi pariobservoinnin toiseksi observoijaksi. Kaikilla observoijilla oli ennen pariobservointia jonkin verran kokemusta syys- ja lokakuun ajalta: yhdellä observoijalla oli takanaan yksi observointikerta, kolmella kaksi kertaa ja lopuilla kolme tai neljä. Jokaisen kunnan hankeyhdyshenkilö oli lähettänyt pariobservointiin arvotuille observoijille ohjeistuskirjeen.

Kehittävä palaute -hankkeelle on saatu Helsingin yliopiston ihmistieteiden eettisen ennakoarvioinnin toimikunnan hyväksyntä ja kuntien tutkimusluvut. Kaikilta hankkeeseen osallistuneiden lasten vanhemmilta on saatu lupa tutkimukseen osallistumiseen. Tämä tutkimus ja sen aineisto noudattavat hankkeen aineistonkeruuperiaatteita, eikä aineistosta ole mahdollista nostaa esiin yksilöiviä lapsen, observoijan tai päiväkodin tietoja.

5.3 Tutkimusaineiston keruu ja analysointimenetelmät

Aineiston analysointiin käytettiin IBM SPSS Statistics -ohjelman versiota 24 Windowsille. Tavoitteena oli reliabiliteettianalyysin kautta saada tietoa havainnointien yhteneväisyydestä. Reliabiliteettianalyysin tavoitteena on Rentzoun (2015) mukaan tuottaa tilastollinen indeksi tai luotettavuuskerroin käytetyn testin tai mittarin luotettavuudesta. Indeksillä viitataan mittarin, testin tai minkä tahansa mittausten menetelmän kykyyn tuottaa sama tulos toistuvasti. (Renzou, 2015, s. 4.)

Analysoinnin alussa aineistona oli 853 observointitilannetta, jotka oli siirretty hankkeen salasanasuojatusta pilvestä salasanasuojattuun excel-tilaukkuun. Ensimmäiseksi aineisto käytiin läpi manuaalisesti siten, että aineistosta poistettiin parittomat ja virheelliset (esimerkiksi väärä observoija- tai päiväkotikoodi) merkinnät sekä tutkimuksen kannalta epäoleelliset sarakkeet (esimerkiksi observoinnin aloitusaika ja observoinnin tunnistenumero). Sellaiset merkinnät, joissa selvästi oli kyse yksittäisestä lyöntivirheestä (numerosarjasta puuttui yksi numero) ja joiden oikea pari tai koodi pystyttiin luotettavasti päättämään (muut tunnistetiedot olivat identtiset), korjattiin parin löytymiseksi. Lopulta virheellisiä tai parittomia merkintöjä jäi yhteensä 31 kappaletta eli 3,6 % kaikista merkinnöistä. Nämä merkinnät poistettiin taulukosta.

Tämän jälkeen loput merkinnät siirrettiin excel-tilaukon samalle riville manuaalisesti observoijien yhteisen päiväkotikoodin, observoinnin lopettamisajan ja lapsen tunnistenumeron perusteella. Mukana siirrettiin aina myös observoijan oma koodi, jolloin saman observoijan kaikki merkinnät pysyivät helposti tarkasteltavissa omalla rivillään. Lopuksi aineisto tarkistettiin sort by -komennolla eri muuttujien perusteella ja käymällä lista uudestaan läpi poikkeuksien tai puuttuvien merkintöjen osalta. Aineiston luonteesta ja suljetuista vastausvaihtoehdoista sekä alustavasta läpikäynnistä johtuen puuttuvia arvoja ei enää tässä vaiheessa ollut.

Näiden toimenpiteiden jälkeen SPSS-ohjelman havaintomatriisiin siirrettiin lopulta 822 yksittäistä observointitapahtumaa eli 411 pariobservointia. Pariobservointien määrä vaihteli pareittain välillä 35–56 ja suurimmalla osalla pareista observointeja oli 44–53. Aineisto siirrettiin excel-tilaukosta SPSS-ohjelmaan yksi sarake kerrallaan. Jokaiselle

sarakkeelle annettiin nimi ja sarakkeen arvot määritettiin observointimittareiden kategorioiden mukaisesti. Lopuksi matriisiin lisättiin sarake, johon merkittiin observoinnin tehneen parin numero. Tätä tarvittiin myöhemmin parikohtaisten tulosten laskemisessa. Lopulta sarakkeita oli yhteensä 25: observointiaika, lasten numerot, parien observointinumerot, parinnumero sekä parien merkinnät kaikilla mittareilla. Jokainen observointitapahtuma piti sisällään 12 merkintää, jolloin raakamatriisin kaikkien alkioden kokonaismäärä oli 10275.

Tutkimuskysymyksiin saatiin vastaus ristiintaulukoinnin ja Cohenin kappa -kertoimen avulla. Ensin aineistosta laskettiin yksitellen jokaisen mittarin kohdalla jokaiselle parille oma kappa, jolloin saatiin jokaisen parin yhteneväisyyden aste jokaisella mittarilla (taulukko 2). Tavoitteena oli tarkastella hankkeen observointien yhteneväisyyttä kokonaisuutena, joten taulukossa 2 esitetyt parien ja mittarien kertoimet lisättiin omiksi sarakkeiksi matriisiin, mikä mahdollisti keskiarvon, keskivirheen ja luottamusvälin laskemisen. Tätä kautta saatiin yhteneväisyyden keskiarvot jokaiselle parille (taulukko 3) ja jokaiselle mittarille (taulukko 4). Tutkimuksessa haluttiin myös tarkastella, miten yhteneväiset ja ei-yhteneväiset observoinnit jakautuvat aineistossa. Tämä tehtiin ristiintaulukoinnin kautta mittari kerrallaan (taulukko 6–taulukko 14). Näiden tulosten analysoinnin kautta pystyttiin saamaan käsitys sekä yksittäisten parien, yksittäisten mittareiden sekä koko aineiston yhteneväisyydestä. Tulokset myös havainnollistivat Cohenin kappan vahvuuksia ja heikkouksia tämän tyyppisen aineiston analysoinnissa.

6 Tulokset

Cohenin kappan ja ristiintaulukoinnin kautta saadut tulokset käydään läpi osio kerrallaan. *Tutkimuskysymyksiä 1–3* vastaukset löytyvät luvusta 6.1 ja *tutkimuskysymyksen 4* vastaukset luvusta 6.2. Osassa taulukoista on paljon dataa, mutta taulukot itsessään ovat yksinkertaisia lukea.

6.1 Observoinnin yhteneväisyys

Tutkimuskysymyksiin 1–3 saatiin vastaus Cohenin kappan avulla. Taulukossa 1 on nähtävissä Landisin ja Kochin (1977, s. 165) määritelmä siitä, miten kappa tulisi tulkita merkintöjen yhteneväisyyden tunnuslukuna. Banerjee ym. (1999, s. 6) ovat esittäneet oman, hieman laveamman määritelmän Landisin ja Kochin taulukkoon perustuen, ja näitä molempia tullaan käyttämään viitattaessa yhteneväisyyden asteeseen.

Taulukko 1

Kappa ja yhteneväisyyden aste

Kappa	Yhteneväisyyden aste
<0.00	Huono
0.00–0.20	Heikko
0.21–0.40	Kohtalainen
0.41–0.60	Keskinkertainen
0.61–0.80	Huomattava
0.81–1.00	Lähes täydellinen

Huom. Landis & Koch, 1977, s. 165 (suomennokset Siren, 2008, s. 11).

Taulukossa 1 on Landisin ja Kochin (1977, s. 165) määritelmät Cohenin kappalle ja yhteneväisyyden asteelle: <0.00 huono, 0.00–0.20 heikko, 0.21–0.40 kohtalainen, 0.41–0.60 keskinkertainen, 0.61–0.80 huomattava ja 0.81–1.00 lähes täydellinen. Banerjee ym.

(1999, s. 6) toteavat samaan taulukkoon perustuen, että useimmissa tapauksissa kappan ollessa yli 0.75 voidaan arvioitsijoiden välistä yhteneväisyyttä pitää erittäin hyvänä, kappan ollessa välillä 0.40–0.75 yhteneväisyyttä voidaan pitää kohtalaisena–hyvänä ja kappan ollessa alle 0.40 yhteneväisyyttä voidaan pitää heikkona. McHugh (2012, s. 279) on määrittelyssään muita ankarampi, todeten kappan 0.60 olevan hyväksyttävän yhteneväisyyden alaraja. Graham, Milanowski ja Miller (2012, s. 9) määrittelevät samantyyppisesti hyväksyttävän yhteneväisyyden alarajaksi kappan 0.61.

Kappan raja-arvot ja niihin liittyvät kuvaukset eivät siis ole määriteltyjä standardeja, ja se mikä on keskinkertainen, heikko tai tarpeeksi hyvä tulos on riippuvainen siitä, mitä mitataan (Brennan & Hays, 1992, s. 158; Hallgren, 2012, s. 27). Tämän tutkimuksen tulososiossa on päädytty kuvailemaan tuloksia Landisin ja Kochin (1977) taulukon mukaisesti sen tunnustetun (ks. McHugh, 2012; Siren, 2008; Sim & Wright, 2005) yleisyyden vuoksi. Samaten aikasempien tutkimusten (ks. McHugh, 2012; Graham ym., 2012) perusteella myös tässä tutkimuksessa kappan ≥ 0.60 tulkitaan hyväksyttävän yhteneväisyyden alarajana. Tulosten pohdinta -osiossa tuloksia tarkastellaan myös Landisin ja Kochin (1977) asteikkoa kriittisemmin McHugh'n (2012) määritelmien mukaisesti. Muissa tutkimuksissa kappan on usein ilmoitettu muodossa 0.xx mutta tässä tutkimuksessa nolla on jätetty pois APA:n ohjeiden mukaisesti (ks. Paiz ym., 2018).

Tämän tutkimuksen tapauksessa yksi yhteinen luku yhteneväisyydestä jättäisi paljon tietoa huomiotta, sillä on todennäköistä, että mittareiden ja parien välillä on edes jonkinlaista eroa. Tutkimuskysymysten kannalta on siis keskeistä käydä tulokset läpi vaihe kerrallaan. Ensiksi tarkastellaan kaikkien mittareiden kappan-kertoimia parikohtaisesti. Tätä kautta on mahdollista vertailla parien välisiä eroja ja myöhemmin verrata parien mittarikohtaisia kertoimia parien keskiarvojen kanssa. Tämän jälkeen tarkastellaan jokaisen mittarin keskiarvoa, joka kertoo tiivistettynä sen, miten yhteneväisesti mittaria on käytetty tässä aineistossa. Lopuksi tarkastellaan ristiintaulukoinnin kautta observointien jakautumista jokaisella mittarilla ja valotetaan mahdollisia syitä kertoimien koolle.

Tilastollista merkitsevyyttä ei ole yleensä merkitty kappan raportoinnissa. Tilastollinen merkitsevyys on vahvasti sidoksissa yhteneväisyyden asteeseen ja otoskoko, jolloin tarpeeksi isolla otoskolla ja jonkinasteisella yhteneväisyydellä tulos on aina tilastollisesti merkitsevä. (Vierra & Garret, 2005, s. 362; Sim & Wright, 2005, s. 265.) Merkitsevyydet

on jätetty pois myös tässä tutkimuksessa, mutta melkein kaikki kappa-kertoimet olivat tilastollisesti erittäin merkitseviä ($p < 0.001$) ja muutama selvästi matala kerroin ei ollut tilastollisesti merkitsevä ($p > 0.05$).

6.1.1 Parien kappa kaikilla mittareilla

Taulukossa 2 nähdään jokaisen parin kappa jokaisella observointimittarilla. P1 on pari numero 1, P2 on pari numero 2 ja niin edespäin. Sarakkeet ovat yhden parin kappa-kertoimia eri mittareilla ja rivit ovat yhden mittarin kappa-kertoimia eri pareilla.

Taulukko 2

Kappa pareittain kaikilla mittareilla

Mittari	P1	P2	P3	P4	P5	P6	P7	P8	P9
A. Yleinen toiminta	.86	.85	.79	.90	.51	.98	1	.92	.97
B. Lapsen oma toiminta	.76	.78	.45	.65	.58	.59	.38	.72	.72
C. Lapsen huomionkohde	.59	.56	.45	.68	.53	.55	.44	.65	.56
D. Lähin lapsikontakti	.75	1	1	-.01 ^a	.45	.68	.67	.50	.80
E. Lapsen fyysinen aktiiv.	.67	.48	.14	.52	.58	.58	.43	.54	.67
F. Lapsen sitoutuneisuus	.30	.27	.15	.53	.18	.04 ^b	.30	.13	.31
G. Lapsen emootio	.34	.30	.18	.49	.44	.34	.19	.35	.45
H. Lapsen sos. orientaatio	.47	.53	.29	.56	.19	.28	.26	.35	.35
I. Lähimmän aik. toiminta	.65	.72	.11	.57	.23	.45	.43	.58	.40

Huom. Solun luku on kappa.^a Yleisyysongelma (ks. luku 8), ei mukana laskuissa.^b Reunafrekvenssiongelma (ks. luku 8), ei mukana laskuissa.

Taulukossa 2 A. *Yleinen toiminta* -mittarin yhteneväisyys on seitsemällä parilla lähes täydellinen, joista parilla numero 7 yhteneväisyys on täydellinen. Parilla numero 3 yhteneväisyys on huomattava ja parilla numero 5 yhteneväisyys on keskinkertainen. B. *Lapsen oma toiminta* -mittarilla viiden parin yhteneväisyys on huomattava, kolmella keskinkertainen ja yhdellä kohtalainen. Mittarilla C. *Lapsen huomion kohde* yhteneväisyys on kahdella parilla huomattava ja seitsemällä keskinkertainen. D. *Lähin lapsikontakti* -

mittarilla kaksi paria on merkinnyt tilanteet täysin yhteneväisesti, neljällä parilla yhteneväisyys on huomattava ja kahdella keskinkertainen.

Parin numero 4 kohdalla kappa on mystisesti ≈ 0.1 . Kyseinen pari ei ole ollut täysin eri mieltä kaikista merkinnöistä, kuten kappa antaisi ymmärtää. Päinvastoin parin yhteneväisyys oli melkein täydellinen: merkinnöistä yhteneväisiä oli 51/53 ja prosentuaalinen yhteneväisyys oli 98 %. Kyseisen parin kohdalla näkyy kappan niin sanottu yleisyysongelma. Toisen kappan kohdalla taulukossa 2 nousi esiin kappan niin sanottu reunafrekvenssiongelma. Molempia aiheita käsitellään tarkemmin luvussa 8. Tässä vaiheessa todetaan kuitenkin, että ongelmat olivat merkittäviä vain näiden kahden kappan kohdalla ja ne jätettiin pois keskiarvolaskuista.

E. Lapsen fyysinen aktiivisuus -mittarilla kahden parin yhteneväisyys on huomattava, kuudella keskinkertainen ja yhdellä heikko. *F. Lapsen sitoutuneisuus* -mittarilla yhden parin yhteneväisyys on keskinkertainen ja neljällä parilla kohtalainen. Kolmella parilla yhteneväisyyden voidaan sanoa olevan heikko. Mittarilla *G. Lapsen emootio* kolmella parilla yhteneväisyys on keskinkertainen, neljällä kohtalainen ja kahdella heikko. Mittarilla *H. Lapsen sosiaalinen orientaatio* kolmella parilla yhteneväisyys on keskinkertainen, viidellä kohtalainen ja yhdellä heikko. Lopuksi mittarilla *I. Lähimmän aikuisen toiminta* kahden parin yhteneväisyys on huomattava, neljällä keskinkertainen, kahdella kohtalainen ja yhdellä heikko.

Hyväksyttävä yhteneväisyyden raja $.60$ näyttäisi ylittyvän useasti tietyillä pareilla (*I*, *2* ja *9*) ja tietyillä mittareilla (*A*, *B* ja *D*). Monen parin kohdalla kappa kuitenkin myös alittaa hyväksyttävän rajan useaan otteeseen. Joillain mittareilla taas näyttäisi siltä, että kappan hyväksyttävä raja ei ylity lainkaan. Taulukkoa 2 tarkastelemalla huomataan, että mittareiden ja parien yhteneväisyydessä on siis selviä eroja. Mittareilla *A*, *B*, *C* ja *D* näyttäisi siltä, että parien kappa-kertoimet ovat yleisesti korkeampia ja toisaalta mittareilla *F*, *G* ja *H* parien kertoimet ovat matalampia. Samaten näyttäisi siltä, että esimerkiksi parien *1* ja *9* kertoimet ovat korkeita, ja parien *3* ja *5* matalia. Pari- ja mittarikohtaiset kertoimet taulukossa 2 ovat hyvää vertailuaineistoa siirryttäessä tarkastelemaan molempien keskiarvoja.

6.1.2 Parien kappan keskiarvo

Taulukossa 3 on parien kappan keskiarvo, kun kyseisen parin kaikkien mittarien kappa-kertoimet on laskettu yhteen. Suositusten mukaisesti (ks. McHugh, 2012, s. 281, Viera & Garrett, 2005, s. 362) jokaisen parin kohdalla on ilmoitettu keskivirhe ja 95 % luottamusvälin ala- ja yläraja. Kappan keskiarvo, keskivirhe ja luottamusvälin ylä- ja alarajat on hankittu SPSS-ohjelman Explore-toiminnolla.

Taulukko 3

Parien kappan keskiarvo

Pari	Kappa	Keskivirhe	95 % Luottamusväli	
1	.60	.064	.450	.756
2	.61	.082	.419	.797
3	.43	.113	.170	.691
4	.61	.048	.499	.725
5	.41	.055	.282	.537
6	.55	.076	.376	.737
7	.46	.082	.266	.644
8	.52	.077	.345	.702
9	.58	.075	.407	.754

Landisin ja Kochin (1977) määrittelyn mukaisesti arvioituna yhteneväisyys on seitsemällä parilla keskiverto ja kahdella parilla yhteneväisyys on juuri rajalla huomattava.

Mielenkiintoisesti aikaisemman taulukon 2 erot näyttävät siis tasaantuneen tarkastellessa parien keskiarvoja. Vaikka suurin osa kappan arvoista sopii Landisin ja Kochin (1977) määrittelyssä samalle alueelle, on niiden välillä silti selviä eroja jopa niin, että kaksi paria ovat keskiarvoiltaan tasan ääripäissä asteikon aluetta .41–.60. Neljän parin kerroin ylittää tai on lähellä ylittää hyväksyttävän yhteneväisyyden rajan .60. Tulen kiinnittämään

enemmän huomiota parien keskivirheiden ja luottamusvälin ylä- ja alarajojen kokoihin tulosten pohdinta -osiossa.

6.1.3 Mittarien kappan keskiarvo

Alla taulukossa 4 on jokaisen mittarin keskiarvo, kun jokaisen parin kyseisen mittarin kappa-kertoimet on laskettu yhteen. Kuten pareille, myös mittareille on ilmoitettu keskivirhe ja luottamusvälin ala- ja yläraja.

Taulukko 4

Mittarien kappan keskiarvo

Mittari	Kappa	Keski- virhe	95 % Luottamusväli	
A. Yleinen toiminta	.86	.050	.749	.977
B. Lapsen oma toiminta	.63	.046	.518	.730
C. Lapsen pääasiallinen huomion kohde	.56	.026	.495	.616
D. Lähin sosiaalinen lapsikontakti	.73	.072	.560	.900
E. Lapsen fyysinen aktiivisuus	.51	.054	.390	.636
F. Lapsen sitoutuneisuus	.27	.045	.164	.377
G. Lapsen emootio	.34	.036	.259	.423
H. Sosiaalinen orientaatio	.36	.043	.262	.464
I. Lähimmän aikuisen toiminta	.46	.066	.307	.609

Taulukossa 4 mittarin *A. Yleinen toiminta* kappa on .86 ja yhteneväisyys voidaan nähdä lähes täydellisenä Landisia ja Kochia (1977) määritelmiä mukailien. Mittarilla *B. Lapsen oma toiminta* kappa on .63 ja yhteneväisyys voidaan nähdä huomattavana. *C. Lapsen pääasiallinen huomionkohde* -mittarin kappa on .56 ja yhteneväisyys voidaan nähdä

keskinkertaisena. Mittarin *D. Lähin sosiaalinen lapsikontakti* kappaa on .73 ja yhteneväisyys on huomattava, ja *E. Lapsen fyysinen aktiivisuus* -mittarin kappaa .51 on yhteneväisyydeltään keskinkertainen. Mittarin *F. Lapsen sitoutuneisuus* kappaa on .27 ja yhteneväisyys on kohtalainen. Samaten mittarin *G. Lapsen emootio* kappaa .33 voidaan nähdä kohtalaisena yhteneväisyytenä. *H. Sosiaalinen orientaatio* -mittarin kappaa on .36, ja yhteneväisyys on myös kohtalainen. *I. Lähimmän aikuisen toiminta* -mittarin kappaa on .46, jolloin yhteneväisyys voidaan nähdä keskinkertaisena.

Taulukossa 4 mittari *A* on ainoa, jonka kappaa saavuttaa Landisin ja Kochin (1977) asteikon ylimmän alueen eli lähes täydellisen yhteneväisyyden. Mittareilla *B, C* ja *D* yhteneväisyys on välillä huomattava–keskinkertainen, ja mittareilla *E, F, G, H* ja *I* keskinkertainen–kohtalainen. Mittareiden välillä on selviä eroja, alhaisimman kappan ollessa .27 ja korkeimman kappan ollessa .86. Hyväksyttävän yhteneväisyyden raja .60 ylittyy mittareilla *A, B* ja *D*.

Tutkimuskysymyksiin 1–3 saatiin vastaus Cohenin kappan avulla. Taulukossa 2 parien mittarikohtaisissa yhteneväisyyksissä on selkeitä eroja korkeimman mittarin parienvälisen eron ollessa .45–1 ja parin mittarinvälisen eron ollessa .11–1. Observointiparien yhteneväisyyden keskiarvo taulukossa 3 on Landisin ja Kochin (1977) asteikolla keskinkertainen .41–.60. Hyväksyttävän yhteneväisyyden raja .60 ylittyy tai on hyvin lähellä ylittyä neljän parin kohdalla.

Taulukosta 4 nähdään, että yhteneväisyys on korkeampi niillä mittareilla, joiden kautta havainnoidaan lapsen ympäristöä (*A* ja *B*) ja lapsen huomion kohdetta (*C* ja *D*). Yhteneväisyys näillä mittareilla on pääosin .60–.80. eli Landisin ja Kochin (1977) mukaan huomattava. Yhteneväisyys on matalampi mittareilla, joiden kautta havainnoidaan lapsen olotilaa (*E, F, G, H*) ja aikuisen roolia (*I*). Yhteneväisyys näillä mittareilla on keskinkertainen .41–.60 tai kohtalainen .21–.40. Hyväksyttävän yhteneväisyyden raja .60 ylittyy mittareilla *A. Yleinen toiminta, B. Lapsen oma toiminta* ja *D. Lähin sosiaalinen lapsikontakti*.

Kappa-kertoimien kautta saatiin yksi näkökulman observointien yhteneväisyyteen. Kappa ei kuitenkaan kerro meille, miten merkinnät jakautuvat mittareiden kategorioihin, tai anna lisätietoa siitä, miksi kerroin on korkea tai matala. Mittareiden sisällöt ovat kovin erilaisia,

ja niitä tulee tarkastella yksityiskohtaisemmin, jotta saisi käsityksen yhteneväisyyden mahdollisista syistä. Lopulta ainoastaan tilanteessa itse olleet observoijat tietävät, millaisissa tilanteissa merkinnät on tehty, emmekä voi varmuudella tietää syitä tilanteiden yhteneväisyydelle tai eriävyydelle. Voimme kuitenkin tarkastella yhteneväisyyteen vaikuttavien merkintöjen jakautumista ristiintaulukoinnin avulla.

6.2 Merkintöjen jakautuminen

Tutkimuskysymykseen 4 saatiin vastaus ristiintaulukoinnin kautta. Vastaus ei kuitenkaan ole yhtä helposti tiivistettävissä kuin tutkimuskysymykset 1–3, ja tulokset on jaettu tämän luvun alaotsikoihin. Yhteenvedo mittareiden ja niiden kategorioiden toimivuudesta sekä yhteneväisyyteen keskeisesti vaikuttavista tekijöistä löytyy luvusta 7.2.10.

Mittareiden ristiintaulukoinneissa parien merkinnät on jaettu riveihin ja sarakkeisiin. Yhteneväisesti arvioitujen merkintöjen määrä on kyseisen kategoriarivin ja -sarakkeen risteyskohdassa, ja muut merkinnät rivin soluissa kuvaavat ei-yhteneväisiä merkintöjä. Jokaisessa solussa on merkitty solun prosentuaalinen osuus kaikista rivin merkinnöistä ja solufrekvenssi. Viiva tarkoittaa, ettei solulle ole tehty merkintöjä. Taulukon viimeinen sarake sisältää riviprosentin r % ja rivin reunajakauman f_i . Soluista on jätetty pois prosentti- ja frekvenssisymbolit lukukokemuksen helpottamiseksi.

Taulukoissa 2, 3 ja 4 luvut olivat kertoimia ja seuraavissa taulukoissa luvut ovat prosentteja ja lukumääriä. Luvut eivät siis ole suoraan vertailukelpoisia keskenään eikä Landisin ja Kochin (1977) määritelmät taulukossa 1 ole yhteensopivia ristiintaulukoinnin lukujen kanssa. Ristiintaulukointiin voidaan kuitenkin soveltaa prosentuaalisen yhteneväisyyden kuvaamiseen liittyviä termejä. Tosin tämäkin tapa on väärä siinä mielessä, että toisin kuin prosentuaalisen yhteneväisyyden tapauksessa ristiintaulukoinnin lukuja ei ole jaettu kokonaismerkintöjen määrällä. Luvun kuvailemisessa tulisi ottaa huomioon myös kyseessä oleva mittari, kategoria, solufrekvenssi ja reunajakauma. Näiden kaikkien huomioonottaminen olisi kuitenkin erittäin aikaavievää ja veisi huomiota pois siitä, että tutkimuksen tavoitteena on tarkastella observoinnin yhteneväisyyttä kokonaisuutena.

Lopulta tässä tutkimuksessa päädyttiin käyttämään prosentuaalisen yhteneväisyyden termejä, sillä Landisin ja Kochin (1977) kappaan liittyvä määrittely koettiin ristiriitaiseksi erityisesti alhaisten prosenttilukujen kohdalla, ja koska kappa ei ole prosenttiluku.

Prosentuaalisen yhteneväisyyden kuvaamiseen ei kuitenkaan löytynyt yhtä yksityiskohtaista asteikkoa kuin Landisin ja Kochin kappaan liittyvä asteikko. Lähteet (Shweta, Bajpai, & Chaturvedi, 2015, s. 26; Stemler, 1988, s. 2; Graham ym., 2012, s. 9) kuitenkin määrittelevät yleisesti hyväksyttävän prosentuaalisen yhteneväisyyden rajaksi tutkimuksessa 75–90 %. Tähän perustuen myös tässä tutkimuksessa yli 75 % voidaan tulkita hyvänä tuloksena, ja jos luku on alle 75 % käytän muita termejä kuten kohtalainen, heikko tai huono oman harkinnan mukaan. Alla oleva taulukko 5 havainnollistaa ristiintaulukointitulosten tulkitsemista.

Taulukko 5

Esimerkki observointien ristiintaulukoinnista

	Opetus	Tuettu leikki	Vapaa leikki	Ulkoilu	Perushoito	Ruokailu	Lepohetki	r % f_i
Opetus	85,5 (47)	3,6 (2)	1,8 (1)	-	5,5 (3)	3,6 (2)	-	100 (55)
Tuettu leikki	11,1 (2)	50 (9)	33,3 (6)	-	-	5,6 (1)	-	100 (18)

Huom. Solun ylempi luku on % ja alempi luku on f .

Esimerkkinä toimivassa taulukossa 5 *tuettu leikki* -sarakkeen ja *tuettu leikki* -rivin risteämiskohdassa on luku 50 (9). Ensimmäinen luku kuvaa merkintöjen prosentuaalista osuutta sen rivin kaikista merkinnöistä, ja suluissa oleva luku kuvaa merkintöjen lukumäärää. Eli kaikista mittarin merkinnöistä, joissa jompikumpi observoijista on merkinnyt tilanteen tuetuksi leikiksi, on 50 %:ssa kerroista myös toinen observoija merkinnyt tilanteen tuetuksi leikiksi. Kahdessa tilanteessa (11,1 %) toinen observoijista on nähnyt tilanteen tuetun leikin sijaan opetuksena, kuudessa (33,3 %) vapaana leikkinä ja niin edespäin. Eli tuloksissa ilmotettaisiin, että *tuettu leikki* (50 %, $f = 9$) -kategorian yhteneväisyys on huono, riviprocentti on 100 % ja reunajakauma on 55. Helpointa on aina

valita ensin riviltä kategoria, etsiä vastaava sarake, vertailla niitä keskenään ja katsoa, mihin eriävät merkinnät kyseisellä rivillä jakautuvat.

Kaikki mittarit läpi käydään samalla logiikalla ja jokaisesta mittarista nostetaan esiin sen keskeiset luvut. Tulosten analysointiin ja pohdintaan siirrytään tulososion jälkeen.

Informatiivisinta (ja mielenkiintoisinta) olisi ollut tehdä ristiintaulukointi jokaiselle parille erikseen, mutta tämä olisi valitettavasti lisännyt sivumäärää huomattavasti.

Ristiintaulukoinnissa on siis päädytty kuvaamaan kaikkia aineiston merkintöjä kootusti kyseisellä mittarilla.

6.2.1 A. Yleinen toiminta

Ensimmäisenä on ristiintaulukointi mittarista *A. Yleinen toiminta*. Taulukossa 4 mittarin *A* kappi oli .86, 95 % CI [.749, .977]. Landisin ja Kochin (1977) mukaan mittarin yhteneväisyys on lähes täydellinen ja Banerjeen ym. (1999) mukaan yhteneväisyys on hyvä.

Taulukko 6

Ristiintaulukointi A. Yleinen toiminta

	Opetus	Tuettu leikki	Vapaa leikki	Ulkoilu	Perushoito	Ruokailu	Lepohetki	<i>r % fi</i>
Opetus	85,5 (47)	3,6 (2)	1,8 (1)	-	5,5 (3)	3,6 (2)	-	100 (55)
Tuettu leikki	11,1 (2)	50 (9)	33,3 (6)	-	-	5,6 (1)	-	100 (18)
Vapaa leikki	1,4 (1)	21,4 (15)	74,3 (52)	-	1,4 (1)	1,4 (1)	-	100 (70)
Ulkoilu	-	-	-	100 (102)	-	-	-	100 (102)
Perushoito	1,4 (1)	5,8 (4)	2,9 (2)	-	88,4 (61)	1,4 (1)	-	100 (69)
Ruokailu	-	1,1 (1)	-	-	3,4 (3)	95,5 (85)	-	100 (89)
Lepohetki	-	-	-	-	-	-	100 (8)	100 (8)

Huom. Solun ylempi luku on % ja alempi luku on *f*.

Taulukossa 6 näkyy observointien ristiintaulukointi mittarilla *A. Yleinen toiminta*.

Opetustilanteet (85,5 %, *f* = 47) on merkitty hyvin yhteneväisesti ja eriävät merkinnät hajaantuvat tasaisesti tuetun leikin, vapaan leikin, perushoidon ja ruokailun välillä. *Tuetun leikin* (50 %, *f* = 9) merkintöjen yhteneväisyys on heikko, ja eriävät merkinnät jakautuvat vapaan leikin, opetuksen ja ruokailun kesken. *Vapaa leikki* (74,3 %, *f* = 52) on merkitty yhteneväisesti, ja suurin osa eriävistä merkinnöistä on tuettu leikki -kategoriassa sekä yksittäisinä opetus-, perushoito- ja ruokailumerkintöinä. *Ulkoilun* (100 %, *f* = 102) merkitsemisessä observoijat ovat olleet yhteneväisesti sitä mieltä, että lapset todella ovat ulkona.

Perushoitotilanteet (88,4 %, *f* = 61) on merkitty yhteneväisesti ja loput merkinnät jakaantuvat tuetun leikin, vapaan leikin, opetuksen ja ruokailun välille. *Ruokailutilanteet* (95,5 %, *f* = 85) on merkitty lähes täysin yhteneväisesti ja vain muutama merkintä on

eriävästi perushoito- tai tuettu leikki -kategoriassa. *Lepohetket* (100 %, $f = 8$) on merkitty ulkoilun tavoin selkeästi omana sisältönään.

Taulukko 6 osoittaa, että eriävien merkintöjen jakautuminen on vähäistä mittarilla A, ja suurin eriävien merkintöjen määrä on tuetun leikin ja vapaan leikin välillä. Tulosten pohdintaosiossa nostetaan esiin mahdollisia syitä mittarin yhteneväisyyden vahvuudelle sekä tuetun ja vapaan leikin sekoittumiselle.

6.2.2 B. Lapsen oma toiminta

Seuraavaksi tarkastellaan mittaria *B. Lapsen oma toiminta*. Taulukossa 4 *B. Lapsen toiminta* -mittarin kappia oli .63, 95 % CI [.575, .681]. Yhteneväisyys on Landisin ja Kochin (1977) mukaan keskinkertainen ja Banerjeen ym. (1999) mukaan kohtalainen. Mittari on sisällöltään kaikkein laajin, joten merkintöjen hajonta on myös muihin mittareihin verrattuna laajempaa. Taulukko on iso ja sen lukeminen voi tuntua hankalalta, mutta taulukkoa seuraava teksti selventää sen sisältöjä.

Taulukko 7 Ristiintaulukointi B. Lapsen oma toiminta

	Rooli- leikki	Sääntö- leikki	Esine- leikki	Fyysinen leikki	Luke- minen	Tehtävä	Toiminta il. fokusta	Yhdessä- olo	Kielletty toiminta	Yleinen kehys	Muu toiminta	$r\%$ f_i
Rooli- leikki	40 (12)	-	53,3 (16)	3,3 (1)	-	-	-	3,3 (1)	-	-	-	100 (30)
Sääntö- leikki	-	60 (3)	-	20 (1)	-	-	-	-	-	-	20 (1)	100 (5)
Esine- leikki	5,2 (3)	3,4 (2)	60,3 (35)	1,7 (1)	1,7 (1)	12,1 (7)	1,7 (1)	3,4 (2)	-	1,7 (1)	8,6 (5)	100 (58)
Fyysinen leikki	2,3 (1)	-	6,8 (3)	75 (33)	-	-	11,4 (5)	2,3 (1)	-	-	2,3 (1)	100 (44)
Luke- minen	-	-	-	-	75 (3)	-	-	-	25 (1)	-	-	100 (4)
Tehtävä	-	10 (3)	-	6,7 (2)	3,3 (1)	73,3 (22)	-	-	-	6,7 (2)	-	100 (30)
Toiminta il.fokusta	-	2,1 (1)	-	-	-	2,1 (1)	59,6 (28)	10,6 (5)	-	23,4 (11)	2,1 (1)	100 (47)
Yhdessä- säolo	3,6 (1)	-	3,6 (1)	17,9 (5)	-	-	7,1 (2)	53,6 (15)	3,6 (1)	7,1 (2)	3,6 (1)	100 (28)
Kielletty toiminta	-	-	-	-	-	-	11,1 (1)	11,1 (1)	66,7 (6)	11,1 (1)	-	100 (9)
Yleinen kehys	-	-	0,7 (1)	1,4 (2)	0,7 (1)	0,7 (1)	4,2 (6)	-	2,1 (3)	88,1 (126)	2,1 (3)	100 (143)
Muu toiminta	-	-	15,4 (2)	-	-	-	30,8 (4)	7,7 (1)	-	23,1 (3)	23,1 (3)	100 (13)

Huom. Solun ylempi luku on % ja alempi luku on *f*.

B. Lapsen oma toiminta -mittarin merkinnöistä taulukossa 7 nähdään, että *roolileikit* (40 %, $f = 12$) ja *esineleikit* (53,3 %, $f = 16$) näyttävät sekoittuvan niin, että yli puolissa tilanteista roolileikki on merkitty esineleikkinä. Molempien kategorioiden yhteneväisyys on heikko. *Sääntöleikki* (60 %, $f = 5$) -kategorian yhteneväisyys on vähäinen, mutta merkintöjä on myös vähän. *Esineleikki* (60 %, $f = 35$) -kategorian yhteneväisyys on vähäinen, ja eriävät merkinnät ovat jakautuneet laajasti melkein kaikille kategorioille, joista suurimpia ovat tehtävä- ja muu toiminta -kategoriat. *Fyysinen leikki* (75 %, $f = 33$) on merkitty yhteneväisesti, ja eriävät merkinnät jakautuvat ei fokusta - ja esineleikki-kategorioiden alle. *Lukemista* (75 %, $f = 3$) on vähän, ja yksi observoija on mielenkiintoisesti merkinnyt saman tilanteen kiellettynä toimintana.

Tehtävät (73 %, $f = 22$) on merkitty melko yhteneväisesti, ja eriävät merkinnät ovat jakautuneet tasaisesti leikin eri kategorioihin ja muu kehys -kategoriaan. *Toiminta ilman fokusta* (59,6 %, $f = 28$) on merkitty jossain määrin yhteneväisesti, ja iso osa eriävistä merkinnöistä sijoittuu samankaltaisiin muu toiminta - ja yleinen kehys -kategorioihin. *Yhdessäolon* (53,6 %, $f = 15$) merkintöjen yhteneväisyys on heikko. Eriävät merkinnät ovat jakautuneet tasaisesti muille kategorioille, joista suurin on ollut fyysinen leikki. *Kielletty toiminta* (66,7 %, $f = 6$) on merkitty jossain määrin yhteneväisesti, ja eriävät merkinnät jakaantuivat ei fokusta-, yhdessäolo- ja yleinen kehys -kategorioiden kohdille.

Yleinen kehys (88,1 %, $f = 126$) on merkintämäärältään suurin kategoria. Observoijat ovat merkinneet tilanteet hyvin yhteneväisesti, ja eriävät merkinnät ovat jakautuneet usealle kategorialle. *Muu toiminta* (23,1 %, $f = 3$) on yhteneväisyydeltään huonosti merkitty kategoria, jossa suurin osa eriävistä merkinnöistä on jakautunut ei fokusta - ja yleinen kehys -kategorioihin. Analyysiosiossa tullaan pohtimaan tulosten valossa mittarin kategorioiden määrän vaikutusta merkintöjen yhteneväisyyteen, esine- ja roolileikin sekoittumista, tiettyjen kategorioiden päällekkäisyyttä sekä yleinen kehys -kategorian merkitystä.

6.2.3 C. Lapsen pääasiallinen huomion kohde

Taulukossa 4 mittarin C. *Lapsen pääasiallinen huomion kohde* kappa oli .56, 95 % CI [.495, .616]. Mittarin yhteneväisyys on Landisin ja Kochin (1977) mukaan keskinertainen ja Banerjeen ym. (1999) mukaan kohtalainen.

Taulukko 8

Ristiintaulukointi C. Lapsen pääasiallinen huomion kohde

	Ei- sos.	Aikuinen	Lapsi	Useampi lapsi	Kokonais- tilanne	r % f _i
Ei- sosiaalinen	72,4 (63)	5,7 (5)	2,3 (2)	6,9 (6)	12,6 (11)	100 (87)
Aikuinen	9,6 (5)	50 (26)	5,8 (3)	3,8 (2)	30,8 (16)	100 (52)
Lapsi	5,3 (3)	5,3 (3)	77,2 (44)	1,8 (1)	10,5 (6)	100 (57)
Useampi lapsi	12,9 (8)	-	6,5 (4)	61,3 (38)	19,4 (12)	100 (62)
Kokonais- tilanne	5,9 (9)	11,8 (18)	4,6 (7)	7,8 (12)	69,9 (107)	100 (153)

Huom. Solun ylempi luku on % ja alempi luku on *f*.

Taulukosta 8 nähdään, että observoijat ovat merkinneet *ei-sosiaalinen* (72,4 %, *f* = 63) vaihtoehdon melko yhteneväisesti, ja eriävät merkinnät ovat jakautuneet tasaisesti kaikkiin kategorioihin, joista eniten kokonaistilanne-kategoriaan. *Aikuinen* (50 %, *f* = 26) -kategorian merkintöjen yhteneväisyys on heikko, ja iso osa eriävistä merkinnöistä sijoittuu kokonaistilanne-kategoriaan. Observoijat ovat käyttäneet yhteneväisesti kategoriata *toinen lapsi* (77,2 %, *f* = 44) ja suurin osa muista merkinnöistä on kokonaistilanne-kategoriassa. *Useampi lapsi* (61,3 %, *f* = 38) -kategorian merkintöjen yhteneväisyys on vähäinen, ja suurin osa eriävistä merkinnöistä on kokonaistilanne- ja ei-sosiaalinen -kategorioiden kohdalla. *Kokonaistilanne* (69,9 %, *f* = 107) on merkinnöiltään

suurin kategoria, ja observoijat ovat käyttäneet sitä melko yhteneväisesti muiden merkintöjen jakautuessa tasaisesti kaikille muille kategorioille. Pohdintaosiossa tullaan hahmottelemaan syitä aikuinen-kategorian alhaiselle yhteneväisyydelle sekä kokonaistilanne- ja useampi lapsi -kategorioiden päällekkäisyydelle.

6.2.4 D. Lähin sosiaalinen lapsikontakti

Taulukossa 4 mittarin *D. Lähin sosiaalinen lapsikontakti* kappa oli .73, 95 % CI [.560, .900]. Landisin ja Kochin (1977) mukaan mittarin yhteneväisyys on huomattava ja Banerjeen ym. (1999) mukaan kohtalainen. *D*-mittarin merkintöjä pitää taulukossa 9 tarkastella hieman eri tavalla, sillä kyseessä on käytännössä ordinaaliasteikko muiden mittareiden ollessa nominaaliasteikollisia. Eri vaihtoehtoja on siis vaikea vertailla tai tulkita, sillä ne eivät pidä sisällään mitään toimintaa tai määritelmää. Järkevintä on tarkastella kokonaisuudessaan merkintöjen hajontaa ja *ei kontaktia* -vaihtoehdon painottumista.

Taulukko 9

Ristiintaulukointi D. Lähin sosiaalinen lapsikontakti

	Lapsi 1	Lapsi 2	Lapsi 3	Lapsi 4	Lapsi 5	Ei kontak.	<i>r % f_i</i>
Lapsi 1	30 (3)	-	-	-	-	70 (7)	100 (10)
Lapsi 2	-	69,7 (23)	3,0 (1)	-	-	27,3 (9)	100 (33)
Lapsi 3	6,1 (2)	6,1 (2)	66,7 (22)	-	-	21,2 (7)	100 (33)
Lapsi 4	-	-	-	60 (6)	-	40 (4)	100 (10)
Lapsi 5	-	3,3 (1)	-	-	53,3 (16)	43,3 (13)	100 (30)
Ei kontaktia	0,3 (1)	1,4 (4)	1,4 (4)	0,7 (2)	1,4 (4)	94,9 (280)	100 (295)

Huom. Solun ylempi luku on % ja alempi luku on *f*.

Taulukossa 9 *lapsi 1* (30 %, *f* = 3) -kategoriaa lukuun ottamatta kaikkien *lapsi*-kategorioiden kohdalla merkintöjen yhteneväisyys on vähäistä tai keskinkertaista (53,3–69,7 %). *Ei kontaktia* (94,9 %, *f* = 280) -kategorian kohdalla yhteneväisyys on lähes täydellinen. Prosenttilukuja tarkastellessa on hyvä ottaa huomioon, että joissain kategorioissa merkintöjä on selvästi vähemmän, jolloin pienikin ei-yhteneväisyys aineistossa voi siis näkyä suurena prosenttilukuna. Pohdintaosiossa tullaan keskittymään *D*-mittarin osalta *ei-kontaktia* -kategorian lukuihin sekä neutraalin vastausvaihtoehdon merkitykseen observoinnissa. Tämän lisäksi pohdintaosiossa tarkastellaan mittarin merkintöjen perusteella tehtäviin johtopäätöksiin liittyviä mahdollisia riskejä.

6.2.5 E. Lapsen fyysinen aktiivisuus

Taulukossa 10 näkyy merkintöjen jakautuminen mittarilla *E. Lapsen fyysinen aktiivisuus*. Taulukossa 4 mittarin kappia oli .51, 95 % CI [.390, .636]. Landisin ja Kochin (1977) mukaan mittarin yhteneväisyys on keskinkertainen ja Banerjeen ym. (1999) mukaan kohtalainen.

Taulukko 10

Ristiintaulukointi E. Lapsen fyysinen aktiivisuus

	Liikkumaton	Paik. tapaht. liike	Hitaat ja kevyet liikkeet	Koht. kuormitt. liikkeet	Nopea ja kuormitt. liike	r % f_i
Liikkumaton	51,7 (15)	44,8 (13)	3,4 (1)	-	-	100 (29)
Paikallaan tapahtuva liike	12,1 (27)	78,6 (176)	8 (18)	0,9 (2)	0,4 (1)	100 (224)
Hitaat ja kevyet liikkeet	1,7 (2)	24,3 (28)	67 (77)	5,2 (6)	1,7 (2)	100 (115)
Kohtalaisesti kuormitt. liikkeet	-	11,1 (3)	29,6 (8)	40,7 (11)	18,5 (5)	100 (27)
Nopea ja kuormittava liike	-	-	6,3 (1)	31,3 (5)	62,5 (10)	100 (16)

Huom. Solun ylempi luku on % ja alempi luku on f .

Taulukossa 10 *liikkumaton* (51,7 %, $f = 15$) -merkintöjen yhteneväisyys on vähäistä ja suurin osa eriävistä merkinnöistä on paikallaan tapahtuva liike -kategoriassa. *Paikallaan tapahtuvan liikkeen* (78,6 %, $f = 176$) -merkintöjen yhteneväisyys on hyvä. Katteoria on mittarin lukumäärällisesti eniten merkintöjä saanut vaihtoehto, ja eriävät merkinnät ovat jakautuneet enimmäkseen liikkumaton- ja hitaat ja kevyet liikkeet -kategorioihin. *Hitaat ja kevyet liikkeet* (67 %, $f = 77$) -merkintöjen yhteneväisyys on kohtalainen ja eriävät merkinnät painottuvat paikallaan tapahtuva liike -kategoriaan. *Kohtalaisesti kuormittavien liikkeiden* (40,7 %, $f = 11$) -kategorian merkintöjen yhteneväisyys on heikko, ja eriävät

merkinnät painottuvat hitaat ja kevyet liikkeet- ja nopea ja kuormittava liike -kategorioihin. *Nopea ja kuormittava liike* (62,5 %, $f = 10$) -merkintöjen yhteneväisyys on vähäinen ja suuressa osassa eriävistä merkinnöistä tilanne on nähty sijoittuvan kohtalaisesti kuormittavat liikkeet -kategoriaan.

Aikaisempi tutkimus -osiossa CARS-menetelmän havainnoijien välisen prosentuaalisen yhteneväisyyden ilmoitettiin olleen $84,1 \% \pm 10,1 \%$ (Puhl ym., 1990, s. 29) ja 88 % (Noland ym., 1990, 148). Pohdintaosiossa vastaava luku lasketaan tämän tutkimuksen aineistolle vertailun vuoksi. Osiossa vertaillaan myös observoinnin haastavuutta suhteessa CARS-menetelmään sekä pohditaan syitä merkintöjen ryhmittymiselle lähelle toisiaan.

6.2.6 F. Lapsen sitoutuneisuus

Seuraavaksi tarkastellaan taulukkoa 11 ja merkintöjen jakautumista mittarilla *F. Lapsen sitoutuneisuus*. Taulukossa 4 mittarin kappi oli .27, 95 % CI [.164, .377]. Landisin ja Kochin (1977) mukaan mittarin yhteneväisyys on vähäinen ja Banerjeen ym. (1999) mukaan yhteneväisyyden voidaan nähdä olevan heikko.

Taulukko 11

Ristiintaulukointi F. Lapsen sitoutuneisuus

	Yksinkertainen	Usein keskeytyvä	Enim. jatkuva	Jatkuva, jossa int. toimintaa	Kestävä, int. toiminta	<i>r %</i> <i>f_i</i>
Yksinkertainen	64,3 (18)	7,1 (2)	17,9 (5)	7,1 (2)	3,6 (1)	100 (28)
Usein keskeytyvä	11,3 (7)	35,5 (22)	43,5 (27)	9,7 (6)	-	100 (62)
Enimmäkseen jatkuva	10,1 (15)	12,8 (19)	45,3 (67)	27,7 (41)	4,1 (6)	100 (148)
Jatkuva, jossa int. toimintaa	5,4 (8)	6,8 (10)	23,6 (35)	52 (77)	12,2 (18)	100 (148)
Kestävä, int. toiminta	4 (1)	8 (2)	16 (4)	52 (13)	20 (5)	100 (25)

Huom. Solun ylempi luku on % ja alempi luku on *f*.

Taulukossa 11 *yksinkertainen* (64,3 %, *f* = 18) -kategorian merkinnät ovat kohtalaisen yhteneväisiä. Eriävät merkinnät ovat jakautuneet melko tasaisesti kaikkiin kategorioihin joista useimmin enimmäkseen jatkuva -kategoriaan. *Usein keskeytyvä* (35,5 %, *f* = 22) -kategorian merkintöjen yhteneväisyys on heikko, ja eriävistä merkinnöistä suurin osa on enimmäkseen jatkuva -kategoriassa. *Enimmäkseen jatkuva* (45,3 %, *f* = 67) -merkintöjen yhteneväisyys on myös heikko, ja eriävät merkinnät jakautuvat kaikkiin kategorioihin, joista suurin on jatkuva, jossa intensiivistä toimintaa -kategoria. *Jatkuva, jossa intensiivistä toimintaa* (52 %, *f* = 77) -kategorian merkintöjen yhteneväisyys on välttävä, ja jälleen merkinnät ovat jakautuneet kaikille vaihtoehdoille, joista suurin on enimmäkseen jatkuva -kategoria. *Kestävä, intensiivinen toiminta* (20 %, *f* = 5) -kategorian merkintöjen yhteneväisyys on huono, ja suurin osa eriävistä merkinnöistä on painottunut jatkuva, jossa intensiivistä toimintaa -kategoriaan.

Mittari *F* oli yhteneväisyydeltään heikoin mittari, ja merkintöjen hajonta on laajaa.

Pohdintaosiossa tullaan tarkastelemaan sitoutuneisuuden havainnoinnin haastavuutta ja koulutuksen merkitystä luotettavan havainnoinnin toteuttamisessa. Tämän tutkimuksen

tuloksia tullaan myös vertailemaan muiden LIS-YC-menetelmää käyttäneiden tutkimusten tulosten kanssa.

6.2.7 G. Lapsen emootio

Seuraavaksi tarkastellaan *G. Lapsen emootio* -mittarin merkintöjä taulukossa 12.

Taulukossa 4 mittarin kappi oli .34, 95 % CI [.259, .423]. Landisin ja Kochin (1977) määritelmän mukaan mittarin yhteneväisyys on kohtalainen ja Banerjeen ym. (1999) mukaan heikko.

Taulukko 12

Ristiintaulukointi G. Lapsen emootio

	Ärtymys, turh.	Suru, masennus	Ilo, riemu	Onni Tyytyväisyys	Hämmästys	Neutraali	Muu tunne	<i>r</i> % <i>f</i> _i
Ärtymys, turhauma	72,7 (8)	-	-	-	-	27,3 (3)	-	100 (11)
Pelko, ahdistus	-	-	-	-	-	100 (1)	-	100 (1)
Suru, masennus	-	-	-	-	-	100 (2)	-	100 (2)
Ilo, riemu	-	-	73 (27)	13,5 (5)	8,1 (3)	5,4 (2)	-	100 (37)
Onni, tyytyväisyys	1,4 (1)	-	19,2 (14)	35,6 (26)	12,3 (9)	30,1 (22)	1,4 (1)	100 (73)
Hämmästys	-	-	13,2 (5)	36,8 (14)	28,9 (11)	21,1 (8)	-	100 (38)
Neutraali	1,2 (3)	0,4 (1)	2,1 (5)	11,2 (27)	10,4 (25)	73,9 (178)	0,8 (2)	100 (241)
Muu tunne	12,5 (1)	-	-	12,5 (1)	12,5 (1)	37,5 (3)	25 (2)	100 (8)

Huom. Solun ylempi luku on % ja alempi luku on *f*.

Taulukkoa 12 yleisesti tarkasteltaessa voi nähdä, että negatiivisia tunteita, varsinkin yhteneväisiä, on merkitty selvästi vähemmän kuin positiivisia. Mittarin vähiten merkintöjä saanut kategoria on *pelko, ahdistus* (100 %, $f = 1$) ja se puuttuu sarakkeena kokonaan, eli yhtenevää parimerkintää ei ole. Samaten *suru, masennus* (100 %, $f = 2$) -kategorian merkinnöillä ei ole yhtenevää paria. *Ärtymys, turhauma* (72,7 %, $f = 8$) on merkitty melko yhteneväisesti, ja eriävät merkinnät sijoittuvat neutraali-kategoriaan. *Ilo, riemu* (73 %, $f = 27$) -kategoria on merkitty melko yhteneväisesti, ja eriävät merkinnät sijoittuvat muihin positiivisten tunteiden kategorioihin. *Onni, tyytyväisyys* (35,6 %, $f = 26$) -kategorian merkintöjen yhteneväisyys on heikko, ja eriävät merkinnät sijoittuvat kaikkiin muihin paitsi negatiivisiin kategorioihin. *Hämmästys* (28,9 %, $f = 11$) -kategorian merkinnöissä yhteneväisyys on huono. Itse asiassa yhteneväisiä merkintöjä enemmän on tilanteita, joissa toinen observoiija on merkinnyt emootioksi *onni, tyytyväisyys* (36,8 %, $f = 14$).

Neutraali (73,9 %, $f = 178$) -kategorian merkintöjen yhteneväisyys taulukossa 12 voidaan nähdä hyvänä. Neutraali on merkintämäärältään suurin kategoria mittarilla, ja sen yhteneväiset merkinnät kattavat 43 % kaikista mittarin merkinnöistä. Eriävien merkintöjen jakautuneisuus käsittää kaikki muut kategoriat, joista suurimpina hämmästys- ja onni, tyytyväisyys -kategoriat. *Muu tunne* (25 %, $f = 2$) -kategorian yhteneväisyys on huono, ja eriävät merkinnät jakautuvat useaan eri kategoriaan. Pohdintaosiossa tullaan muun muassa tarkastelemaan merkintöjen jakautumista erityisesti mittarin positiivisten emootioiden kohdalla sekä neutraali-kategorian problemaattisuutta niin merkintöjen yhteneväisyyden kuin mittarin toimivuuden kannalta.

6.2.8 H. Sosiaalinen orientaatio

Alla olevassa taulukossa 13 nähdään observointien jakautuminen mittarilla *H. Sosiaalinen orientaatio*. Taulukosta 4 voidaan tarkistaa mittarin kappale, joka oli .36, 95 % CI [.262, .464]. Mittarin yhteneväisyys on Landisin ja Kochin (1977) mukaan kohtalainen ja Banerjeen ym. (1999) mukaan heikko.

Taulukko 13

Ristiintaulukointi H. Sosiaalinen orientaatio

	Mukautuva, avoin	Osallistuva, vuorovaik.	Dominoiva, itsepintainen	Vetäytyvä, ei-sosiaalinen	Ei voi määrittää	<i>r</i> % <i>f_i</i>
Mukautuva, avoin	50 (80)	33,8 (54)	1,3 (2)	12,5 (20)	2,5 (4)	100 (160)
Osallistuva, vuorovaikut.	27,6 (37)	64,9 (87)	3 (4)	3 (4)	1,5 (2)	100 (134)
Dominoiva, itsepintainen	-	40 (6)	46,7 (7)	6,7 (1)	6,7 (1)	100 (15)
Vetäytyvä, ei-sosiaalinen	29,3 (22)	1,3 (1)	2,7 (2)	60 (45)	6,7 (5)	100 (75)
Ei voi määrittää	14,8 (4)	-	7,4 (2)	55,6 (15)	22,2 (6)	100 (27)

Huom. Solun ylempi luku on % ja alempi luku on *f*.

Taulukossa 13 *mukautuva, avoin* (50 %, *f* = 80) -kategorian merkintöjen yhteneväisyys on alhainen ja melkein kaikki eriävät merkinnät ovat kategorioissa *osallistuva, vuorovaikutteinen* ja *vetäytyvä, ei-sosiaalinen*. *Osallistuva, vuorovaikutteinen* (64,9 %, *f* = 87) -kategorian merkintöjen yhteneväisyys on kohtalainen ja suurin osa eriävistä merkinnöistä on kategoriassa *mukautuva, avoin*. *Dominoiva, itsepintainen* -kategorian merkintöjen osuus kaikista merkinnöistä on pieni (3,7 %, *f* = 15). Kategorian yhtenevien merkintöjen määrä (46,7 %, *f* = 7) on alhainen, ja iso osa eriävistä merkinnöistä sijoittuu mittarille *osallistuva, vuorovaikutteinen*. Yhteneväisyys kategoriassa *vetäytyvä, ei-sosiaalinen* (60 %, *f* = 45) on kohtalainen, ja iso osa eriävistä merkinnöistä on kategoriassa *mukautuva, avoin*. *Ei voi määrittää* (22,2 %, *f* = 6) -kategoriaa on käytetty melko harvoin, ja yhteneväisyys on huono eriävien merkintöjen painottuessa kategoriaan *vetäytyvä, ei-sosiaalinen*. Tulosten pohdinnassa tullaan tarkastelemaan kategorioiden vähyyden merkitystä merkintöjen yhteneväsyydelle sekä *mukautuva-* ja *osallistuva-* kategorioiden päällekkäisyyttä. Osiossa tullaan myös pohtimaan *ei voi määrittää* -kategorian merkitystä ja määrittelyä.

6.2.9 I. Lähimmän aikuisen toiminta

Alla olevan taulukon 14 kautta tarkastellaan observointilomakkeen viimeistä mittaria *I. Lähimmän aikuisen toiminta*. Taulukossa 4 mittarin kappa oli .46, 95 % CI [.307, .609]. Mittarin yhteneväisyys on Landisin ja Kochin (1977) mukaan keskinkertainen ja Banerjeen ym. (1999) mukaan kohtalainen.

Taulukko 14

Ristiintaulukointi I. Lähimmän aikuisen toiminta

	Vastaan- ottava	Osallis- tuva	Tavoit- telee	Pinnal- linen	Nega- tiivinen	Ei huom.	r % f_i
Vastaan- ottava	12,9 (4)	41,9 (13)	22,6 (7)	22,6 (7)	-	-	100 (31)
Osallistuva	8,6 (5)	55,2 (32)	29,3 (17)	6,9 (4)	-	-	100 (58)
Tavoittelee	2 (1)	18,4 (9)	69,4 (34)	10,2 (5)	-	-	100 (49)
Pinnallinen	2,9 (4)	13,6 (19)	6,4 (9)	62,1 (87)	0,7 (1)	14,3 (20)	100 (140)
Negatiivinen	-	16,7 (1)	33,3 (2)	33,3 (2)	16,7 (1)	-	100 (6)
Ei huomiota	-	2,4 (3)	6,3 (8)	13,4 (17)	-	78 (99)	100 (127)

Huom. Solun ylempi luku on % ja alempi luku on f .

Taulukossa 14 *Vastaanottava* (12,9 %, $f = 4$) -kategorian merkinnät ovat yhteneväisyydeltään huonoja, ja kategoria on merkinnöiltään kaikista observoinnin kategorioista vähiten yhteneväinen. Suurin osa eriävistä merkinnöistä on painottunut osallistuva-kategoriaan. *Osallistuva* (55,2 %, $f = 32$) -kategorian merkintöjen yhteneväisyys on alhainen ja iso osa eriävistä merkinnöistä on tavoitteleva-kategoriassa. *Tavoitteleva* (69,4 %, $f = 34$) -kategorian merkintöjen yhteneväisyys voidaan nähdä kohtalaisena. *Pinnallinen* (62,1 %, $f = 87$) -kategorian merkintöjen yhteneväisyys on

vähäinen. *Negatiivinen*-kategoriassa merkintöjä on kokonaisuudessaan vähän ($f = 6$) ja vain yksi niistä on yhteneväinen. *Ei huomiota* (78 %, $f = 99$) -kategorian merkintöjen yhteneväisyys voidaan nähdä hyvänä. Viimeisen mittarin kohdalla pohdintaosiossa aiotaan tarkastella vastaanottava-kategorian merkintöjen vähyttä, ei huomiota -kategorian merkitystä kappan koon muodostumiselle sekä merkintöjen painottumista eri kategorioissa.

7 Tulosten pohdinta

Kappa-kertoimet ja merkintöjen ristiintaulukointi osoittivat, että observoinnin yhteneväisyydessä on selviä eroja. Kappa-kertoimien lisäksi taulukoissa 2, 3 ja 4 on mielenkiintoisia eroja, jotka jäävät helposti huomioimatta vain keskiarvoja tarkasteltaessa. Näiden huomioiden lisäksi kappa-tulosten pohdinnassa tarkastellaan seuraavaksi kappan keskivirheen ja luottamusvälin merkitystä sekä Landisin ja Kochin (1977) taulukon sopivuutta tulosten kuvailemisessa.

7.1 Kappa-tulosten pohdinta

Taulukossa 2 parien kappa-kertoimien välillä on merkittäviä eroja. Esimerkiksi mittarilla *A* yleinen kappan taso on korkea $>.80$, mutta yhdellä parilla selvästi muita alhaisempi $.51$. Samanlaisia, suuria yksittäisten parien eroja yleisestä tasosta, löytyy myös mittareilla *E*, *F* ja *I*. Kaikkien parien kohdalla keskiarvoa nostaa mittari *A*, jonka merkitseminen yhteneväisesti on huomattavasti helpompaa verrattuna muihin mittareihin. Vastaavasti mittari *F* laskee jokaisen parin keskiarvoa, vaikka parin yhteneväisyys muilla mittareilla olisikin korkea. Tämän lisäksi voidaan nähdä taulukossa 2, että tietyillä mittareilla kertoimien arvot ovat enemmän hajaantuneita kuin toisilla. Esimerkiksi mittareilla *A*, *C* ja *E* suurin osa parien kertoimista on lähellä toisiaan, kun taas mittareilla *B*, *D*, *F*, *H* ja *I* parien kertoimet jakaantuvat laajasti. Yksittäiset parikohtaiset erot kuitenkin tasaantuvat, kun kaikki parien kappa-kertoimet lasketaan yhteen keskiarvoiksi (taulukko 3).

Keskivirheen määrä taulukossa 3 osoittaa, että parien yhteneväisyyden johdonmukaisuudessa on eroja. Esimerkiksi parilla 3 keskivirhe on yli kaksinkertainen pariin 4 verrattuna, ja keskivirheet ovat yleisesti suuria ottaen huomioon, että kappa-

kertoimet ovat melko alhaisia. Voidaan siis nähdä, että yhteneväisyys ei ole vakaata, ja on altis vaihteluille parista ja mittarista riippuen.

Luottamusvälin ala- ja ylärajan erot ja keskivirheen määrä taulukossa 3 osoittavat, että parien yhteneväisyys vaihtelee selvästi mittareittain. Yhdellä parilla ala- ja ylärajan välinen erotus on pyöristettynä .52, kuudella .30–.38 ja kahdella .23–.26. Usean observoijaparin kohdalla yhteneväisyyden keskiarvo ei siis kuvaa kovin totuudenmukaisesti sitä, miten kyseinen pari on käyttänyt observoinnin kaikkia mittareita. Mittareiden keskivirhe ja luottamusvälin ala- ja ylärajat taulukossa 4 eivät ole niin suuria kuin parien keskiarvoissa taulukossa 3. Mittarien pyöristetty luottamusvälin ala- ja ylärajan erotus on kahdella .30–.34, viidellä .20–.25 ja kahdella .12–.17. Mittarikohtaiset keskiarvot siis kuvaavat, ainakin parikohtaisia mittareita totuudenmukaisemmin, niitä lukuja joista keskiarvo muodostuu.

Tämän aineiston kontekstissa on mielestäni tarpeellista pohtia Landisin ja Kochin (1977) määritelmän sopivuutta hyväksyttävästä yhteneväisyyden asteesta. Ristiintaulukointi osoitti useiden mittareiden merkintöjen olevan jakautuneita, ja kappat osoittavat yhteneväisyyden olevan vaihtelevaa parien ja mittareiden välillä. Kuten sanottua, yhteneväisyyttä tulisi arvioida tutkimuksen ja aineiston kontekstissa.

McHugh'n (ks. 2012, s. 279) tavoin myös tässä tutkimuksessa Landisin ja Kochin määritelmät kappasta tuntuvat ristiriitaisilta tilanteissa, joissa .30 kappa tulisi kuvailla sanalla kohtalainen, tai .41 kappa sanalla keskinkertainen. Landisin ja Kochin (1977) määritelmät taulukossa 1 antavat sen kuvan, että jopa .21 kappa olisi kohtalainen tulos. Ristiintaulukointi osoitti, että näin alhainen kappa mittarilla tarkoittaisi sitä, että merkinnät ovat hyvin eriäviä. On siis keskeistä ymmärtää, että Landisin ja Kochin asteikkoa tulisi tulkita vain suuntaa-antavasti, ja että yhteneväisyydeltään kohtalaiseksi kuvailtu aineisto voi silti sisältää paljon eriäviä merkintöjä.

McHugh (2012) huomauttaa, että yhteneväisyyden aste kuvaa yksimielisyyden lisäksi myös sen vastakohtaa eli erimielisyyttä. Kappa on eräänlainen korrelaatiokerroin, joten sitä ei voi suoraan tulkita prosentuaalisena kuvauksena luotettavan datan osuudesta aineistossa. Kappasta voidaan kuitenkin laskea neliöjuuri, joka puolestaan voidaan tulkita yhteneväisen aineiston prosentuaalisena osuutena koko aineistosta. (McHugh, 2012, s. 279.) Taulukosta

15 on nähtävissä, kuinka luotettavan datan määrää aineistossa tulisi McHugh'n mukaan tulkita suhteessa kappakertoimen kokoon.

Taulukko 15

Luotettavan datan määrä aineistossa kappan perusteella

Kappa	Yhteneväisyyden aste	Luotettavan datan määrä aineistossa
0–.20	Olematon	0–4 %
.21–.39	Minimaalinen	4–15 %
.40–.59	Heikko	15–35 %
.60–.79	Keskiverto	35–63 %
.80–.90	Vahva	64–81 %
Yli 0.90	Melkein täydellinen	82–100 %

Huom. McHugh, 2012, s. 279.

Taulukossa 15 olevaa määritelmää voi soveltaa sekä parien että mittareiden kappaan. Tämän tutkimuksen aineiston kohdalla tulisi nimenomaan vaatia korkeaa yhteneväisyyttä ja mittareiden toimivuutta, sillä aineistoa käytetään varhaiskasvatuksen arviointiin. Ristiintaulukoinnin kautta tarkastellaan esimerkiksi, millainen oppimisympäristö on yhteydessä lasten korkeaan sitoutumiseen, ja aineiston perusteella saatetaan tehdä johtopäätöksiä lasten tuntemuksista, sosiaalisista suhteista ja kasvattajan roolin merkityksestä lapsiryhmätilanteissa.

Taulukon 15 perusteella virheen määrää tämän aineiston merkinnöissä voi usean parin ja mittarin kohdalla luonnehtia huomattavaksi, ja virheen määrä vaikuttaa merkintöjen sisältämän aineiston luotettavuuteen. Lähteet (Graham ym., 2012, s. 9; Shweta ym., 2015, s. 26; Byrt, 1996, s. 561) määrittelevät hyväksyttävän kappan raja-arvon välillä .60–.61, ja myös tässä tutkimuksessa .60 tulkitaan hyväksyttävänä raja-arvona. Jos tarkastellaan tuloksia kriittisesti taulukon 15 perusteella vain mittarit *A. Yleinen toiminta*, *B. Lapsen oma toiminta* ja *D. Lähin sosiaalinen kontakti* saavuttavat kyseisen tason, jossa virheen määrä

aineistossa ei ole merkittävä. Observoijaparien kohdalla raja .60 ylittyy vain kolmella parilla. McHugh'n kriteerien perusteella muiden parien merkinnöissä on siis niin paljon virhettä, ettei aineistoa tulisi tulkita luotettavana. Samaten mittareilla *F. Lapsen sitoutuneisuus*, *G. Lapsen emootio*, *H. Sosiaalinen orientaatio* ja *I. Lähimmän aikuisen toiminta* kertoimet ovat McHugh'n kriteerien perusteella tasolla, jossa aineistossa olevan virheen määrän voi nähdä merkittävänä. Kohtalainen kappa ei siis suoraan tarkoita kohtalaista tulosta, ja virheen määrään osassa tuloksia tulisi suhtautua vakavasti tulosten käyttötarkoitusta pohdittaessa.

7.2 Ristiintaulukoinnin pohdinta

Ristiintaulukoinnin tulokset osoittivat, että merkintöjen jakautuminen vaihtelee huomattavasti mittarin aiheen ja rakenteen mukaan. Ilman lisäaineistoa on mahdotonta määrittellä vaihtelevuuden syitä, mutta yhdistämällä ristiintaulukoinnin tulokset observoinnin luonteeseen on mahdollista hahmotella syitä merkintöjen yhteneväsyydelle ja eriäväsyydelle.

7.2.1 A. Yleinen toiminta

A. Yleinen toiminta (taulukko 6) on moneen muuhun verrattuna yksinkertainen mittari, sillä havainnoitava aihe (mitä tehdään) on selkeä ja vaihtoehdot eroavat toisistaan, lukuun ottamatta *tuettu leikki* - ja *vapaa leikki* -kategorioita. Näiden kahden sekoittumista voi selittää leikin luonne ja erojen hahmottamisen haastavuus; lasten leikkien sisällöt ja painotukset ovat hienovaraisia ja dynaamisia asioita, joiden täsmällinen määrittely voi olla vaikeaa. Näiden kahden kategorian erotteluun tulisi kiinnittää koulutuksissa huomiota. Muut mittarin asiat taas ovat tulosten perusteella ilmeisen helppoja havainnoida: ulkoilu on ulkoilua, ja ruokailu on ruokailua.

7.2.2 B. Lapsen oma toiminta

Taulukossa 7 *B. Lapsen oma toiminta* -mittarin kategorioiden määrä näyttäisi vaikuttavan negatiivisesti yhteneväsyyteen – observoijilla on enemmän vaihtoehtoja, joista valita. Yhtenä huomiona voisi nostaa esiin kategoriat *toiminta ilman fokusta*, *yhdessäolo* ja *muu*

toiminta. Näihin kategorioihin oli tehty paljon merkintöjä, mutta merkintöjen yhteneväisyys vaihteli kohtalaisesta (59,6 %) huonoon (23,1 %) ja iso osa eriävistä merkinnöistä painottui johonkin edellä mainituista kolmesta kategoriasta. Näiden kategorioiden käytön voi siis nähdä olevan ei-yhteneväistä ja päällekkäistä, jolloin joko niiden relevanssia tai selkeyttämistä tulisi pohtia. Useamman parin kohdalla *roolileikki* oli merkitty eriävästi *esineleikki*-kategoriaan. Tämä on ongelmallista, jos aineiston pohjalta tehdään päätelmiä leikin ja jonkin toisen muuttujan yhteydestä.

Usean kategorian kautta syntyvä ongelma ilmenee hyvin tarkasteltaessa *esineleikki*-kategoriaa taulukossa 7. Sama toiminta oli tulkittu usealla eri tavalla, jolloin merkinnät olivat eriäviä ja jakautuivat useaan eri kategoriaan. Haastavaksi osoittautuvan aiheen havainnointi yhdistettynä useaan vaihtoehtoon vaikuttaa siis sekä mittarin reliabiliteettiin (mittaako mittari samaa asiaa luotettavasti) että validiteettiin (mittaako mittari sitä mitä sen on tarkoitus mitata).

Taulukossa 7 toisena ääripäänä samasta asiasta tuntuu olevan kategoria *yleinen kehys*, jonka kaikista merkinnöistä ($f_i = 143$) yhteneväisiä oli 88 % ($f = 122$).

Observointikoulutuksen käyneenä ja observointia itse tehneenä uskoisin kategorian määritelmän vaikuttavan tulokseen: yleinen kehys on tilanne jolloin ”lapsi tekee sitä mitä hänen kuuluukin tehdä”. Määritelmä on laaja, joten tilannetta havainnoitaessa moni näkemys sopii saman kategorian alle. Neutraali kategoria on siis helpompi merkitä yhteneväisesti.

7.2.3 C. Lapsen pääasiallinen huomion kohde

Taulukossa 8 *C. Lapsen pääasiallinen huomion kohde* -mittarilla *ei-sosiaalinen* -kategoria eroaa dikotomisuuksellaan huomion kohteen määrittelystä ja on ehkä tästä syystä selkeä ja toimiva kategoria. Yhteneväisten merkintöjen vähyys *aikuinen*-kategoriassa saattaa johtua varhaiskasvatuksen tilanteiden luonteesta: tilanteet, joissa yksi aikuinen on yhden lapsen kanssa, ovat harvinaisia. Toisaalta lapsen näkökulmasta tarkasteltuna on harvoin tilanteita, jolloin ympäristö olisi vapaa muusta toiminnasta ja lapsen huomio kohdistuisi vain aikuiseen. Useimmiten ympärillä on myös muita lapsia ja tilanteet ovat, kuten tulokset osoittavat, dynaamisia. Tämä ei kuitenkaan muuta sitä, että kategorian käyttö ei ollut yhteneväistä. Kuten mittarissa *B*, myös mittarissa *C* yleisluontoinen vaihtoehto

kokonaistilanne kattoi ison osan merkinnöistä. On siis hyvä pohtia, kuinka paljon tarkkaa tietoa havainnoitavasta aiheesta mittari sisältää, jos iso osa yhteneväisistä merkinnöistä on tämäntyyppisessä yleiskategoriassa.

7.2.4 D. Lähin sosiaalinen lapsikontakti

Merkintöjen painottuminen *ei-kontaktia* -kategoriaan mittarilla *D. Lähin sosiaalinen lapsikontakti* (taulukko 9) voi tulkita johtuvan observoitavien lasten valintaan liittyvästä satunnaisuudesta sekä varhaiskasvatuksen luonteesta. Observoitavat lapset ovat samasta ryhmästä, mutta voivat olla observointiaikana eri pienryhmissä, eri leikissä tai eri toiminnassa. Observoitavilla lapsilla ei siis ole välttämättä mahdollisuutta tai syytä olla kontaktissa toisiinsa observoinnin aikana. Kuten niin sanotut yleiskategoriat mittareissa *B* ja *C*, myös *D*-mittarin *ei-kontaktia* -kategorian voi nähdä olevan yleinen vaihtoehto. Vaihtoehdon voi valita, jos ei ole varma asiasta. Mittarin kappan vahvuus perustuu pitkälti tälle kategorialle, jonka korkea yhteneväisten merkintöjen määrä ($f = 280$) oli 68 % kaikista merkinnöistä. Eli lopulta iso osa mittarin yhtenevistä merkinnöistä perustui sille, että observoitavaa ei ole.

Lapsikontaktimerkinnöissä parit näkivät harvoin eri lapsen olevan observoitavan lapsen lähin kontakti. Sen sijaan kaikkien lapsikontaktien kohdalla useampi observoija oli merkinnyt tilanteen eriävästi *ei-kontaktia* (taulukko 9). Voisi siis tulkita, että tilanteet ovat dynaamisia ja että tarkan sosiaalisen kontaktin määrittäminen näissä tilanteissa on vaikeaa. Tästä päästään näkökulmaan, joka liittyy mittarin validiuteen. Mittarin *D* tavoitteena on hankkeen aineistossa tarkastella sitä, viettävätkö erilaiset tai samanlaiset lapset aikaa toistensa kanssa. Olisi siis tärkeää, että merkintä lapsen sosiaalisesta kontaktista perustuisi varmaan havaintoon. Tulosten tulkinnan kautta näyttäisi siltä, että monet näistä tilanteista ovat niin dynaamisia ja muuttuvia, että kahden lapsen yhteyden täsmällinen määrittely on vaikeaa. Tämä taas voi johtaa merkityn kontaktin ylikorostamiseen tuloksissa ja ristiintaulukoinnissa. Tuloksien perusteella mittarin pohjalta tehdyissä johtopäätöksissä tulisi siis olla varovainen.

7.2.5 E. Lapsen fyysinen aktiivisuus

E. Lapsen fyysinen aktiivisuus -mittari (taulukko 10) on ensimmäinen yhtä tiettyä asiaa lapsessa havainnoiva mittari. Eriävien merkintöjen sijoittuminen lähelle yhteneväisiä merkintöjä voi tulkita niin, että lähekkäin olevien kategorioiden jaottelu on observoijille haastavaa. *Liikkumaton*-kategorian tulisi määritelmän perusteella olla melko selvä, mutta sen sekoittuminen *paikallaan tapahtuva liike* -kategorian kanssa kertoo siitä, että koulutuksissa tulisi kiinnittää huomiota kategorioiden selkiyttämiseen.

Aikaisemmin esitelty CARS-menetelmä, johon mittarin kategoriat pohjautuvat, perustuu tarkkaan ja pitkäaikaiseen mittaamiseen. Mittarin kategoriat perustuvat liikkeiden kalorinkulutukseen ja hapenottoon, joita mitattiin erityisesti siihen tarkoitettujen välineiden avulla. Lisäksi CARS keskittyy vain lapsen fyysisen aktiivisuuden havainnointiin, kun tässä aineistossa se on yksi yhdeksästä havainnoitavasta asiasta. Puhl ym. (1990, s. 29) ilmoittivat prosentuaaliseksi yhteneväisyydeksi 84,1 % \pm 10,1 % ja Noland ym. (1990, s. 148) ilmoittivat yhteneväisyydeksi 88 %. Tässä aineistossa samalla menetelmällä mittarin yhteneväisyydeksi tuli $289 / 411 = 70$ %. Näin laskettuna tulos ei siis oikeastaan ole juurikaan alhaisempi. Prosentuaalisen yhteneväisyyden käyttäminen ei kuitenkaan tässä tapauksessa sovi tutkittavaan menetelmään, jossa jokainen merkintä sen tarkkuudesta riippumatta päätyy osaksi aineistoa.

7.2.6 F. Lapsen sitoutuneisuus

Mittarin *F. Lapsen sitoutuneisuus* merkinnät taulukossa 11 osoittavat, että sitoutuneisuuden havainnoinnin yhteneväisyys tässä aineistossa oli heikkoa ja eriävät merkinnät jakautuivat kaikkiin kategorioihin. Selvästi alhaisen kertoimen voidaan olettaa johtuvan joko ongelmasta mittarin käytössä tai mittarin rakenteesta.

Hautamäki ja Laevers (1997, s. 18) ilmoittivat reliabiliteetiksi kertoimen Spearman 0.90, ja Ebbeck ym. (2012, s. 612) ilmoittivat prosentuaaliseksi yhteneväisyydeksi 90 %. Tässä aineistossa Spearmanin rho oli 0.475, ($p < .000$) ja prosentuaalinen yhteneväisyys oli $189 / 411 = 46$ %. Molemmat luvut ovat siis korkeampia kuin kappa .27, mutta selvästi matalia verrattuna lähteiden tuloksiin. Kalliala (2011, s. 253) ilmoitti luotettavuuden kappan sijaan korrelaationa ”inter-rater correlation 0.93”. Kallialan tutkimuksesta ei käy ilmi, mitä

korrelaatiokerrointa laskemiseen on käytetty, mutta kahden nominaaliasteikollisen ei-dikotomisen muuttujan korrelaatiokertoimeksi suositellaan Cramerin V -kerrointa (Sivonen, 2004). Tässä aineistossa Cramerin V oli F -mittarin merkinnöillä pareittain välillä 0.26–0.63 ja mittarin keskiarvo oli 0.44. Seitsemän parin sekä mittarin keskiarvon kertoimet olivat tilastollisesti erittäin merkitseviä ($p < .000$). Kertoimen vahvuus on siis selvästi Kallialan ilmoittamaa kerrointa heikompi. On tosin huomioitava, että Kallialan tutkimuksessa otanta on huomattavasti pienempi, ja se koostuu vain yhden parin merkinnöistä.

Muiden tutkimuksien perusteella näyttäisi siis siltä, että LIS-YC-menetelmän käytössä on mahdollista saavuttaa korkea yhteneväisyys. Voidaan siis olettaa, että mittarin kategoriat mittaavat niitä ilmiön ulottuvuuksia, joita sen tulisi mittarin teoreettisessa kontekstissa mitata. Matala kappa ei siis todennäköisesti johdu mittarista itsestään, vaan siitä miten mittaria käytetään.

Laevers (1994, s. 164) määrittelee LIS-YC-ohjeistuksessa seuraavat seikat, joihin havainnoinnissa tulisi kiinnittää huomiota: keskittyminen, energia, toiminnan monimutkaisuus, ilmeet ja asento, sitkeys, tarkkuus, reaktioaika, kieli ja tyytyväisyys. Laeversin (2006) mukaan havainnoijan empatiakyky on keskeisessä asemassa sitoutuneisuuden havainnoinnissa ja LIS-YC-menetelmässä. Havainnoijan tulisi hänen mukaansa pyrkiä näkemään tilanne lapsen silmin ja ”olla lapsi”. (Laevers, 2006, s. 3.) Laeversin ohjeistusta on käytetty myös observointikoulutuksessa, mutta tulosten perusteella on aiheellista pohtia, onko ohjeistus toteutunut käytännössä. On myös perusteltua pohtia, olisiko tämä edes realistista. Muissa LIS-YC-menetelmää käyttäneissä tutkimuksissa on keskitytty vain yhden tai kahden asian havainnointiin (ks. Knoors & Renting, 2000, s. 271; Ebbeck ym., 2012, s. 612) kun *Kehittävä palaute* -hankkeen tapauksessa havainnoitavia asioita on yhdeksän.

Sitoutuneisuuden käsite liitetään usein flow-tilaan ja nähdään laadukkaana toiminnan tunnusmerkkinä (ks. Pascal, Bertram, Mould, & Hall, 1998, s. 55; Goldspink & Foster, 2013, s. 294). Flow-tilan voidaan nähdä tarkoittavan yksilön kykyjen ja ympäristön haasteiden välillä syntyvää suhdetta, joka näkyy intensiivisenä ja keskittyneenä toimintana (Nakamura & Csikszentmihalyi, 2014, s. 240). Sama yksilön ja ympäristön välinen suhde voidaan myös nähdä keskeisenä osana lähikehityksen vyöhykkeen käsitettä (ks. Daniels,

2005, s. 5). On siis perusteltua tutkia lasten sitoutuneisuutta osana varhaiskasvatuksen kehittämistä ja pyrkiä nostamaan esiin se toiminta, jonka yhteydessä esiintyy korkeaa sitoutuneisuutta. Samalla on kuitenkin tärkeää arvioida, millä tarkkuudella sitoutuneisuutta on mahdollista tarkastella niin, että varmistetaan samalla merkintöjen ja niistä tehtävien johtopäätösten luotettavuus.

Tulosten perusteella on selvää, että sitoutuneisuuden havainnointia tulisi painottaa enemmän observointikoulutuksessa. Tämän lisäksi tulisi pohtia mittarin muuttamista yksinkertaisemmaksi. Observoinnissa käytetään useaa mittaria lyhyen ajan sisällä ja observoinnin kohdetta vaihtaen, ja tarkan tiedon merkitseminen on ilmeisen haastavaa. Mittarin kategoriat eivät siis ole tasapainossa observoinnin tarkkuuden kanssa. Lopulta sitoutuneisuudessa ollaan todennäköisesti eniten kiinnostuneita siitä, missä tilanteissa sitoutuneisuus voidaan nähdä korkeana tai matalana. Tällä hetkellä viisiportaisen asteikon käytössä on paljon eriäviä merkintöjä, joten yksinkertaisempi matala-keskitaso-korkea -asteikko voisi toimia paremmin. Tällaisella asteikolla tuloksissa todennäköisesti näkyisi edelleen mielenkiintoisia eroja, mutta virheen määrä olisi pienempi ja mittarin käyttö ei olisi niin vaativaa.

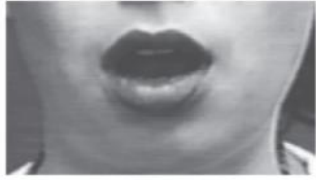
7.2.7 G. Lapsen emootio

Taulukosta 12 voidaan nähdä, että *G. Lapsen emootio* mittarin alhainen yhteneväisyys johtui merkintöjen jakautumisesta useassa eri kohdassa. Näyttäisi siltä, että kategorioita on liian paljon ja ne muistuttavat liikaa toisiaan. Erityisesti positiivisten tunteiden jakaminen useaan eri kategoriaan näyttäisi luovan eriäväisyyttä merkinnöissä. *Ilo, riemu* -kategoria oli observoijien mielestä vielä melko selkeä, mutta *onni, tyytyväisyys* -kategorian kohdalla merkinnät jakautuivat selvästi. Tämän perusteella voi siis olettaa, että kyseinen tunnetila määritellään liian subjektiivisesti tai tilanteiden dynaamisuudesta johtuen tarkan havainnon tekeminen on vaikeaa. *Hämmästys* -kategorian kohdalla yhteneväisiä merkintöjä oli jopa vähemmän kuin tilanteita, joissa toinen oli merkinnyt tilanteen *onni, tyytyväisyys* -kategoriaan. Tuloksen perusteella on siis aiheellista harkita sitä, onko kategoria tarpeellinen tällä mittarilla tai sitä, onko kyseessä liian tarkka tunnetila havainnoitavaksi. *Neutraali* -kategorian kohdalla ollaan jälleen tilanteessa, jossa mittarin merkinnät painoutuivat yhden, yleisen kategorian ympärille. *Neutraali* -merkintöjen suuri määrä ja muiden merkintöjen eriäväisyys iloa lukuun ottamatta luovat lopulta tilanteen, jossa

luotettavan tai syvällisen tiedon osuus havainnoitavasta aiheesta on lopulta melko vähäinen.

Tunteiden ilmaisuun ja havaitsemiseen liittyvää tutkimusta on tehty valtavasti (ks. Matsumoto, 1993; Righart & de Gelder, 2008; Gomez-Garibello & Young, 2018) ja kysymykseen oikeasta menetelmästä tunteiden havainnointiin ei ole yhtä helppoa vastausta. *Kehittävä palaute* -hankkeen emootiomittarin pohjana toimiva Ekmanin alkuperäinen, jo aikanaan kritisoitu (ks. Ortony & Turner, 1990; Russell, 1994) näkemys perustunteista tulisi siis nähdä vain yhtenä suuntauksena ja tunteita tulisi, myös Ekmanin (2005, s. 624) itsensä mielestä, tarkastella moniulotteisena aiheena. Tunteidenilmaisua ja -havainnointia on tutkittu paljon esimerkiksi eri kulttuurien ja aivotoiminnan (ks. Elfenbein & Ambady, 2002; Barrett, 2017) merkityksen konteksteissa. Tunteiden määrittely ja havaitseminen luotettavasti ovat siis monimutkaisia konsepteja, jolloin tunnetilojen havainnointiin osana aineistonkeruumenetelmää tulisi suhtautua kriittisesti.

Ekmanin tunteidenilmaisuun liittyvä tutkimus painottaa kasvojen liikkeiden yhteyttä tiettyjen tunteiden ilmaisemiseen, ja aihetta varten on kehitetty oma menetelmä FACS (Facial Action Coding System). FACS-menetelmässä kasvojen eri lihasryhmät ja niiden liikkeet on koodattu niin, että esimerkiksi kulmakarvojen tietty asento voidaan nähdä liittyvän tiettyyn tunteeseen (Bakeman & Gottman, 1986, s. 23–24). Alla kuvassa 1 on osa FACS-menetelmän koodausasteikkoa.

AU	Description	Facial muscle	Example image	Interrater agree (Kappa coeffic (tolerance window in :		
		<i>Orbicularis oris</i>				
26	Jaw Drop	<i>Masseter, relaxed Temporalis and internal Pterygoid</i>		.65	.72	.76

Kuva 1. FACS-koodausasteikon osa (Cohn, Ambadar, & Ekman, 2007, s. 209)

Kasvonliikkeiden havainnoinnissa käytetään hidastettua, tarkkaa videokuvaa, ja menetelmä on laajasti käytetty ja tieteellisesti testattu (Cohn, Ambadar, & Ekman, 2007, s. 203). Mielenkiintoisena huomiona Cohn ym. (2007) ilmoittavat artikkelissaan erään tutkimuksen FACS-menetelmän luotettavuuden yksityiskohtaisina kappa-kertoimina. Kappan taso vaihtelee välillä .31–.86 ja he toteavat, että matalan kertoimen osa-alue menetelmässä vaatii kehittämistä. (Cohn ym., 2007, s. 213–214.)

FACS-esimerkki osoittaa, kuinka tarkkoja menetelmiä tunteiden havainnoinnissa saatetaan käyttää luotettavan havainnon takaamiseksi. Samanlaista tarkkuutta ei voi vaatia tietenkään jokaiselta menetelmältä, mutta kaikkien menetelmien tulisi pyrkiä samanlaiseen tasapainoon mittauksen luotettavuuden, ilmiön tarkastelun ja tuloksista tehtävien johtopäätösten kanssa.

Tutkimuksen ote -osiossa viitattiin Erätuulen ym. (1994, s. 10–11) määritelmään, jonka mukaan kvantitatiivisten menetelmien tavoitteena on saada mitattava ilmiö mittauksen kautta numeeriseen hallintaan. Tulosten perusteella tässä tutkimuksessa tarkasteltavan emootiomittarin ilmiön numeerinen hallinta on melko kapea, sillä suurin osa yhteneväisistä merkinnöistä on vain kahdessa kategoriassa. Luotettavan havainnon tekeminen toisen henkilön tunteesta on vaikea konsepti, jonka tiivistäminen yhteen merkintään ei oheisten esimerkkien perusteella ole välttämättä edes mahdollista. On tärkeää miettiä, miten paljon yhden merkinnän perusteella voidaan tehdä johtopäätöksiä ja toisaalta miten tämä merkintä voisi kuvata mahdollisimman totuudenmukaisesti kyseistä havaintoa.

Mielestäni olisi keskeistä pohtia kriittisesti, miten mittarin saisi toimimaan luotettavasti ja keräisi samalla eksaktisti sen tiedon, joka on hankkeen tavoitteiden kannalta tärkeää. Näyttäisi siltä, että ainakin kategorioiden yhdistämisen ja tiivistämisen tulisi olla osa tätä pohdintaa. Voi olla, että yksinkertainen ilo, suru, pelko, viha ja neutraali -mittari antaisi yhtä paljon tai jopa enemmän tietoa, ollen samalla merkinnöiltään luotettavampi. Toisaalta tulisi myös miettiä voisiko jokin muu, jo toimivaksi todettu, lasten tunnetilojen havainnointiin tarkoitettu mittari olla parempi ratkaisu.

7.2.8 H. Sosiaalinen orientaatio

Taulukossa 13 nähtiin, kuinka mittarin *H. Sosiaalinen orientaatio* tapauksessa *mukautuva, avoin* -kategoria ja *osallistuva, vuorovaikuttava* -kategoria sekoittuivat keskenään. Vaihtoehtojen erot ovat siis observoijille epäselvät, tai tarkan merkinnän tekeminen kyseisestä aiheesta on tilanteissa liian vaativaa. Näyttäisi muutenkin siltä, että eri kategorioiden erot eivät ole observoijilla selvästi tiedossa. *Dominoiva, itsepintainen* -kategoria sekoittui *osallistuva*-kategorian kanssa. Tosin merkintöjä oli vähän, joten niiden vaikutus yhteneväisyyteen ei ollut suuri. Ehkä erikoisinta on, että *vetäytyvä, ei-sosiaalinen* -kategoria ja *mukautuva, avoin* -kategoria sekoittuivat. Jos sama tilanne voidaan nähdä niin eri tavoin, on aiheellista pohtia, kuinka hyvin mittarin kategoriat vastaavat sitä, mitä mittarin olisi tarkoitus mitata. Mittarin vaihtoehdot pohjautuvat teoriaosiossa esitettyyn orientaatioteoriamalliin, ja tulosten perusteella mallin nelikentän siirtäminen teoriasta käytäntöön on haastavaa.

Muilla mittareilla neutraalit kategoriat painoutuivat merkinnöissä, mutta sosiaalisen orientaation kohdalla *ei voi määrittää* -kategoriassa oli vain vähän merkintöjä. Näistä merkinnöistä myös vain harva oli merkitty yhteneväisesti. Voisi siis olettaa, että observoijat eivät itse koe lapsen sosiaalisen orientaation havainnointia vaikeana asiana. Tulosten perusteella asia ei kuitenkaan näytä olevan näin yksinkertainen, ja mittarin luotettavuutta tulisi pyrkiä kehittämään. Mittarin kategoriat perustuvat orientaatioteoriamalliin ja tulosten arvioinnissa tullaan samaan tilanteeseen kuin sitoutuneisuuden ja emotion kohdalla: luotettavan havainnon tekeminen abstraktista aiheesta näyttäisi olevan vaikeaa. Samaan aikaan on myös tärkeää pohtia, onko tämä mittari ja nämä kategoriat se tapa, jolla kyseinen aihe saadaan tallennettua luotettavasti ja ilmiön eri ulottuvuudet hahmottaen.

Bakeman ja Gottman (1986) toteavat mittarin ja sen kategorioiden olevan linssi, jonka läpi tutkittavaa ilmiötä päätetään tarkastella. Linssin läpi välittyvä tieto on selkeää, jos linssi on hyvin rakennettu ja oikein suunnattu. Jos linssi rakenne taas ei ole kunnossa, on sen läpi välittyvän tiedon jälkeensä korjaaminen vaikeaa. (Bakeman & Gottman, 1986, s. 19.) Sosiaalinen orientaatio on erittäin abstrakti käsite, jolla pyritään tiivistämään lapsen olemus neljään toisensa poissulkevaan kategoriaan. Observoijien koulutus on keskeinen osa luotettavuuden parantamista, mutta sitäkin tärkeämpää on, että käytettävä mittari ja sen

kategoriat ovat sopivat kyseisen ilmiön mittaamiseen. Luotettavuuden vahvistaminen ei siis tarkoita vain observoidijien koulutusta mittarin käytössä vaan myös mittarin muokkausta ilmiön tarkastelun parantamiseksi.

7.2.9 I. Lähimmän aikuisen toiminta

Taulukosta 14 nähtiin, että *I. Lähimmän aikuisen toiminta* -mittarin *vastaanottava*-kategorian merkinnöistä vain häviävän pieni osa oli yhteneväisiä. Observoidijat eivät siis näe kategorian määritelmää selvästi tai olennaisena osana havainnoitavaa aihetta. Muutenkin eri kategoriat sekoittuivat keskenään melko paljon. On tarpeellista pohtia, onko havainnoitava aihe liian dynaaminen luotettavan merkinnän tekemiseen vai onko kyseessä kategorioiden määrittelyyn liittyvät ongelmat. Toisaalta yhteneväisyys parani kategorioissa *pinnallinen* ja *ei huomiota*, joten kyseessä näyttäisi olevan enemmän mittarin vaihtoehtoihin ja niiden täsmälliseen määrittelyyn liittyvä ongelma. Yhteneväisyys oli kaikkein korkein *ei huomiota* -kategoriassa. Mittarin yhteneväisyyden kannalta paras kategoria oli siis jälleen se, jonka määrittely oli dikotominen ja helppo.

I. Lähimmän aikuisen toiminta -mittarin kategoriat perustuvat samaan teoriaosiossa esiteltyyn orientaatioteoriamalliin kuin aikaisempi *H. sosiaalinen orientaatio* -mittari. Mallin näkemyksen mukaan aikuisen toiminta on mahdollista määritellä mittarin kategorioiden mukaisesti. Näyttäisi kuitenkin siltä, että observoidijat eivät erota kaikkia kategorioita selvästi toisistaan tai että he näkevät tilanteen painotuksen keskenään hyvin eri tavoin. Nykyisessä muodossaan mittari siis mittaa heikosti sitä, mitä sen kautta pyritään mittaamaan. Kuten lapsen sosiaalinen orientaatio, myös aikuisen toiminnan määrittely on dynaaminen tilanne. Näiden dynaamisten tilanteiden määrittely valmiisiin kategorioihin luotettavasti on haasteellista, jopa ongelmallista, jos merkintöjen pohjalta tehdään päätelmiä tai syy-seuraussuhteita.

Kuten sitoutuneisuuden, emotion ja sosiaalisen orientaation mittareissa, myös aikuisen toiminnan mittarin kategorioiden sopivuutta kyseisen ilmiön mittaamiseen tulisi tarkastella tulosten valossa. Tällä hetkellä voisi kuvitella, että luotettavampaa observointia aikuisen toiminnasta voisi tehdä helpommalla asteikolla kuten paikalla–ei paikalla tai pinnallinen osallistuminen–syvämpi osallistuminen. Tämän tyyppinen, yksinkertaisempi mittari on

ollut käytössä myös *Kehittävä palaute* -hanketta edeltäneessä *Orienteatioprojektissa* (ks. Reunamo ym., 2014, s. 622).

7.2.10 Ristiintaulukoinnin pohdinnan yhteenveto

Mittarien ristiintaulukoinnin yhteenvetona voidaan todeta, että mittareiden *A. Yleinen toiminta*, *B. Lapsen oma toiminta* ja *D. Lähin sosiaalinen lapsikontakti* kategoriat näyttävät toimivan melko hyvin sen asian mittaamisessa, mitä niiden on tarkoitus mitata. Muilla mittareilla on yksi tai useampi kategoria, joita ei syystä tai toisesta käytetä yhteneväisesti. Ristiintaulukointi myös paljastaa, että usean mittarin kohdalla merkittävä osa yhteneväisistä merkinnöistä on kategorioissa, jotka tarkastelevat asiaa yleisellä tasolla, määrittelevät tilanteen neutraalina tai havainnoitavan asian poissaolevana. Kolmesta yllämainitusta mittarista esimerkiksi *D*-mittarin yhteneväisyys koostuu pääosin *ei-kontaktia* -merkinnöistä. Yhteneväisyydeltään vahva mittari ei siis tarkoita automaattisesti sitä, että observoinnin aiheita havainnoitaisiin moniulotteisesti tai että observoijat käyttäisivät koko mittaria yhteneväisesti.

Bakemanin ja Gottmanin (1986) mukaan mittarin rakentuminen on tärkein yksittäinen tekijä onnistuneen havainnointitutkimuksen suorittamisessa. Mittarin rakentaminen voi olla työläs prosessi, koska se sisältää aineistonkeräystä, väittelyä oikeasta mallista ja eri menetelmien toimivuuden kokeilemistä. Mittari ja sen koodaus voidaan nähdä hypoteesina tutkittavasta aiheesta, jolloin tutkimus jossa mittaria käytetään, toimii tämän hypoteesin testaajana. Jos hypoteesi on vahva, tulisi tulosten olla selvästi tulkittavissa. Jos tulokset taas ovat epäjohdonmukaisia tai hämmentäviä, tulisi mittaustilanteen toimivuuden lisäksi pohtia sitä, onko mittarin asettama kysymys havainnoitavasta asiasta selkeä ja sopiiiko mittarin rakenne mittarin kysymyksenasetteluun. (Bakeman & Gottman, 1986, s. 46–47.)

Tässä aineistossa usean mittarin kohdalla kategorioita on paljon, jolloin yhteneväisyyden saavuttaminen on vaikeampaa, koska mittarin käyttäjä joutuu valitsemaan usean, tilanteeseen sopivan vaihtoehdon välillä. Toisaalta joillakin mittareilla kategorioita on liian vähän, ja havainnoitava asia näyttäisi olevan liian dynaaminen määriteltäväksi yksinkertaisesti yhteen kategoriaan. Tulosten valossa mittareiden toimivuutta tulisi arvioida eri näkökulmista.

Ensinnäkin tulisi pohtia sitä, onko kaikkia mittareita mahdollista käyttää luotettavasti nykyisellä koulutuksella ja aineistonkeräysmenetelmällä. Mittareiden uudelleenkalibroimista tulisi harkita, ja lisätä tai poistaa kategorioita vastaamaan paremmin havainnoitavan aiheen luonnetta. Koulutuksen sisältöjä ja kestoja tulisi arvioida uudelleen, ja kiinnittää erityisesti huomiota sitoutuneisuuden ja sosiaalisen orientaation havainnointiin. Lopuksi olisi ehkä aiheellista pohtia observoijilta jonkinlaista vähimmäisvaatimusta mittareiden yhtenevässä käytössä ennen kenttäobservointien aloittamista.

8 Kappan käyttöön liittyvät ongelmat

Kappaan liittyviä ongelmia on tutkittu laajalti (ks. Vach, 2005; Siren, 2008; Feinstein & Cicchetti, 1990), ja tutkimukset osoittavat kappan käyttäytyvän epä johdonmukaisesti tietyissä tilanteissa. Kappan ongelmallisuus näyttäytyi myös tässä aineistossa kahden tyyppisissä tapauksissa.

8.1 Merkintöjen esiintyvyys ja reunafrekvenssit

Graham ym. (2012) toteavat kappan heikkoutena sen, kuinka kerroin voi olla harhaanjohtavan matala tilanteissa, joissa iso osa merkinnöistä sijoittuu mittarin ääripäihin. Samaten he huomauttavat kappan saattavan olla matala, jos iso osa merkinnöistä sijoittuu vain yhteen mittarin kategoriaan. (Graham ym., 2012, s. 8–9.) Jos siis esimerkiksi molempien observoijien merkinnöistä puuttuisi jokin mittarin kategoria kokonaan, saattaisi se madaltaa yhteneväisyyttä. Paradoksaalisesti kappaa voi siis olla huonompi, jos parien merkinnät ovat ”väärällä” tavalla samankaltaisia.

Tässä aineistossa oli yksi tapaus, jossa korkea yhteneväisyys johti täysin vastakkaiseen tulokseen ja kappan arvoon -0.1 (taulukko 2, D4). Kyseisen parin merkintöjen painottuminen vahvasti yhdelle kategorialle vaikutti sattuman merkityksen korostumiseen kappan kaavassa tuottaen erityisen matalan kertoimen. Parin merkintöjen ristiintaulukointi osoitti, että kaikki 51 yhteneväistä merkintää oli yhdessä kategoriassa (*ei-kontaktia*), ja kaksi eriävää merkintää oli eri kategorioissa (*lapsi 3* ja *lapsi 4*). Näin ollen merkintöjen esiintyvyyden epätasaisuus johti todennäköisesti eriävien merkintöjen ja sattuman

merkityksen vääränlaiseen korostumiseen yhtälössä. Erikoinen tulos johtui siis mittarin rakenteesta, arvioijien merkintöjen painottumisesta ja käytettävien kategorioiden esiintyvyyden epätasaisuudesta.

Byrt, Bishop ja Carlin (1993, s. 425) ovat esittäneet PABAK (Prevalence Adjusted Bias Adjusted Kappa) -version kappasta yllä mainitun ongelman korjaamiseksi, mutta versiota ei löydy SPSS-ohjelmasta. Niin sanottu yleisyysongelma (engl. prevalence) oli lopulta vakava vain aikaisemmin mainitun yhden luvun kohdalla, joten se päätettiin jättää tässä tutkimuksessa pois laskettaessa parien ja mittarien keskiarvoja.

Kappa-yhtälössä yhteneväisyyteen vaikuttaa siis yhteneväisten merkintöjen lisäksi se, miten mittarin merkinnät jakautuvat. Kappan toinen ongelma liittyy aikaisemmin mainittuihin reunafrekvensseihin, ja siihen miten jo valmiiksi alhainen yhteneväisyys yhdistettynä reuna-arvojen epätasaiseen symmetriaan observoiden välillä laskee kerrointa entisestään (taulukko 2, F6). Feinstein ja Cicchetti (1990, s. 548) toteavat, että reunafrekvenssien erilainen jakautuminen voi tuottaa jopa puolet pienemmän kappan, vaikka yhteneväisten merkintöjen määrä, eli $Pr(a)$ kaavassa 2, olisi kahdella parilla sama.

Kaavassa 2 $Pr(e)$, eli satunnaisuudesta johtuvan yhteneväisyyden määrä, muodostuu reunafrekvensseistä. Reunafrekvenssit taas voivat sijoittumisellaan kasvattaa tai pienentää $Pr(e)$ tulosta. Sattuman merkitykseen yhtälössä vaikuttaa siis se, mihin kategorioihin eriävät merkinnät sijoittuvat. Feinstein ja Cicchetti (1990) huomauttavat, että kahden arvioijan eriävien merkintöjen reunafrekvenssien ollessa epäsymmetriset saattaa kappalla olla korkeampi verrattuna tilanteeseen, jossa eriävien merkintöjen reunafrekvenssit olisivat symmetriset. Epäsymmetrisyydellä tarkoitetaan sitä, että toisen arvioijan reunafrekvenssit eroavat toisen arvioijan reunafrekvensseistä kontingenssitaulukossa. (Feinstein & Cicchetti, 1990, s. 546.) Käytännössä tämä tarkoittaa sitä, että huomattava eriävien merkintöjen erilaisuus pienentää sattuman vaikutusta yhtälössä tuottaen paremman tuloksen. Eli jälleen paradoksaalisesti kappalla huononee, koska merkinnät ovat väärällä tavalla samankaltaisia.

Tässä aineistossa ongelma näyttäisi koskevan niitä parikertoimia, joissa lähtökohtaisesti parien yhteneväisyys kyseisellä mittarilla on jo matala. Kun parit tekevät yksittäisiä eriäviä

merkintöjä eri kategorioihin, jo valmiiksi matalat kertoimet laskevat entisestään. Alla oleva taulukko 16 havainnollistaa ongelmaa.

Taulukko 16

Reunafrekvenssien vaikutus kappan kokoon

	Suru	Ilo, riemu	Hämmästys	Neutraali	Muu tunne	Ob2 f_i
Ärtymys	-	-	-	1	-	1
Pelko	-	-	-	1	-	1
Ilo, riemu	-	-	1	-	-	1
Onni	-	2	-	1	1	4
Hämmästys	-	1	-	-	-	1
Neutraali	1	-	4	29	-	34
Ob1 f_i	1	3	5	32	1	42

Huom. Solun luku on f .

Taulukossa 16 on ristiintaulukointi parin numero 3 merkinnöistä mittarilla *G. Lapsen emotionio*. Taulukosta 2 nähdään, että parin kappa tällä mittarilla oli .18. Taulukon 16 sarakkeet ovat observoijan 1 merkintöjä ja rivit observoijan 2 merkintöjä. Taulukossa on keskeistä se, mihin ei-yhteneväiset ja yhteneväiset merkinnät sijoittuvat. Esimerkiksi taulukossa on vain yksi merkintä kategorioissa *suru*, *ärtymys* ja *pelko*. Merkinnät ovat eri kategorioissa ja eri observoijilla, joten molempien observoijien sattuman reuna-arvot yhtälössä kasvavat eri suuntiin. Eriävien merkintöjen jakautuneisuuden lisäksi parin kaikki yhteneväiset merkinnät ovat *neutraali*-kategoriassa, jolloin havaitun yhteneväisyyden vaikutus yhtälössä ei ole niin suuri kuin jos samat merkinnät olisivat useammassa kategoriassa.

Yleisyysongelman tavoin reunafrekvenssiongelma oli lopulta vakava vain yhden kappan kohdalla, ja se jätettiin pois laskuista. Kertoimen vääristyminen kuitenkin osoitti, että Cohenin kappan painottamaton versio ei välttämättä ollut paras vaihtoehto kaikkien tämän aineiston mittareiden yhteneväisyyden laskemiseen. Eri vaihtoehtojen kokeileminen koko

aineistolle ei ollut enää tässä vaiheessa tutkimusta realistista, varsinkin kun kappan eri versiot on laskettava SPSS-ohjelmassa yksi kerroin kerrallaan. Matalien kertoimien kohdalla pystyi kuitenkin melko helposti kokeilemaan, miten ne reagoivat kappan eri versioihin.

8.2 Kappan eri versioiden vaikutus mataliin tuloksiin

Kappan eri versioiden tulokset ovat alla taulukossa 17. Reunafrekvenssien symmetrioiden vaikutusta tuloksiin ei ollut mahdollista testata tällä aineistolla, sillä testaus olisi vaatinut reunafrekvensseiltään symmetrisesti erilaista mutta muuten identtistä aineistoa.

Taulukko 17

Kappan eri versioiden vaikutus alhaiseen kertoimeen

Mittari/pari	Cohenin painottamaton kappa	Cohenin painotettu kappa	Fleissin kappa	Yhteneväisyys %
E3	.14	.23	.04	41
F3	.15	.20	.13	41
G3	.18	.18	.17	69
I3	.11	.31	.10	41
F5	.18	.35	.16	36
H5	.19	.36	.11	61
F6	.04	.17	.03	39
G7	.19	.23	.18	41
F8	.13	.26	.09	32

Taulukosta 17 nähdään, että painotetun kappan (Cohen, 1968) käyttö nostaa melkein kaikkia kertoimia jonkin verran. Kolmella parilla tuloksen kasvu vaikuttaa sen asteen tulkintaan hieman vahvempana, mutta yhteneväisyyden aste on silti vain kohtalainen.

Fleissin kappa (Fleiss, 1971) ei selvästi olisi ollut oikea versio tälle aineistolle, sillä usean parin kappa itse asiassa pienenee. Ainakaan taulukossa 17 olevien parien kohdalla kappan toisen version valitseminen ei siis olisi muuttanut sitä, että parien yhteneväisyys oli selvästi alle hyväksyttävän tason.

Prosentuaalinen yhteneväisyys taulukossa 17 on laskettu havainnollistamaan kahden laskentatavan eroja. Esimerkiksi parin G3 yhteneväisyys oli 69 %, ja olisi houkuttelevaa todeta parin yhteneväisyys kohtalaiseksi. Katsomalla kuitenkin taulukkoa 16 nähdään, että sama pari merkitsi usean tilanteen eriävästi ja oli yhteneväinen vain neutraaleissa merkinnöissä. Eli lopulta he ovat olleet suurimman osan ajasta vain sitä mieltä, että lapsen tunnetta ei voi tulkita suuntaan tai toiseen. Tämä osoittaa, kuinka pelkkä prosentuaalinen yhteneväisyys jättää huomiotta paljon tietoa ja saattaa helposti vääristää tutkittavaa aihetta.

8.3 Otoksoon vaikutus kappaan

On mahdollista, että parien kappa-kertoimiin saattaa vaikuttaa myös pieni otoskoko. Nummenmaa (2004, s. 280) huomauttaa, että tulomomenttikertoimet ovat erityisesti pienissä otoksissa herkkiä poikkeaville havainnoille. McHugh (2012, s. 281) toteaa yleisenä sääntönä, että verrattaessa kahta aineistoa otoksen tulisi käsittää yli 30 vertailtavaa paria. Tässä aineistossa kaikilla pareilla oli merkintöjä 35–56, joten McHugh'n määritelmä täyttyy.

Otoksoon vaikutusta tuloksiin testattiin luokittelemalla aineisto uudelleen. Aineistosta yhdistettiin kaksi alhaisen kappan saanutta paria niin, että SPSS-ohjelma laski nämä merkinnät yhden parin merkintöinä. Tällainen aineisto ei siis ollut totuudenmukainen, mutta simuloi tilannetta, jossa heikosti yhteneväinen pari olisi suorittanut enemmän observointeja. Yhdistetyt parit olivat F3 (.15) + F5 (.18) ja G3 (.18) + G7 (.19) taulukossa 2. Yhdistettynä parin F3 + F5 kappa oli .17, joten yhteneväisyys olisi ollut samalla tasolla, vaikka merkintöjen määrä olisi ollut kaksinkertainen. Sen sijaan yhdistettynä parin G3 + G7 kappa oli .24, eli hieman korkeampi. On siis mahdollista, että aineiston pieni koko aiheutti herkkyyttä poikkeavuuksille, mikä vaikutti osaan matalista kertoimista tässä aineistossa. Tämän selvittäminen tarkemmin vaatisi lisätutkimusta.

Kappan ongelma on Cicchettin ja Feinsteinin (1990) mukaan sen sisäänrakennettu tavoite luoda standardisoitu tapa vertailla eri tutkimusten havaintojen varianssia. Kuten muutkin niin sanotut kokoavat kertoimet, kappa tarjoaa yhteenvedon aineiston yhteneväisyydestä yhden luvun muodossa. Yksi luku voi riittää joissain tapauksissa, mutta se ei anna lisätietoa siihen vaikuttavista syistä. (Cicchetti & Feinstein, 1990, s. 553.) Cicchetti ja Feinstein (1990) esittävät ratkaisuksi p_{pos} ja p_{neg} arvojen laskemista kappan yhteydessä. Arvot mahdollistaisivat havaintojen ääriarvojen tarkastelun sekä tässä luvussa mainittujen paradoksien välttämisen. (Cicchetti & Feinstein, 1990, s. 554.) Kaikkien yhteneväisyyteen vaikuttavien muuttujien huomioon ottaminen vaatisi siis laajempaa, kontrolloitua aineistoa, useampia menetelmiä sekä laajempaa ymmärrystä menetelmien hyödyistä ja rajoitteista.

Lopulta valinta painottamattoman tai painotetun kappan välillä riippuu siitä, mitä tuloksilla halutaan tehdä. Painotettu kappa kertoo, ovatko eriävät merkinnät lähellä vai kaukana yhteneväisistä merkinnöistä. Tämä on hyödyllistä tietoa niiden mittareiden kohdalla, joissa samantyyppiset kategoriat ovat lähellä toisiaan. Samaten tieto on hyödyllistä, jos mittarilla havainnoidaan yhtä ilmiötä eri asteisena. Painotetun kappan avulla voidaan arvottaa jokin vaihtoehto ”vähemmän vääränä” kuin toinen. Esimerkiksi *E. Lapsen fyysinen aktiivisuus* -mittarilla tämä olisi nostanut yhteneväisyyttä, sillä useat observoijat olivat eriävissä merkinnöissä arvioineet lapsen fyysisen aktiivisuuden yhden tason alempana tai ylempänä.

Tämän aineiston kohdalla painottamaton kappa oli kuitenkin johdonmukaisempi siinä mielessä, että merkintöjä kohdeltiin samalla tavalla kuin hankkeessa. Hanke tarkastelee merkintöjä vain yksiulotteisesti, ilman tietoa siitä, olisiko merkintä ollut lähellä toista vaihtoehtoa. Eli totuudenmukaisinta on tarkastella yhteneväisyyttä yhtä ”armottomasti” joko yhteneväisenä tai eriävänä. Hankkeen kannalta on toisaalta paikallaan pohtia tämän tutkimuksen tulosten valossa, onko observoinnin toteuttaminen ja tulosten tulkitseminen nykyisessä yksinkertaisessa muodossaan paras ratkaisu. Samaten tilanteessa, jossa tavoitteena olisi kalibroida ja suunnitella mittareita paremmiksi, auttaisi painotettu kappa kertomaan, miten mittarin eri osa-alueet toimivat ilmiön havainnoinnissa.

9 Johtopäätökset

Tieteellisen tutkimuksen tavoitteena on tuottaa uutta tietoa. Tieto on uutta myös silloin, kun tutkija yhdistää jo olemassa olevaa tietoa uudella tavalla tai osoittaa uusia tapoja käyttää tietoa jonkin toiminnan kehittämisessä. (Vilkkä, 2015, s. 33.) Tässä tutkimuksessa tuotettiin uutta tietoa lapsihavainnoinnin yhteneväisyydestä, ja tiedolla on toivottavasti merkitystä hankkeen, varhaiskasvatuksen tutkimuksen ja havainnoinnin luotettavuuden tutkimuksen kannalta. Tämän tutkimuksen tieto on tuotettu järjestelmällisesti, ja menetelmien hankkimiseen ovat olleet tietoisesti valittuja ja perusteltuja. Tutkimuksen otanta oli perusjoukkoon verrattuna pieni mutta monipuolinen, ja sen voidaan nähdä edustavan satunnaista kuvausta hankkeen observoijista.

Aineistonkeruutilanteeseen vaikuttaneita muuttujia ei voi tutkimusasetelman takia tarkasti määrittellä, ja tämä tulisi ottaa huomioon tulosten yleistettävyydessä. Emme myöskään voi varmuudella tietää jokaisen aineistomerkin totuudenmukaisuuden astetta, koska aineisto koostuu 18 eri ihmisen merkinnöistä. Aineiston samankaltaisuus kuitenkin puoltaa oletusta, että aineisto ei ole ollut satunnaista: esimerkiksi lapsen numero oli sama molempien observoijien merkinnöissä. Voidaan siis olettaa, että observoijat ovat toimineet ohjeistuksen mukaisesti ja toteuttaneet pariobservointia toivotulla tavalla. Mittauksen reliabiliteettiin liittyy useita tekijöitä, ja tämä tutkimus keskittyy vain havainnoijien yhteneväisyyteen rajatusta aineistosta ja rajatuilla menetelmillä. Aineistoon ja menetelmiin liittyi paljon muuttujia. Kaikkia niitä ei ollut mahdollista ottaa huomioon tässä tutkimuksessa, koska se olisi vaatinut edistyneempää tietoa reliabiliteetilaskennasta, kvantitatiivisista menetelmistä sekä aineistoon vaikuttavista muuttujista.

Cohenin kappan käyttö tässä tutkimuksessa ei ollut ongelmaton, ja joissakin tilanteissa observoinnin rakenne ja mittareiden merkintöjen jakautuminen vaikuttivat kappan kokoon. Ongelma koski kuitenkin vain osaa aineistosta, ja suurin osa tuloksista olisi sijoittunut samalle kohdalle arviointiasteikkoa kappan versiosta riippumatta. Kappan rajoitukset tulisi ottaa huomioon tulevissa tutkimuksissa, ja painottaa usean eri menetelmän käyttöä tasapainoisen tuloksen takaamiseksi. Kappan jo aikaisemmin esiteltyjen eri versioiden lisäksi viimeisessä luvussa tarkastellaan kahta menetelmää mittareiden reliabiliteetin ja validiteetin tutkimiseen.

Kehittävä palaute -hankkeen tavoitteena on saada observoinnin kautta luotettavaa tietoa lasten yksilöllisestä toiminnasta. Pelkästään tämän tutkimuksen perusteella ei voida määrittellä hankkeen observoinnin kokonaisreliabiliteettia, joka on moniulotteinen käsite. Tulosten perusteella voi kuitenkin todeta, että observoijien yhteneväisyydessä on selviä eroja.

Sekä observoijien että mittareiden luotettavuudessa on epätasaisuutta, ja epätasaisuus on joillakin osa-alueilla merkittävää. Mittarit, jotka tarkastelevat selkeitä, vähemmän dynaamisia asioita kuten lapsen päiväjärjestystä, tuottavat myös selkeämpää tietoa. Useat observoinnin sisällöt pohjautuvat kuitenkin abstrakteihin käsitteisiin, joiden konkreettinen määrittely käytännössä näyttäisi olevan vaikeaa. Joillakin mittareilla on pyritty toistamaan monimutkaisia sitoutuneisuuden tai fyysisen aktiivisuuden havainnointimenetelmiä yksinkertaistettuna, mikä on johtanut merkintöjen epätarkkuuteen. Tulosten perusteella näyttäisi siis siltä, että tiedonhankintamenetelmät ovat epätasapainossa sen tarkkuuden kanssa, jolla tietoa halutaan tarkastella. Tasapainon saavuttamiseksi on olemassa useita eri ratkaisuja, ja lopulta iso osa epätarkkuudesta voitaisiin todennäköisesti korjata lisäkoulutuksella ja mittareiden kalibroimisella.

Graham ym. (2012) esittelevät useita havainnoinnin yhteneväisyyden kehittämisen kannalta keskeisiä tekijöitä. He pitävät tärkeänä havainnoijien pitkäkestoista ja monipuolista koulutusta, sillä sen kautta havainnoijille muovautuu yhteinen käsitekehys (engl. frame of reference). Käsitekehys vähentää merkintöjen epätarkkuutta, eriävyyttä ja ennakoasenteita. Toisena tekijänä he nostavat esiin menetelmän jatkuvan laaduntarkkailun, joka sisältää esimerkiksi havainnoijien yhteneväisyyden kartoittamista ja uudelleen kouluttamista. Kaikkein tärkeimpänä he kuitenkin pitävät sitä, että menetelmää testataan monipuolisesti ennen sen käyttöönottoa, jolloin keskeiset heikkoudet voidaan huomioida ja korjata. (Graham ym., 2012, s. 15–21.) Tämän tutkimuksen tulosten perusteella yllä mainitut tekijät tulisi ottaa huomioon hankkeen observoinnin kehittämisessä.

Tulosten pohdinta -osiossa on esitelty mittarikohtaisia kehitysehdotuksia tuloksiin perustuen. Yleisenä kehitysehdotuksena kaikille mittareille on *en tiedä* -kategorian lisääminen. Tällä hetkellä osassa kategorioista tämän tyyppinen vaihtoehto on, osassa ei, ja

osassa jokin muu vaihtoehto saattaa tietämättämme toimia sellaisena. Tämä olisi alkeellinen mutta helppo tapa erotella aineistosta epävarmat merkinnät.

On selvää, että observoinnin harjoittelua tulisi lisätä yhteneväisyyden parantamiseksi. Samalla olisi hyvä pohtia, tulisiko observoijan saavuttaa tietty yhteneväisyyden raja harjoitteluaineistossa ennen observoinnin suorittamista kentällä kuten esimerkiksi CLASS- tai ECERS -menetelmien kohdalla (ks. Ishimine & Tayler, 2014, s. 281; Sheridan, 2000, s. 69). Esimerkiksi Semmelroth ja Johnson (2014, s. 132) toteavat, että pätevyysvaatimukset havainnoinnin suorittamiselle ovat keskeisessä osassa luotettavien tulosten aikaansaamista.

Suosituksat vaadittavasta kappan arvosta vaihtelevat lähteestä riippuen, ja lopulta arvo tulisi määritellä tutkittavan aiheen perusteella. *Kehittävä palaute* -hankkeessa pyritään havainnoimaan lasten toimintaa ja yksilöllisiä ominaisuuksia, joten hyväksyttävän yhteneväisyyden asteen asettamisessa tulisi olla vaativa. Jos tavoitteena on tarkastella observoinnin tuloksia kootusti ja yleistetysti, tulisi .80 olla alin hyväksyttävä kappan raja-arvo mittarille. Tällä hetkellä useat mittarit ja observoijat eivät saavuta hyväksyttävää kertoimia yhteneväisyydelle. Hankkeessa tulisi siis arvioida uudestaan observoinnin tavoitteita ja pohtia, miten yhteneväisyyden saisi hyväksyttävälle tasolle.

Yhteneväisyyden kannalta olisi parasta, jos observointia toteutettaisiin aina pareittain tai videoidun aineiston pohjalta. Esimerkiksi Semmelroth ja Johnson (2014, s. 132) toteavat useamman observoijan käytön olevan keskeistä luotettavuuden saavuttamisessa. Pariobservoinnin keskiarvolle voisi määritellä alarajan .80, jonka ylittyessä aineiston voitaisiin nähdä täyttävän luotettavan datan tunnusmerkit. Yhteneväisyydeltään alemmat observoinnit jätettäisiin pois, tai, jos ne sisällytettäisiin tuloksiin, merkittäisiin niiden kohdalla selvästi yhteneväisyyden aste. Alle .60 tuloksia ei käytettäisi perustuen McHugh'n (2012) kriteereihin taulukossa 15.

Toinen vaihtoehto tarkkuuden lisäämiseksi voisi olla observointitilanteiden videointi. Videoituja tilanteita voisi katsoa useampi observoija, ja he voisivat vertailla tuloksiaan ja keskustella valintojensa perusteista, kehittäen samalla omaa osaamistaan. Liian eriävät mielipiteet tarkoittaisivat, ettei tilannetta voi määritellä tarkasti. Molemmat menetelmät ovat nykyiseen verrattuna hitaita ja työläitä, mutta tässä tapauksessa laadun tulisi korvata määrä.

Kehittävä palaute varhaiskasvatuksessa on tärkeä projekti, jossa pyritään tarkastelemaan varhaiskasvatusta monimenetelmällisesti ja nostamaan samalla lapsen yksilölliset prosessit yhdeksi toiminnan kehitystä ohjaavaksi suuntaviivaksi. Projektin kautta hankitulla tiedolla voi olla suuri merkitys varhaiskasvatuksen kehittämässä, ja hankkeen historian aikana on jo nyt luotu poikkeuksellisen laaja ja pitkäkestoinen aineistonkeruumenetelmä. Aineistonkeruumenetelmän yhdellä osa-alueella eli observoinnilla on kuitenkin tämän tutkimuksen tulosten perusteella selviä kehittämiskohteita, jotka tulisi ottaa huomioon pohdittaessa hankkeen luotettavuuden lisäämistä.

10 Ehdotuksia havainnointipohjaisten menetelmien luotettavuuden tutkimukseen

Tämän tutkimuksen aineisto osoittautui hyvin riittoisaksi, ja useat tutkimuksen aikana esiin nousseet tienhaarat karsiutuivat pois niiden laajuuden takia. Nämä tienhaarat vaativat kuitenkin myös rinnalleen monipuolisempaa kvantitatiivista ja kvalitatiivista dataa, jonka keräämisessä tulisi käyttää useampaa eri menetelmää. Samalla nämä aiheet liittyvät ”vain” yksittäisen menetelmän kehittämiseen aineistopohjaisen tiedon kautta, eivätkä ne tarkastele laajemmin havainnointipohjaisten arviointimenetelmien luotettavuuden kehittämistä. Reliabiliteetin ja validiteetin tutkimiseen on kehitetty useita erilaisia malleja, ja nostan niistä esiin kaksi mielestäni mielenkiintoisinta. Molemmat mallit tarjoavat mielestäni omanlaisensa näkökulman havainnointipohjaisten menetelmien luotettavuuden tutkimukseen.

10.1 Hankkeen observoinnin laajempi tutkiminen

Tässä tutkimuksessa ei saatu tietoa observojien observointiin liittyvistä näkemyksistä tai siitä, millä tavalla taustatekijät vaikuttavat observojan näkemyksiin tai mittareiden käyttöön. Olisi ollut mielenkiintoista tutkia tarkemmin myös observojan sisäistä yhteneväisyyttä, eli sitä ovatko tietyt observoijat jatkuvasti enemmän yhteneväisiä kuin toiset. Tätä kautta olisi voitu saada lisätietoa havainnoinnin luotettavuuteen positiivisesti vaikuttavista taustatekijöistä.

Puhuttaessa tietyn menetelmän kehittamisestä on tärkeää tarkastella tuloksia mahdollisimman läheltä, sillä tärkeät erot ja painotukset katoavat, jos aineistoa tarkastelee vain yleisellä tasolla. Tässä tutkimuksessa parikohtaiset ristiintaulukoinnit olivat hyvin mielenkiintoisia, ja ne olisivat tarjonneet tarkemman näkökulman parikohtaisiin eroihin. Taulukot ja niitä avaavat tekstit olisivat kuitenkin nostaneet sivumäärän liian suureksi. Havainnointiajan vaikutusta observoinnin tarkkuuteen olisi ollut mahdollista tutkia vertailemalla yhteneväisten merkintöjen määrää suhteessa havainnointikellonaikaan. Observoinnin videointi olisi mahdollistanut sekä observoijan että observeitavan tilanteen yksityiskohtaisen tarkastelun ja antanut molemmista aiheista lisätietoa.

Tässä tutkimuksessa käytetty Cohenin kappa on vain yksi monista tavoista arvioida luotettavuutta. Kappan etu on sen tapa laskea sattuman vaikutus yhtälössä, mutta sen ongelmana voidaan nähdä merkintöjen sijoittumisesta johtuva epäjohdonmukaisuus tuloksissa. Kappan tulos on siis lopulta melko tapauskohtainen, ja kahden tutkimuksen tuloksia tulisi verrata rinnakkain vain, jos tutkimusasetelmat ovat hyvin samanlaiset. Luotettavuuden arviointiin ei ole yhtä standardia, ja voi olla, että tämän tutkimuksen tulokset olisivat olleet erilaiset, jos olisi käytetty jotain toista tapaa laskea yhteneväisyys. On myös mahdollista, että aineiston erilaisella koodauksella tai muuttujamuunnoksilla olisi voitu välttää kappan yhtälöongelmat.

Lopulta yhtälön valinta ja se, miten tuloksia tulkitaan, riippuvat siitä, mitä halutaan selvittää. Tässä tutkimuksessa aiheena olleet yhdeksän mittaria ovat rakenteeltaan erilaisia, joten mittarin kehittämiseen tähtäävässä tutkimuksessa voisi olla perusteltua valita käytettävä yhtälö mittarin mukaan. Esimerkiksi mittareilla *A*, *B* ja *C* painottamaton kappa näyttäisi olevan paras vaihtoehto, sillä mittarissa on erilaisia asioita sekalaisessa järjestyksessä. Mittarit *E* ja *F* puolestaan mittaavat yhtä asiaa eri asteisena, ja vastausvaihtoehdot ovat nousevassa järjestyksessä. Tässä tapauksessa painotettu kappa voisi olla järkevä, koska useat merkinnät sijoittuvat lähelle toisiaan. Painottamalla vaihtoehdot voitaisiin arvioida sitä, ovatko havainnoijat suunnilleen samaa mieltä vai täysin eri mieltä, mikä taas auttaisi mittarin kehittämisessä. Tämä voisi olla mielenkiintoinen jatkotutkimuksen aihe.

Puhuttaessa mittauksesta ja arvioinnista varhaiskasvatuksessa on mielestäni keskeistä kysyä ensisijaisesti ”mitä” ja ”miksi”, ennen kuin voidaan miettiä ”miten hyvin”. En siis

usko, että varhaiskasvatuksen kaikkia tärkeitä ulottuvuuksia voisi mitata jotenkin ”paremmin” jonkin menetelmän kautta; se voiko kaikkia tärkeitä ulottuvuuksia mitata alkujaankaan on oma, mielenkiintoinen aiheensa. Puhuttaessa jatkotutkimuksesta olisi kuitenkin erikoista olla mainitsematta seuraavia malleja. Mallit nousivat tutkimuksen aikana esiin useassa lähdeaineistossa, joten ne voidaan nähdä keskeisenä osana tutkimuksen aihetta eli havainnointipohjaisten menetelmien luotettavuudentutkimusta.

Yleistettävyysteoria on reliabiliteetin laskemiseen liittyvä malli, jonka keskeisiksi kehittäjiksi on usein (ks. Lakes & Hoyt, 2009, s. 146; Webb & Shavelson, 2005, s. 718; Musquash & Connor, 2006, s. 542) mainittu Cronbach, Gleser, Nanda ja Rajaratnam (1972). Argumenttipohjaisen validiteetin malli vuorostaan tarkastelee validiteetin rakentumista argumenttipohjaisen päättelyprosessin kautta, ja Kane (2006) tunnustetaan usein mallin keskeiseksi kehittäjäksi (ks. Cook, Brydges, Ginsburg, & Hatala, 2015, s. 560; Bell ym., 2012, s. 63). Molempien mallien juuret tuntuvat olevan syvällä kovin erilaisessa tutkimuksessa verrattuna suomalaisen varhaiskasvatukseen. Mallien lähdeaineistoissa puhutaan kokeista, testeistä ja pisteistä pääosin yhdysvaltalaisen peruskoulun kontekstissa. Ei siis tule olettaa, että alla olevat esimerkit olisivat suoraan siirrettävissä suomalaisen varhaiskasvatuksen kontekstiin. Tämän alustuksen myötä näen kuitenkin seuraavien mallien voivan olla hyödyksi myös varhaiskasvatuksen tutkimuksessa.

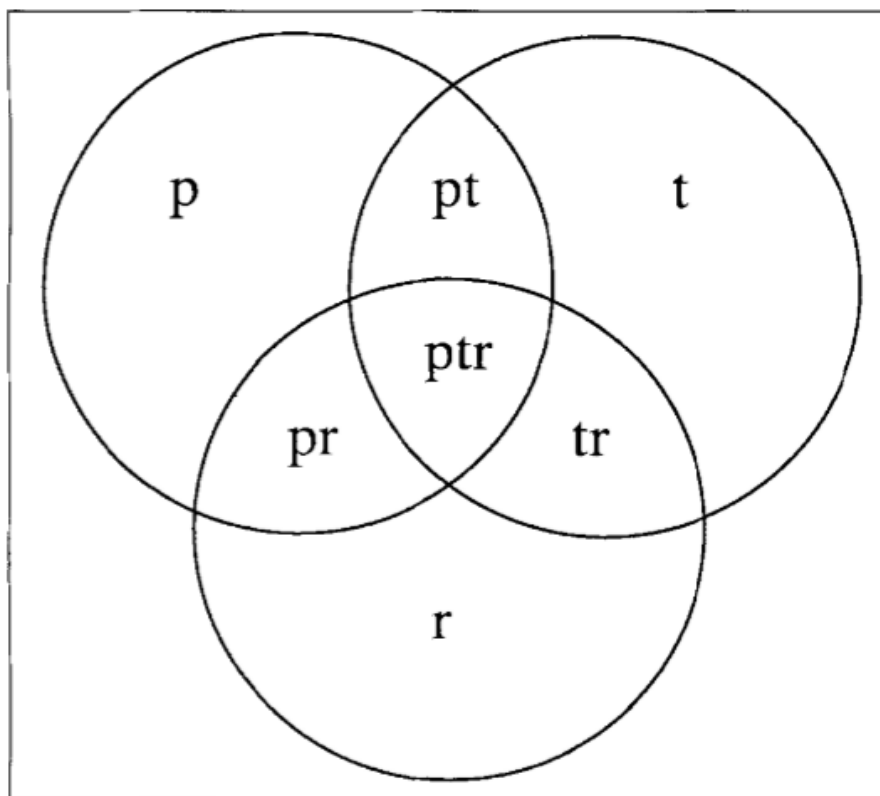
10.2 Yleistettävyysteoria

Yleistettävyysteoria on käyttäytymistä mittaavien menetelmien luotettavuuden tarkastelua varten luotu tilastotieteellinen teoria (Webb & Shavelson, 2005, s. 718). Teorian nimi pohjautuu näkemykseen tiedon luonteesta ja siitä, kuinka tiedon yleistettävyydelle ei ole vain yhtä yhteistä kehystä. Tutkijan on siis tutkimuksensa kautta määriteltävä se ulottuvuus, johon tutkimuksen tulokset voidaan nähdä yleistettävän. (Brennan, 2001, s. 9.)

Yleistettävyysteorian kehittymisen syynä voidaan Webbin ja Shavelsonin (2005) mukaan nähdä tässäkin tutkimuksessa käytetty klassinen testiteoria, ja erityisesti sen määritelmä mittausvirheestä E yhtälössä $X = T + E$. Kuten reliabiliteettia käsittelevässä luvussa todettiin, klassisessa testiteoriassa mittausvirhe E on syy siihen, että havainnoitua tulosta X ei voida suoraan laskea. Tätä ongelmaa on ratkaistu erilaisilla virhettä arvioivilla

kertoimilla kuten esimerkiksi tässäkin tutkimuksessa käytetyllä Cohenin kappalla. Tällainen tapa tarkastella mittausvirhettä ei kuitenkaan ota huomioon potentiaalisen ja todellisen virheen määrää, virheen lähteitä tai eri lähteiden suhteita. (Webb & Shavelson, 2005, s. 718.) Myös tässä tutkimuksessa ollaan tilanteessa, jossa voidaan tulosten perusteella todeta mittausvirheen olemassaolo aineistossa, mutta ei pystytä tarkkaan määrittelemään sen syytä, sijaintia tai mittakaavaa.

Yleistettävyysteorian kautta on mahdollista määritellä tarkemmin, missä kohtaa mittausprosessia mittausvirhettä on ja kuinka paljon mittauksen eri osa-alueiden sisällöt vaikuttavat toisiinsa ja mittarin kokonaisluotettavuuteen. Tämä saavutetaan jakamalla vaikuttavat muuttajat omiin osioihinsa (engl. facet), jolloin jokaista osiota ja niiden sisältämää mittausvirhettä voidaan mitata erikseen ja yhdessä. (Musquash & O'Connor, 2006, s. 542.) Osiolla tarkoitetaan tässä kontekstissa yhtä tai useampaa asiaa, joiden mitattavat ominaisuudet ovat samanlaiset (Brennan, 1992, s. 27). Alla olevassa *kuviossa 16* on tutkimusasetelma, jossa on määritelty mittausvirhettä sisältävät osat ja niiden risteävyydet yleistettävyysteorian mukaisesti.



Kuvio 16. Mittausvirheen osat (Brennan, 1992, s. 27)

Kuviossa 16 on määriteltynä mittausvirhettä aiheuttavat osat ja niiden suhteet. Brennanin (1992) esittämässä kuvitteellisessa tilanteessa tutkija haluaa luoda menetelmän kirjoitustaidon arviointiin. Kuvion kolme ympyrää kuvaavat mittauksen kolmea osaa: kirjoittajia p , kirjoittamisen arvioijia r sekä kirjoitustaitokäsitettä t . Ympyröiden risteyskohdat ovat tilanteita joissa nämä osat (ja niiden mittausvirhe) kohtaavat. Kuvion seitsemän aluetta kuvaavat niitä osioita mittauksessa, joissa mittausvirhettä syntyy. Keskellä oleva osio ptr kuvaa kaikkien osien yhteistä mittausvirhettä, sekä kuvion ulkopuolella olevaa, ei-määriteltävissä olevaa mittausvirhettä. (Brennan, 1992, s. 27–28.)

Yleistettävyysteoriassa tehdään ensin niin sanottu G-tutkimus (generalizability), jossa alkuperäistä aineistoa käyttämällä saadaan tietoa varianssin määrästä mittauksen eri osissa ja niiden välillä. G-tutkimuksesta saadun tiedon perusteella on mahdollista tehdä seuraavaksi D-tutkimus (decision), jossa voidaan tutkia, miten mittauksen osien muuttaminen vaikuttaa osien mittausvirheeseen. (Musquash & O'Connor, 2006, s. 543.) Yleistettävyysteorian kautta mittausvirheen syyt voidaan siis löytää ja määrittää tarkemmin

sekä arvioida eri muuttujien vaikutusta mittausvirheeseen. Käytännössä tällä tiedolla on mahdollista laskea esimerkiksi, johtuuko mittauksen virhe havainnoijasta vai havainnoinnin kohteesta, tai tarkastella, miten havainnoijien tai havainnoitavien kohteiden määrän muuttaminen vaikuttaa mittauksen luotettavuuteen.

Semmelroth ja Johnson (2014) tutkivat yleistettävyysteorian sopivuutta RESET-menetelmän luotettavuudenarvioinnissa. RESET on erityisopetuksen opettajien arviointiin tarkoitettu menetelmä, jossa havainnointi toteutetaan videoidun luokkahuonetilanteen tarkastelun kautta. Aikaisempien tutkimusten perusteella he toteavat perinteisten havainnoijien välistä yhteneväisyyttä mittaavien kertoimien olevan ongelmallisia RESET-menetelmän arvioinnissa. Kertoimet ovat matalia, sillä kertoimet tarkastelevat mittausvirhettä yhtenä massana erottelematta siihen liittyviä tekijöitä kuten kohde, tilanne tai aika. (Semmelroth & Johnson, 2014, s. 133.) Yleistettävyysteorian kautta luodussa kaksipuolisessa tutkimusasetelmassa Semmelroth ja Johnson (2014, s. 143) pystyivät tarkastelemaan eri tekijöiden (havainnoitavat, havainnointitilanteet, havainnoijat) vaikutusta havainnoinnin luotettavuuteen, ja laskivat tarvittavan havainnoijien ja havainnointikertojen määrän tyydyttävän reliabiliteetin saavuttamiseksi.

Semmelrothin ja Johnsonin tutkimuksen tavoin olisi mielenkiintoista tutkia yleistettävyysteorian sopivuutta erilaisten havainnointipohjaisten arviointimenetelmien kehittämisessä. Artikkeleiden perusteella yleistettävyysteoria voisi mahdollistaa paremman tavan arvioida havainnointipohjaisen menetelmän luotettavuutta. Pelkästään uudelleenmittaamisen tuloksiin perustuva palaute mittarin osa-alueiden luotettavuudesta on hapuilevaa ja työlästä, ja yleistettävyysteoria voisi tarjota järkevän tavan vertailla eri toteutustapoja osana menetelmän luotettavuuden tutkimista.

10.3 Argumenttipohjainen validiteetti

Sirecin (2009) mukaan validiteettia on kuvailtu psykometrisen tutkimuksen tärkeimmäksi käsitteeksi. Hänen mukaansa validiteetin käsitteellä viitataan mittaustuloksen pohjalta tehtyyn tulkintaan tai toimintaan, jonka arvioinnin tulisi perustua testin tarkoitukseen ja käyttötapaan. (Sireci, 2009, s. 20.) Käsitevaliditeetti nähdään usein muita validiteetin näkökulmia yhdistäväksi määritelmäksi (ks. Sireci, 2009, s. 25; Carmines & Zeller, s.

1979, s. 27), mutta sen yleistä sopivuutta tutkimuksessa on myös kritisoitu (ks. Borsboom, Cramer, Kievit, Scholten, & Franić, 2009). Perinteinen, tässäkin tutkimuksessa käytetty, käsitevaliditeetin näkökulma ei Bellin ym. (2012) mukaan välttämättä ole riittävä tarkasteltaessa opetuksen laatua arvioivien havainnointipohjaisten menetelmien validiteettia. Heidän mukaansa havainnointipohjaisten menetelmien rakenteet, pisteytys ja arviointi saattavat erota huomattavasti perinteisistä laadunarviointimenetelmistä, jotka perustuvat yksilön tuloksien arviointiin. (Bell ym., 2012, s. 63.)

Kane (2016) näkee validiteetin määrittelyn olevan riippuvainen kahdesta asiasta: tuloksien ehdotetusta tulkinnasta ja tuloksien ominaisuuksista. Puhuessaan argumenttipohjaisesta validiteettimallista (Kane, 1992; 2006; 2017) hän toteaa, ettei tavoitteena ole niinkään määrittää tai uudelleenmäärittää validiteettia, vaan tulkita ja selvittää validiteetin käsitettä. (Kane, 2016, s. 198–199.) Argumenttipohjainen validiteettimalli koostuu kahdesta osasta: tulkinta- tai käyttöargumentista (interpretation/use argument) sekä sitä seuraavasta validiusargumentista.

Kaksiosaisessa mallissa tulkinta-argumentit pitävät sisällään ne päätelmät ja oletukset, joita saaduista tuloksista voidaan tehdä. Validiusargumentti puolestaan arvioi näitä päätelmiä ja oletuksia empiirisen tiedon ja analyttisen järkeilyn kautta. Kane on määritellyt neljä keskeistä tulkinta-argumenttia, joiden kautta valitun menetelmän validiutta voidaan lähteä arvioimaan tuloksiin perustuen: pisteytys (tehdyn havainnon muuttaminen pisteiksi), yleistettävyyys (kuinka hyvin annetut pisteet kuvaavat tilannetta), yhdistettävyyys (kuinka hyvin pisteet kuvaavat jo todistettua) sekä johtopäätökset (pisteiden käyttö). Jokaista tulkinta-argumenttia kohden tulisi olla sitä tukevia todisteita eli validiusargumentteja. (Bell ym., 2012, s. 65.)

Toisin sanoen validoinnissa on Cookin ym. (2015) mukaan kyse niiden väitteiden, tulkintojen ja päätelmien arvioinnista, jotka yhdistävät saadut tulokset niistä tehtäviin johtopäätöksiin. Tätä validointiprosessia voi verrata oikeudenkäyntiin, jossa asianajajan on vakuutettava valamiehistö esitetyn todistusaineiston ja siihen liittyvän argumentoinnin kautta. (Cook ym., 2015, s. 560–562.) Argumenttipohjaista validiteettimallia voi Kanen (2016, s. 202) mukaan soveltaa monipuolisesti eri aiheisiin, ja silti tarkastella valitun aiheen validiteettia perusteellisesti. Alla olevassa *kuviossa 17* on Bellin ym. (2012, s. 66) muokattu esimerkki (olen poistanut tutkimuskohtaisia sanoja) argumenttipohjaisesta

validiteettimallista tutkimuksessa, jossa aiheena oli opetuksen laatua ja opettajan toimintaa arvioivan havainnointipohjaisen CLASS-S-menetelmän validiteetti. Kuvion tarkoituksena on konkretisoida argumenttipohjaisen mallin lähtökohtia validiteetin arvioinnille.

1. Pisteytys
 - 1.1 Pisteytysjärjestelmä on sopiva
 - 1.2 Pisteytysjärjestelmää käytetään tarkasti ja johdonmukaisesti
 - 1.3 Pisteytys on puolueetonta
 - 1.4 Aineisto sopii kokonaisuuteen
2. Yleistettävyys
 - 2.1 Aineisto kuvaa riittävästi yleistä oppituntien laatua
 - 2.2 Satunnainen virhe on riittävästi huomioitu
3. Yhdistettävyys
 - 3.1 Tuloksen yhteys mitattavaan aiheeseen on todettavissa
 - 3.2 Ei ole systemaattista virhettä joka estäisi tulosten yleistyksen
4. Johtopäätökset
 - 4.1 Johtopäätökset ovat sopivia
 - 4.2 Mitatut ominaisuudet tukevat päätelmiä

Kuvio 17. Argumenttipohjainen validiteetti tutkimuksessa (Bell ym., 2012, s. 66 muokattu)

Kuviossa 17 on Bellin ym. (2012) versio argumenttipohjaisesta validiteettimallista. Kuviossa esitetyt päätelmät ja niiden sisällöt eivät ole pysyviä ominaisuuksia, ja menetelmän arvioinnin runko määrittyy menetelmän sisältöjen mukaan. Tavoitteena on esittää tulkintoja menetelmän validiudesta näiden väittämien perusteella ja myöhemmin esittää näitä tulkintoja tukevia argumentteja. Pisteytys-otsikon alla on pisteytyksen validiteettiin vaikuttavat argumentit: sopivuus (käytetäänkö järjestelmän kaikki ulottuvuuksia niin kuin on tarkoitettu), tarkkuus (käytetäänkö pisteitä oikein ja johdonmukaisesti), puolueettomuus (pisteytykseen vaikuttavien asioiden huomioonottaminen) ja sopivuus (aineiston sopivuus osana isompaa mitattavaa kokonaisuutta tai muita pisteitä). Yleistettävyys-otsikon argumentit liittyvät siihen, kuinka hyvin aineisto kuvaa mitattavaa asetelmaa sekä siihen liittyviä tiedossa olevia muuttujia. (Bell ym., 2012, s. 66–67.)

Yhdistettävyyso-otsikon alla *kuviossa 17* ovat ne argumentit, joiden perusteella saadut tulokset voidaan nähdä yhteydessä aikaisempaan tietoon tutkittavasta aiheesta. Bell ym. (2012) käyttävät esimerkkinä opetuksen laadun eri määritelmiä ja sitä, ovatko testattavan menetelmän sisällöt yhdistettävissä vain yhteen vai useampaan määritelmään opetuksen laadusta. Samaten yhdistettävyydessä tulee ottaa huomioon se, kuinka testin tulokset ottavat huomioon mitattavan ominaisuuden systemaattiset vaihtelut. Lopulta johtopäätökset-otsikon alla ovat ne argumentit, joiden perusteella voidaan olettaa tuloksien perusteella tehtyjen johtopäätösten olevan sopivia. (Bell ym., 2012, s. 68.)

Esiteltyjä argumentteja käytetään lähtökohtana monimenetelmälliselle analyysille, joka vahvistaa tai heikentää jokaista argumenttia menetelmän sopivuudesta. Bell ym. (2012, s. 63) toteavat, ettei tutkimuksen tavoitteena ole arvioida CLASS-S-menetelmää itsessään, vaan tarkastella argumenttipohjaisen validiteettimallin mahdollisuuksia havainnointiin perustuvan menetelmän arvioinnissa. En käy tutkimuksen argumentointiosiota läpi, mutta nostan esiin pari esimerkkiä prosessista.

Pisteytysosiossa Bell ym. (2012) toteavat argumenttien olevan vahvoja liittyen pisteytysjärjestelmän sopivuuteen, sillä menetelmä on laajasti käytetty ja huolellisesti kalibroitu. Argumentit pisteytyksen johdonmukaisuudesta ja luotettavuudesta taas ovat heikompia, sillä havainnoijien välinen yhteneväisyys on useassa kategoriassa alle toivotun tason. Yleistettävyyso-otsiossa he vuorostaan toteavat menetelmän olevan siihen liittyvien argumenttien osalta puutteellinen määrittelemättömän mittausvirheen koon ja aineiston otantaan liittyvien tulosten vuoksi. (Bell ym., 2012, s. 78–80.) Kaikkia *kuviossa 17* esitetyjä argumentteja tarkastellaan tämän tyyppisesti käyttäen muun muassa faktorianalyysia, yleistettävyysteoriaa ja aikaisempaa tutkimustietoa argumenttien vahvuuden selvittämiseksi.

Yhteenvedossaan Bell ym. (2012) toteavat, että argumenttipohjaisen validiteettimallin rakenne pakottaa tutkijat tarkastelemaan oman menetelmänsä suhdetta havainnoinnin kohteisiin, havainnoinnin välineisiin sekä prosessiin vaikuttaviin muuttujiin. Tämän lisäksi malli korostaa heidän mielestään validiteetin arvioinnissa implementointiin liittyviä tekijöitä, kuten havainnoijien koulutusta ja aineiston otantamenetelmiä. Malli myös nostaa selkeästi esiin menetelmän perustelut ja väitteet sen validiteetista, ja täten se on helposti

myös muiden tarkasteltavissa. He kritisoivat argumenttipohjaista validiteettimallia siitä, että malli on työläs toteuttaa ja se vaatii perusteellista analyysia argumenttien arvioimiseksi. Lopuksi he ehdottavat mallin sopivan parhaiten sellaisten menetelmien arvioimiseen, joista on jo aikaisempaa tutkimustietoa. (Bell ym. 2012, s. 84–85.)

Argumenttipohjaisen validiteettimallin mahdollisuudet havainnointipohjaisten laadunarviointimenetelmien kehittämisessä on mielestäni mielenkiintoinen, joskin työläältä vaikuttava tutkimusaihe. Kuten Bell ym. (2012) totesivat, mallin toteuttaminen vaatii monenlaista dataa ja analyysia sekä pitkäaikaista paneutumista tutkittavan menetelmän teoreettiseen taustaan. Olisi kuitenkin mielenkiintoista tarkastella, miten esimerkiksi Suomessa käytössä olevat havainnointipohjaiset menetelmät käyttäytyisivät mallin tarkastelun kohteena. Toinen mahdollinen tutkimuskohde, jossa argumenttipohjainen malli voisi olla hyödyllinen, olisi ulkomailla jo käytössä olevien menetelmien sopivuuden arviointi suomalaisen varhaiskasvatuksen kontekstissa. On myös mahdollista, että mallia voisi käyttää myös osissa. Jo argumenttien alustava tarkastelu saattaa näet nostaa esiin menetelmän keskeiset vahvuudet ja kehittämiskohteet. Tällä tavalla malli toimisi siis ikään kuin menetelmän ”valmentaja”.

10.4 Yhteenveto

Hill, Charalambous ja Kraft (2012) tarkastelevat tutkimuksessaan opetuksen laatua arvioivia havainnointimenetelmiä. He painottavat, että havainnointimittareiden kehittämisen sijaan tulisi puhua havainnointisysteemeistä ja avaintekijöistä niiden kehittämisessä. Näissä systeemeissä yhdistyvät heidän mukaansa laadukkaat havainnoinnin mittarit, hyvin koulutetut havainnoijat ja vakaat tulokset. Mittareiden kehittämisessä keskeistä on, että mittareita muokataan vastaamaan paremmin havainnoitavaa muuttujaa. Havainnoijien kohdalla tulisi painottaa sitä, millä perusteella havainnoijat valitaan ja miten heitä koulutetaan. Lopuksi tuloksista puhuttaessa tulisi heidän mielestään keskittyä siihen, kuinka tuloksia hyödynnetään ja kuinka luodaan menetelmä, jossa saavutetaan haluttu reliabiliteetin taso. (Hill, Charalambous & Kraft, 2012, s. 56.)

Bell ym. nostavat yhteenvedossaan (2012) esiin kontekstin merkityksen osana opetuksen laadun mittaamisen validiteettia. Vaikka tässä tutkimuksessa ei ole kyse opetuksen

laadusta, on heidän näkemysensä mielestäni tärkeä: heidän tutkimuksessaan oppilaat olivat yksi opetuksen laatuun vaikuttava tekijä. Oppilaat siis vaikuttivat siihen, miten opettajaa arvioitiin, eli konteksti oli keskeinen osa opetuksen laadun arvioinnin kriteerejä. Havaittu yhteys oli kuitenkin vain yksi niistä konteksteista, jotka vaikuttavat opetustilanteeseen. (Bell ym., 2012, s. 85.) Näin ollen eri kontekstit ovat havainnoinnissa tärkeässä roolissa, sillä ne vaikuttavat menetelmän arviointiin, sen sisältöihin ja käyttäytymiseen. Nämä kontekstit ovat tärkeitä ottaa huomioon, kun pohditaan tiedon yleistettävyyttä ja menetelmän toistettavuutta laajemmassa mittakaavassa.

Kuten aikaisemmin sanoin, mielestäni tärkeintä on kysyä ”mitä” ja ”miksi” ennen kuin voidaan kysyä ”miten hyvin”. Jonkin asian mittaus tai laskeminen ei siis ole itseisarvo. Varhaiskasvatuksen näkökulmasta menetelmät, mittarit ja pisteet voivat kuulostaa kylmiltä, ja suoraan kopioituna monet varmasti ovatkin. Kyse ei kuitenkaan ole mittareista ja pisteistä itsessään, vaan niiden kautta saadun tiedon sisältämästä potentiaalista. Menetelmä, joka onnistuu tallentamaan abstraktin käsitteen luotettavasti, voi olla korvaamaton työkalu käsitteen ymmärtämisessä. Samaan aikaan on kuitenkin tärkeää tiedostaa menetelmän rakenne, mittauksen virhe, toteutuksen mittakaava ja näiden yhdistelmät. On eri asia, onko havainnoinnin kohteena yksi lapsi vai tuhat lasta, ja on eri asia, käytetäänkö havainnoinnin kautta saatua tietoa yksilöllisen tuen tarpeen kehittämiseen vai pyritäänkö havainnoinnin kautta luomaan yleistettävää kuvausta perusjoukosta.

Puhuttaessa havainnointimenetelmien käytöstä ja arvioinnista on mielestäni tärkeää pitää kiinni suomalaisen varhaiskasvatuksen arvoista ja tarkastella kriittisesti, millaisia arvoja näihin menetelmiin sisältyy. Ihmisen luotettava havainto toisesta ihmisestä on monimutkainen konsepti, ja havainnointiin vaikuttavat lukemattomat muuttujat. On tärkeää tiedostaa näiden muuttujien merkitys ja vaikutus, varsinkin jos havainnot yleistetään tai niiden pohjalta tehdään johtopäätöksiä tai tulkintoja laajemmassa mittakaavassa. Menetelmän käyttöön sisältyy ehto sen jatkuvasta arvioimisesta ja kehittämisestä. Menetelmän suunnittelu, toteutus, kehittäminen ja arviointi eivät siis muodosta janaa, jolla on alku ja loppu, vaan kehän, jossa ”mitä”, ”miksi” ja ”miten hyvin” ruokkivat toisiaan ja antavat syvemmän, luotettavamman ymmärryksen tutkittavasta ilmiöstä.

Lähteet

- Anttila, P. (2006). *Tutkiva toiminta ja ilmaisu, teos, tekeminen* (2. painos). Hamina: Akatiimi.
- Arvola, O., Lastikka, A. L., & Reunamo, J. (2017). Increasing immigrant children's participation in the Finnish early childhood education context. *The European Journal of Social & Behavioural Sciences*, 20(3), s. 2539–2548.
- Bakeman, R., & Gottman, J. M. (1986). *Observing interaction: An introduction to sequential analysis*. Cambridge: Cambridge University Press.
- Banerjee, M., Capozzoli, M., McSweeney, L., & Debajyoti, S. (1999). Beyond kappa: A review of interrater agreement measures. *Canadian Journal of Statistics*, 27(1), s. 3–23.
- Barrett, L. (2017). The theory of constructed emotion: An active inference account of interoception and categorization. *Social cognitive and affective neuroscience*, 12(1), s. 1–23.
- Bartko, J., & Carpenter, T. W. (1976). On methods and theory of reliability. *The Journal of Nervous and Mental Disease*, 163(5), s. 307–317.
- Bell, C. A., Gitomer, D. H., McCaffrey, D. F., Hamre, B. K., Pianta, R. C., & Qi, Y. (2012). An argument approach to observation protocol validity. *Educational Assessment*, 17(2–3), s. 62–87.
- Borsboom, D., Cramer, A. O. J., Kievit, R. A., Scholten, A. Z., & Frančić, S. (2009). The end of construct validity. Teoksessa Lissitz, R. W. *The concept of validity: Revisions, new directions, and applications*. Charlotte, NC: Information Age Pub.
- Brennan, P. F., & Hays, B. J. (1992). The kappa statistic for establishing interrater reliability in the secondary analysis of qualitative clinical data. *Research in nursing & health*, 15(2), s. 153–158.
- Brennan, R. L. (1992). Generalizability theory. *Educational Measurement: Issues and Practice*, 11(4), s. 27–34.
- Brennan, R. L. (2001). *Generalizability theory*. New York: Springer.
- Byman, R. (2018). *Määrälliset tutkimusmenetelmät 2*. Luentomateriaalit ja muistiinpanot.

- Byrt, T. (1996). How good is that agreement? *Epidemiology*, 7(5), 561.
- Byrt, T., Bishop, J., & Carlin, J. (1993). Bias, prevalence and kappa. *Journal of Clinical Epidemiology*, 46(5), s. 423–429.
- Caiman, C., & Lundegard, I. (2014). Pre-school children's agency in learning for sustainable development. *Environmental Education Research*, 20(4), s. 437–459.
- Carmines, E. G., & Zeller, R. A. (1979). *Quantitative applications in the social sciences: Reliability and validity assessment* Thousand Oaks, CA: SAGE Publications Ltd.
- Cicchetti, D. V., & Feinstein, A. R. (1990). High agreement but low kappa: II. Resolving the paradoxes. *Journal of clinical epidemiology*, 43(6), s. 551–558.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), s. 37–46.
- Cohen, J. (1968). Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological Bulletin*, 70(4), s. 213–220.
- Cohen, L., Manion, L., & Morrison, K. (2011). *Research methods in education* (7. painos). Abingdon: Routledge.
- Cohn, J. F., Ambadar, Z., & Ekman, P. (2007). Observer-based measurement of facial expression with the Facial Action Coding System. *The handbook of emotion elicitation and assessment*, s. 203–221.
- Cook, D. A., Brydges, R., Ginsburg, S., & Hatala, R. (2015). A contemporary approach to validity arguments: A practical guide to Kane's framework. *Medical Education*, 49(6), s. 560–575.
- Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, 78(1), s. 98–104.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), s. 297–334.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4), s. 281–302.

- Cronbach, L. J., & Shavelson, R.J. (2004). My current thoughts on coefficient alpha and successor procedures. *Educational and Psychological Measurement*, 64(3), s. 391–418.
- Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles*. New York: Wiley.
- Daniels, H. (2005). *An Introduction to Vygotsky*. London: Routledge.
- Davenport, E. C., Davidson, M. L., Liou, P-Y., & Love, Q. U. (2015). Reliability, dimensionality, and internal consistency as defined by Cronbach: Distinct albeit related concepts. *Educational Measurement: Issues and Practice*, 34(4), s. 4–9.
- DeMonbrun, R. M., Finelli, C. J., & Shekhar, P. (2015). Methods for establishing validity and reliability of observation protocols. *ASEE Annual Conference and Exposition, Conference Proceedings*. 122(122), s. 1–10.
- Ebbeck, M., Winter, P., Russo, S., Yim, H., Teo-Zutarte, G., & Goh, M. (2012). Measuring children's involvement as an indicator of curriculum effectiveness: A curriculum evaluation of a selected child study centre in Singapore. *Early Child Development and Care*, 182(5), s. 609–619.
- Ekman, P. (1989). The argument and evidence about universals in facial expressions of emotion. Teoksessa H. Wagner & A. Manstead (toim.), *Handbook of social psychophysiology*. Oxford, England: John Wiley & Sons, s. 143–164.
- Ekman, P. (1992a). An argument for basic emotions. *Cognition & Emotion*, 6(3–4), s. 169–200.
- Ekman, P. (1992b). Facial expressions of emotion: New findings, new questions. *Psychological Science*, 3(1), s. 34–38.
- Ekman, P. (1999). Basic emotions. Teoksessa Dalglish, T. & Power, M., *Handbook of cognition and emotion*. Chichester: John Wiley & Sons.
- Ekman, P. (2005). Conclusion: What we have learned by measuring facial behavior. Teoksessa Ekman, P. & Rosenberg, E. L. *What the face reveals: Basic and applied studies of spontaneous expression using the facial action coding system (FACS)* (2. painos). Oxford: Oxford University Press.

- Ekman, P., & Friesen, W. V. (1976). Measuring facial movement. *Environmental psychology and nonverbal behavior*, 1(1), s. 56–75.
- Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin*, 128(2), s. 203–235.
- Eräutuuli, M., Leino, J., & Yli-Luoma, P. V. J. (1994). *Kvantitatiiviset tutkimusmenetelmät ihmistieteissä*. Helsinki: Kirjayhtymä.
- Eskola, A. (1967). *Sosiologian tutkimusmenetelmät: 2*. Porvoo; Helsinki: WSOY.
- Faherty, V. E. (2008). *Compassionate statistics: Applied quantitative analysis for social services with Exercises and Instructions in SPSS*. Thousand Oaks: SAGE Publications.
- Feinstein, A. R., & Cicchetti, D. V. (1990). High agreement but low kappa: I. The problems of two paradoxes. *Journal of clinical epidemiology*, 43(6), s. 543–549.
- Fenech, M., Sweller, M., & Harrison, L. (2010). Identifying high-quality centre-based childcare using quantitative data-sets: What the numbers do and don't tell us. *International Journal of Early Years Education*, 18(4), s. 283–296.
- Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76(5), s. 378–382.
- Goldspink, C., & Foster, M. (2013). A conceptual model and set of instruments for measuring student engagement in learning. *Cambridge Journal of Education*, 43(3), s. 291–311.
- Gomez-Garibello, C., & Young, M. (2018). Emotions and assessment: Considerations for rater-based judgements of entrustment. *Medical Education*, 52(3), s. 254–262.
- Graham, M., Milanowski, A., & Miller, J. (2012). Measuring and promoting inter-rater agreement of teacher and principal performance ratings. *Verkkojulkaisu, CECR*, s. 1–33. Noudettu: <https://eric.ed.gov/?id=ED532068>
- Green, S., & Yang, Y. (2009). Commentary on coefficient alpha: A cautionary tale. *Psychometrika*, 74(1), s. 121–135.
- Gruber, H. E., & Vonèche, J. J. (1995). *The essential Piaget: [an interpretive reference and guide]*. London: Jason Aronson.

- Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology*, 8(1), s. 23–34.
- Hautamäki, A. & Laevers, F. (1997). *Toimintaan sitoutuneisuuden arviointiasteikko leikkikäisille lapsille: The Leuven Involvement Scale for Young Children, LIS-YC: käsikirja*. Helsinki; Vantaa: Helsingin yliopisto, opettajankoulutuslaitos: Vantaan täydennyskoulutuslaitos, koulutuksen arviointipalvelut.
- Hill, H. C., Charalambous, C. Y., & Kraft, M. A. (2012). When rater reliability is not enough. *Educational Researcher*, 41(2), s. 56–64.
- Hirsjärvi, S., Remes, P. & Sajavaara, P. (2007). *Tutki ja kirjoita* (13. osin uudistettu painos). Helsinki: Tammi.
- Hirsjärvi, S., Remes, P., & Sajavaara, P. (2009). *Tutki ja kirjoita* (15. uudistettu painos). Helsinki: Tammi.
- Ishimine, K., & Tayler, C. (2014). Assessing quality in early childhood education and care. *European Journal of Education*, 49(2), s. 272–290.
- Johnson, B., & Christensen, L. B. (2004). *Educational research: Quantitative, qualitative, and mixed approaches* (2. painos). Boston: Allyn and Bacon.
- Kalliala, M. (2011). Look at me! Does the adult truly see and respond to the child in Finnish day-care centres? *European Early Childhood Education Research Journal*, 19(2), s. 237–253.
- Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin*, 112(3), s. 527–535.
- Kane, M. T. (2006). Validation. Teoksessa: Brennan, R. L. (toim.), *Educational Measurement* (4. painos). ACE/Praeger Series on Higher Education.
- Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50(1), s. 1–73.
- Kane, M. T. (2016). Explicating validity. *Assessment in Education: Principles, Policy & Practice*, 23(2), s. 198–211.

- Kane, M. T. (2017). Loosening psychometric constraints on educational assessments. *Assessment in Education: Principles, Policy & Practice*, 24(3), s. 447–453.
- Karma, K. (1983). *Käyttätymistieteiden metodologian perusteet* (1. painos). Helsinki: Otava.
- Kitchener, R. F. (1986). *Piaget's theory of knowledge: Genetic epistemology & scientific reason*. New Haven: Yale University Press.
- Knors, H., & Renting, B. (2000). Measuring the quality of education: The involvement of bilingually educated deaf children. *American Annals of the Deaf*, 145(3), s. 268–274.
- Laevers, F. (1994). *Defining and assessing quality in early childhood education*. Leuven: Leuven University Press.
- Laevers, F. (2006). Making care and education more effective through wellbeing and involvement. An introduction to Experiential Education. Teoksessa *Working together: making a difference for South Australian children: our children the future early childhood seminar*. (Vol. 8). s. 1–8. Noudettu: https://vorming.cego.be/images/downloads/Ond_DP_IntroductionExpEduc.pdf
- Lakes, K. D., & Hoyt, W. T. (2009). Applications of generalizability theory to clinical child and adolescent psychology research. *Journal of Clinical Child & Adolescent Psychology*, 38(1), s. 144–165.
- Landis, J., & Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), s. 159–174.
- Locasale-Crouch, J., Konold, T., Pianta, R., Howes, C., Margaret, B., Bryant, D., ...Barbarin, O. (2007). Observed classroom quality profiles in state-funded pre-kindergarten programs and associations with teacher, program, and classroom characteristics. *Early Childhood Research Quarterly*, 22(1), s. 3–17.
- Mason, J. (2002). *Qualitative researching* (2. painos). London: SAGE Publications.
- Matsumoto, D. (1993). Ethnic differences in affect intensity, emotion judgments, display rule attitudes, and self-reported emotional expression in an American sample. *Motivation and emotion*, 17(2), s. 107–123.

- McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia medica*, 22(3), s. 276–282.
- Metsämuuronen, J. (2011). *Tutkimuksen tekemisen perusteet ihmistieteissä: Opiskelijalaitos* (E-kirjan 1. painos). Helsinki: International Methelp.
- Mezzich, J. E., Kraemer, H. C., Worthington, D. R. L., & Goffman, G. A. (1981). Assessment of agreement among several raters formulating multiple diagnoses. *Journal of psychiatric research*, 16(1), s. 29–39.
- Mislevy, R. J., Moss, P. A., & Gee, J. P. (2009). On qualitative and quantitative reasoning in validity. Teoksessa Bachman, L. F., Becker, B. J., Eisenhart, M., Ercikan, K., Gee, J. P., Luke, A., . . . Wu, M. *Generalizing from educational research: Beyond qualitative and quantitative polarization*. New York: Routledge.
- Mushquash, C., & O'Connor, B. (2006). SPSS and SAS programs for generalizability theory analyses. *Behavior Research Methods*, 38(3), s. 542–547.
- Nakamura, J., & Csikszentmihalyi, M. (2014) The concept of flow. Teoksessa Csikszentmihalyi, M. *Flow and the foundations of positive psychology: The collected works of Mihaly Csikszentmihalyi*. Berlin: Springer.
- Noland, M., Danner, F., Dewalt, K., McFadden, M., & Kotchen, M. (1990). The measurement of physical activity in young children. *Research Quarterly for Exercise and Sport*, 61(2), s. 146–153.
- Nummenmaa, L. (2004). *Käyttäytymistieteiden tilastolliset menetelmät*. Helsinki: Tammi.
- Ortony, A., & Turner, T. J. (1990). What's basic about basic emotions? *Psychological Review*, 97(3), s. 315–331.
- Paiz, J. M., Angeli, E., Wagner, J., Lawrick, E., Moore, K., Anderson, M., . . . Keck, R. (16.1.2018). *APA Changes 6th Edition*. Noudettu: <https://owl.english.purdue.edu/owl/resource/560/24/>
- Pascal, C., Bertam, T., Mould, C., & Hall, R. (1998). Exploring the relationship between process and outcome in young children's learning: Stage one of a longitudinal study. *International Journal of Educational Research*, 29(1), s. 51–67.

- Pellegrini, A. D., Symons, F. J., & Hoch, J. (2004). *Observing children in their natural worlds: A methodological primer* (2. painos). Mahwah, N.J.: Lawrence Erlbaum Associates.
- Piaget, J. & Inhelder, B. (1977). *Lapsen psykologia*. Jyväskylä: Gummerus.
- Piaget, J. (1976). Piaget's theory. Teoksessa Chipman, H. H., Inhelder, B. & Zwingmann, C. *Piaget and his school: A reader in developmental psychology*. New York, NY: Springer-Verlag.
- Puhl, J., Greaves, K., Hoyt, M., & Baranowski, T. (1990). Children's Activity Rating Scale (CARS): description and calibration. *Research Quarterly for Exercise and Sport*, 61(1), s. 26–36.
- Rentzou, K. (2017). Using rating scales to evaluate quality early childhood education and care: reliability issues. *European Early Childhood Education Research Journal*, 25(5), s. 667–681.
- Reunamo, J. (1998). *Olemassaoleva ja muotoutuva vuorovaikutuksen hahmottaminen: Toimintatutkimus Helsingin lastentarhanopettajaopistossa*. Helsinki: Helsingin yliopiston opettajankoulutuslaitos.
- Reunamo, J. (2000) *Lapsi muuttavana ja muuttavana osallistujana päiväkotitilanteissa*. Noudettu: <http://www.helsinki.fi/~reunamo/tutkimus/tutk98/kansi.htm>
- Reunamo, J. (2007). *Tasapainoinen varhaiskasvatus: Erilaisia tapoja suhtautua muutokseen*. Helsinki: WSOY oppimateriaalit.
- Reunamo, J. (2016). Lapsilähtöinen toimintakulttuuri edellyttää lapsen kuuntelua. Teoksessa A. Sääkslahti (toim.) *Tieteelliset perusteet varhaisvuosien fyysisen aktiivisuuden suosituksille*. Opetus- ja kulttuuriministeriön julkaisuja 2016:22, s. 27–31.
- Reunamo, J. (2017). *Kehittävä palaute varhaiskasvatuksessa: Observoijan koulutus* [Powerpoint]. Noudettu: http://blogs.helsinki.fi/ececfeedback/files/2017/03/Obs.koulutus1_2017.pdf
- Reunamo, J. (2018). *Kehittävä palaute varhaiskasvatuksessa: tietoja menetelmästä*. Noudettu: <http://blogs.helsinki.fi/ececfeedback/tervetuloa/>

- Reunamo, J., Lee, H-C., Wang, L-C., Ruokonen, I., Nikkola, T., & Malmstrom, S. (2014). Children's creativity in day care. *Early Child Development and Care*, 184(4), s. 617–632.
- Righart, R., & de Gelder, B. (2008). Recognition of facial expressions is influenced by emotional scene gist. *Cognitive, Affective, & Behavioral Neuroscience*, 8(3), s. 264–272.
- Russell, J. A. (1994). Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychological Bulletin*, 115(1), s. 102–141.
- Semmelroth, C. L., & Johnson, E. (2014). Measuring rater reliability on a special education observation tool. *Assessment for Effective Intervention*, 39(3), s. 131–145.
- Sheridan, S. (2000). A Comparison of external and self-evaluations of quality in early childhood education. *Early Child Development and Care*, 164(1), s. 63–78.
- Sheridan, S. (2007). Dimensions of pedagogical quality in preschool. *International Journal of Early Years Education*, 15(2), s. 197–217.
- Shweta, R.C., Bajpai, R. C., & Chaturvedi, H. K. (2015). Evaluation of inter-rater agreement and inter-rater reliability for observational data: An overview of concepts and methods. *Journal of the Indian Academy of Applied Psychology*, 41(3), s. 20–27.
- Silverman, D. (2001). *Interpreting qualitative data: Methods for analysing talk, text and interaction* (2. painos). London: SAGE.
- Sim, J., & Wright, C. C. (2005). The kappa statistic in reliability studies: Use, interpretation and sample size requirements. *Physical Therapy*, 85(3), s. 257–268.
- Sireci, S. G. (2009). Packing and unpacking sources of validity evidence: History repeats itself again. Teoksessa Lissitz, R. W. *The concept of validity: Revisions, new directions, and applications*. Charlotte, NC: Information Age Pub.
- Siren, T. (2008). *Yksimielisyyden mittaamisen tilastolliset menetelmät*. Pro gradu -tutkielma. Tilastotiede. Tampere: Tampereen yliopisto.
Noudettu: <http://urn.fi/urn:nbn:fi:uta-1-19384>

- Sivonen, J. (28.1.2004). Korrelaatio ja riippuvuusluvut. *KvantiMOTV - Menetelmäopetuksen tietovaranto* [verkkójulkaisu]. Tampere: Yhteiskuntatieteellinen tietoarkisto. Noudettu: <http://www.fsd.uta.fi/menetelmaopetus/korrelaatio/korrelaatio.html>
- Smith, G. T. (2005). On construct validity: Issues of method and measurement. *Psychological Assessment*, 17(4), s. 396–408.
- Stemler, S. E. (2004). A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. *Practical Assessment*, 9(4), s. 1–11.
- Tang, X., Pakarinen, E., Lerkkanen, M-K., Kikas, E., Muotka, J., & Nurmi, J-E. (2017). Validating the early childhood classroom observation measure in first and third grade classrooms. *Scandinavian Journal of Educational Research*, 61(3), s. 275–294.
- Uusitalo, H. (1991). *Tiede, tutkimus ja tutkielma: Johdatus tutkielman maailmaan*. Porvoo: WSOY.
- Vach, W. (2005). The dependence of Cohen's kappa on the prevalence does not matter. *Journal of clinical epidemiology*, 58(7), s. 655–661.
- Valli, R. (2015). Parivertailu aineistonkeruussa. Teoksessa *Ikkunoita tutkimusmetodeihin: 1, Metodien valinta ja aineistonkeruu: virikkeitä aloittelevalle tutkijalle* (4. uudistettu ja täydennetty painos). Jyväskylä: PS-kustannus.
- Webb, N. M., & Shavelson, R. J. (2005). Generalizability theory: overview. Teoksessa Everitt, B. S. & Howell, D. C. *Encyclopedia of statistics in behavioral science*. Chichester: John Wiley.
- Viera, A., & Garrett, J. (2005). Understanding interobserver agreement: The kappa statistic. *Family Medicine*, 37(5), s. 360–363.
- Vilka, H. (2015). *Tutki ja kehitä* (4. uudistettu painos). Jyväskylä: PS-kustannus.

Liitteet

Liite 1. *Sähköinen observointilomake*, s. 1–4. Noudettu 16.8.2017 osoitteesta:
<https://blogs.helsinki.fi/ecefeedback/tiedostopankki/>

16.8.2017Observointilomake

Observoijan koodi *

Päiväkodin koodi *

Lapsen numero *

Valitse vaihtoehto...
▼

A. Yleinen tilanne: Lapsen toiminnan yleinen kehys (mitä lapsen pitää/kuuluu tehdä, sisältää siirtymät tilanteiden välissä) *

1. Suora kasvatustoiminta sisällä (enimmäkseen aikuisen aloittamaa, suunniteltua toimintaa: opetusta, ohjeistusta, ryhmätoimintaa, kirjan lukua, esityksiä, ryhmäkeskustelua)
2. Tuettu leikki sisällä (lapsilähtöistä, aikuinen havainnoi, osallistuu ja on valmis tarvittaessa tukemaan leikkiprosesseja)
3. Vapaa leikki sisällä (Itsenäisempi leikki yksin tai kavereiden kanssa. Kasvattaja ei osallistu prosessiin)
4. Ulkotoiminta (usein pihalla, puistossa, retkellä)
5. Perushoito (pukeminen, riisuminen, WC, hygienia, voi sisältää odotusta ja valmistautumista)
6. Ruokailu (aamupala, lounas, ruuan odotus, järjestelyt ennen/jälkeen syömisestä)
7. Lepohetki

Valitse vaihtoehto...
▼

https://my.surveypal.com/app/form?_d=0&_sid=552333673&_k=W0VNefm7JG2rJaQq0nXPRiq7Hplyq1Gi42ga5_wBH4d3BfZdV9fkYdeA2elsdeM... 1/4

16.8.2017

Observointilomake

B. Lapsen oma toiminta (mitä lapsi tekee)*

1. Roolileikki tai mielikuvaleikki (lelulla tai itsellä on rooli)
2. Sääntöleikki (esim. pallopelejä, lautapelejä, elektroninen peli, pysyvät säännöt, kilpailu)
3. Esineleikki ja kokeilu materiaaleilla ja leluilla (esim. hiekkalaatikolla, paperilla)
4. Fyysinen leikki tai aktiviteetti (juoksu, keinuminen, hyppiminen, riehuminen, tanssi, kiipeily, harjoitus)
5. Lukeminen, opettaja lukee tai lapsi lukee/selaa kirjaa itse
6. Tehtävä (esim. kotitehtävä, kynä-paperi –työt, pukemisen harjoittelu, lorun opettelu, taidon harjoitus, työ)
7. Toiminta ilman kiintopistettä tai kunnon kontaktia muihin (esim. kuljeskelu, etsiminen, odottelu)
8. Yhdessäolo muiden kanssa (esim. hengaillee/kävelee kavereiden kanssa, juttelee kiipeilytelineellä)
9. Kielletty toiminta (esim. ohjeiden rikkominen, kiusaaminen, häirintä)
10. Lapsi toimii yleisen kehyksen sisällä, mutta ei sovi yllä oleviin luokkiin B1-B9 (esim. lapsi syö ruokailussa, pukeutuu ulos lähettäessä)
11. Muu toiminta (ei sovi muihin luokkiin, esim. sekavuus, paljon vaihdoksia, ei rakennetta)

Valitse vaihtoehto...



C. Lapsen pääasiallinen huomion kohde*

1. Ei-sosiaalinen kohde (esim. lelu, hiekka, auto, palikat, vesi, oma itse)
2. Aikuinen (esim. seuraa aikuisen kertomusta, keskustelee, huomio voi sisältää esim. opetusmateriaalia)
3. Toinen lapsi (mukana voi olla myös leluja ja muita ei-sosiaalisia kohteita esim. lapsen kädessä)
4. Ryhmä lapsia (2 tai useampia lapsia, mukana voi olla myös leluja ym.)
5. Kokonaistilanne (tilanteessa on niin paljon elementtejä, ettei pääasiallista kohdetta voi nimetä, esim. lapsia, aikuisia, materiaaleja, usein dynaaminen tilanne).

Valitse vaihtoehto...



D. Lähin sosiaalinen lapsikontakti*

Jos ei kontaktia tutkittaviin lapsiin, valitse 6

Valitse vaihtoehto...



16.8.2017

Observointilomake

E. Lapsen fyysinen aktiivisuus (painotetaan viim. 10 s.) *

1. Liikkumaton (esim. nukkuminen, seisominen, istuu vaunuissa ja seisominen)
2. Paikallaan tapahtuva raajojen/kehon liike (esim. nouseminen ylös, kevyehkön esineen liikuttaminen)
3. Hitaat ja kevyet liikkeet (esim. käveleminen, kevyt keinuminen)
4. Kohtalaisesti kuormittavat liikkeet (esim. käveleminen ylämäkeen, vähintään 2 hyppelyä, kiipeily telineessä, käsillä roikkuminen ja vauhdinotto jaloilla)
5. Nopeat ja kuormittavat liikkeet (esim. juokseminen, portaiden nousu, vähintään kolme hyppelyä, kiipeilytelineessä roikkuminen jalat liikkeessä).

Valitse vaihtoehto... ▼

F. Lapsen sitoutuneisuus *

1. Yksinkertainen, kaavamainen, toistuva, passiivinen toiminta, ei energiaa, ei älyllistä haastetta
2. Usein keskeytyvä toiminta ja toimintaan kiinnittyminen
3. Enimmäkseen jatkuva toiminta, huomio hajoaa helposti, vähän energiaa
4. Jatkuva toiminta jossa intensiivisiä hetkiä, tarkkaavaisuus ei helposti hajoa
5. Kestävä intensiivinen toiminta, keskittyminen, luovuus, energia ja sitkeys

Valitse vaihtoehto... ▼

G. Emootio *

1. Viha
2. Ärtymys, turhauma
3. Pelko, ahdistus
4. Suru, masennus
5. Ilo, riemu
6. Onnellisuus, tyytyväisyys
7. Hämmästys, valppaus, uteliaisuus, innostus
8. Neutraali, tynni, rauhallinen
9. Muu tunne

Valitse vaihtoehto... ▼

16.8.2017

Observointilomake

H. Sosiaalinen orientaatio*

1. Mukautuva ja avoin, hyväksyy ja huomioi.
2. Osallistuva, vuorovaikutteinen, avoin, tekee yhteistyötä.
3. Dominoiva, itsepintainen, ehkä itsekeskeinen tai määräilevä.
4. Vetäytyy sosiaalisesta tilanteesta, ei-sosiaalinen, ei-vuorovaikutteinen.
5. Ei voida määrittää.

Valitse vaihtoehto... ▼

I. Lähimmän aikuisen toiminta*

1. Vastaanottaa: havainnoi, kuuntelee, on avoin, herkistyy, hyväksyy, huomioi (syventyy lapsen kokemukseen).
2. Osallistuu: vuorovaikutteinen, avoin, tekee yhteistyötä, rikastaa, kokeilee, (syventää lapsen kokemusta).
3. Tavoittelee: opettaa, motivoi, ohjaa, harjoittaa (syventää ennalta määrätyn tavoitteen suuntaan).
4. Pinnallinen: Ei vuorovaikutteinen, valvoo, päättää, rutiini, ei huomioi (ei syvennä kokemusta).
5. Negatiivinen: Arvostelee, mitätöi, syyttää, ivailee, uhkailee (jyrää lapsen kokemuksen)
6. Ei aikuisen huomiota.

Valitse vaihtoehto... ▼

Lähetä vastaukset

Powered by SurveyPal 