

Aikuiskasvatuksen tutkimusmenetelmiä



Aikuiskasvatuksen asiantuntijat esittelevät aikuiskasvatustieteen tutkimusmenetelmiä ja näkökulmia tutkimukseen.

VILMA HÄNNINEN

Kuinka paljon on tarpeeksi? Aineiston määrä laadullisessa tutkimuksessa



MÄÄRÄLLISESSÄ TUTKIMUKSESSA pyritään tuottamaan tilastollisesti yleistettävää tietoa. Sen piirissä on käytettävissä tilastotieteestä johdettuja laskelmia siitä, paljonko havaintoyksiköitä tarvitaan päätelmien tekemiseen erilaisissa tutkimusasetelmissä. Nyrkissäntö on, että mitä suurempi on havaintojen määrä (N), sen parempi. Keskeistä on lisäksi, että otos on poimittu tilastollisesti hallitulla tavalla.

Laadullisessa tutkimuksessa kaikki on (melko lailla) toisin. Jos yleistyksiä pyritään tekemään, ne

eivät perustu määriin. Usein toistettu ajatus on, että laadullisessa tutkimuksessa yksikin havaintoyksikkö voi riittää. Kaiken kaikkiaan niin näytteen poimimiseen (suomen kielessä laadullisen tutkimuksen vastine otannalle; englannin kielessä käytetään termiä *sampling*) on kiinnitetty suhteellisen vähän huomiota sekä tutkimuksissa että menetelmäkirjoissa. Eriytyisen vähän on käsitelty kysymystä havaintoyksiköiden määrästä. Riittääkö aina yksi? Jos ei, niin kuinka paljon tarvitaan?

Riittävyttä on usein määritelty saturaatio-käsitteen avulla, erityisesti *grounded theory*-lähestymistavan mukaisessa tutkimuksessa. Aineisto on saturoitunut (kyllääntynyt) eli riittävä silloin, kun uuden tapauksen mukaan ottaminen ei tuota uutta tietoa. Kuulostaa yksinkertaiselta, mutta käytännössä arviointi on hankalaa. Miten arvioida etukäteen saturaatioon tarvittavien haastattelujen määrä? Miten määritellä ja osoittaa, että saturaatio on saavutettu? Ainakaan tutkijan kyllästymispistettä ei saisi sekoittaa aineiston kyllääntymispisteeseen. Analyysin tekoon kiirehtivällä tutkijalla voi olla kiusaus lopettaa aineiston keruu jo tulosten toistumisen ensimerkkeihin.

Laadullisen tutkimuksen tekijäkin joutuu siis pohtimaan aineiston määrään liittyviä kysymyksiä, ensin tutkimusta suunnitellessaan, sitten toteutunutta aineistoa arvioidessaan. Ratkaisut pohjautuvat usein perinnetietoon eli esimerkiksi perusteluun, jonka mukaan ”tämän verran haastateltavia on aiemmissakin väitöskirjoissa pidetty riittävänä”.

Menetelmäkeskustelussa ja tutkimuksen arvioinnissa on yhä useammin kuitenkin alettu kiinnittää huomiota siihen, että havaintoyksiköiden poiminnan pitäisi olla laadullisessa tutkimuksessa hallittua ja määrän perusteltu. Tässä kirjoituksessa pohdin näitä hallinnan ja perustelun kysymyksiä. Tavoitteenani ei ole niinkään vastata kysymykseen paljonko on riittävästi kuin eritellä sitä, millä keinoilla riittävyttä voi turvata ja arvioida.

Tarkastelen määräkysymystä lähinnä ajatellen sellaisia (laadullisessa tutkimuksessa melko tavanomaisia) tutkimuksia, jotka ovat ikään kuin kyselytutkimuksen laadullisia vastineita. Tämänkaltaisissa tutkimuksissa tarkastellaan tietyn ihmisjoukon kokemuksia tai näkemyksiä jostakin heitä yhdistävästä ilmiöstä (sairaus, työttömyys, yksinäisyys jne). Rajaan tarkastelun (niin ikään tyypillisiin) tilanteisiin, joissa havaintoyksiköt ovat toisistaan riippumattomia tai korkeintaan ketjumaisesti toisiinsa liittyviä, kuten lumipallo-otannassa on laita.

Vaikka laadullisen tutkimuksen menetelmäkeskustelussa usein korostetaan sitä, että määrällisen tutkimuksen arvioinnin käsitteet eivät siihen sovi, sovellan ajatuskokeen omaisesti seuraavassa määrällisestä tutkimuksesta tuttuja käsitteitä ulkoinen

ja sisäinen validiteetti. Maxwellin (2012) tapaan en tarkoita validiteetin käsitteellä aineiston vastaavuutta ”objektiivisen totuuden” kanssa, vaan sitä, että aineisto antaa mahdollisuuden tuottaa päteviä päätelmiä aineiston ulkopuolisesta todellisuudesta tutkimuskysymysten näkökulmasta.

ULKOINEN VALIDITEETTI

Aineiston ulkoisella validiteetilla tarkoitetaan osallistujajoukon suhdetta siihen joukkoon, josta tutkimuksen perusteella halutaan sanoa jotakin. Esimerkkinä voisi olla ”ikäihmiset”. Tämän sisällä pienempi joukko on se, jonka piiristä rekrytointi tehdään (vaikkapa ikäihmisille suunnatun lehden lukijakunta), ja pienimpänä lopullinen tutkimukseen osallistuvien joukko. Vaikka laadullisessa tutkimuksessa ei tilastollista yleistämistä tehdäkään, on syytä aina miettiä, mitkä valikoivat tekijät ovat voineet muovata osallistujajoukosta erityisen verrattuna kohdejoukkoon.

Ulkoisen validiteetin arvioinnissa on otettava huomioon tutkimuskysymykset eli se, millaisia päätelmiä halutaan tehdä. Jos halutaan väittää, että jonkinlainen tapaustyyppi on teoreettisessa kohdejoukossa tyypillinen, olisi aineiston annettava mahdollisimman hyvät eväät yleistysten tekemiseen. Tällöin survey-asetelmasta tuttu satunnaisotanta ja suhteellisen suuri osallistujien määrä antavat parhaan tuen päätelmien tekemiselle.

Jos tutkimuksella halutaan kartoittaa tiettyyn ilmiöön liittyvien kokemusten erilaisia tyyppisiä, satunnaisotannan ohella teoreettinen vaihtelun maksimointi voi tuottaa hyvän tuloksen. Jos halutaan tuoda esiin johonkin ilmiöön liittyviä ääripäitä tai erityistapauksia (esimerkiksi vailla yhtään kontaktia elävät vanhuksat, erityisen aktiiviset ikäihmiset), on syytä käyttää hyväksi ennakkotietoa.

Erilaisia näytteen poiminnan tapoja on usein listattu laadullisessa menetelmäkirjallisuudessa – esimerkiksi Onwuegbuzie ja Leech (2007) esittelevät 24 erilaista poimintamenetelmää. Olipa poimintatapa mikä tahansa, se on syytä pitää sekä mielessä että esillä raportissa ja arvioida sen vaikutuksia tuloksiin. Samaan tapaan kuin survey-tutkimuksessa arvioidaan lopullisen otoksen vinoutuneisuutta, on laadullisessa

tutkimuksessa tärkeää miettiä, missä suhteessa tutkittavien joukko saattaa systemaattisesti poiketa muista teoreettiseen kohdejoukkoon kuuluvista. Erityistä arviointia vaatii se, jos osallistujat ovat itse valikoituneet mukaan haastattelu- tai kirjoittajakutsun perusteella. Lähtökohtaisesti he ovat todennäköisesti keskivertoa aktiivisempia, sanavalmiimpia ja sellaisia, jotka kokevat tutkimuksen aiheen itselleen tärkeäksi.

Havaintoyksiköiden määrällä on laadullisessakin tutkimuksessa merkitystä tutkimuksen validiteetin kannalta – mutta ei yksinkertaisella ”enemmän on enemmän” -logiikalla. Koska laadullisen tutkimuksen perinne on hyvin moninainen, ovat myös käsitykset sopivasta havaintoyksiköiden määrästä vaihtelevia. Mark Mason (2010) on eritellyt laadullisessa menetelmäkirjallisuudessa esiteltyjä suuntaviivoja sopiville näytteen suuruuksille: esimerkiksi *grounded theory* -tutkimuksissa haastateltavien määräksi suositellaan yleensä 20–50 ja fenomenologisissa tutkimuksissa 5–25.

Mason toteaa, että perusteluja näille suosituksille ei juuri esitetä, eikä käy ilmi miksi eri perinteissä suositukset ovat erilaisia. Tutkiessaan englantilaisten laadullisten väitöskirjojen näytteiden kokoja Mason havaitsi, että ne vaihtelivat yhden ja 95 välillä, ja tyypillinen koko on noin 30 osallistujaa. Kiinnostavaa on myös, että näytteiden koot painottuivat tasakymmeniin, mikä kertoo siitä, että ne on päätetty etukäteen. Mason toteaa, että nämä tyypillisyydet kertovat näytteen koon määrääntymisestä pikemmin akateemisten normien kuin tutkimusten omien sisäisten tarpeiden perusteella. Toisaalta voisi ehkä olla niinkin, että kolmekymmentä on kokemusten saatossa osoittautunut usein toimivaksi – riittävän kirjavaksi ja samalla hallittavaksi – näytteen kooksi.

Myös osallistujien määrän arvioinnissa on mahdollista hyödyntää jaottelua erilaisiin tavoiteltuihin päätelmätyyppeihin. Tuntumanomaisesti voisi sanoa, että jos halutaan tehdä ”kevyt” yleistyksen laajempaan perusjoukkoon (esim. ”ikäihmiset”), olisi havaintoyksiköitä oltava noin kaksi–kolmekymmentä ja näytteen poimintatavan satunnainen. Samoin silloin, jos halutaan saada esiin erilaisten keskeisten tyyppien kirjoja.

Mitä heterogeenisempi on tutkimuksen kohdejoukko, sitä laajempi näyte tarvitaan. Jos halutaan verrata ennalta määriteltyjä ryhmiä (vaikkapa naiset

-miehet), tulisi kussakin ryhmässä olla ainakin noin 10–15 henkeä. Jos halutaan vain osoittaa, että kokemukset voivat olla vaihtelevia, riittää muutama tapaus. Tällöin yksikin hyvin valittu esimerkki tuo esiin yhden mahdollisen – vaikkapa jossain mielessä äärimmäisen – tavan kokea, merkityksellistä tai kielellistä kiinnostuksen kohteena oleva ilmiö.

Johtamani *Masennuksen kanssa eläminen sosiaalisessa kontekstissa* -tutkimushankkeen osallistujajoukkoa voi eritellä seuraavaan tapaan. Väestörekisteristä poimituille Helsingissä ja Etelä-Karjalassa asuville naisille (N= 750) ja miehille (N= 1500) lähetettiin kutsu ilmoittautua haastatteluun, jos he ovat olleet masentuneita edellisten kahden vuoden aikana. Tuloksena saimme 31 nais- ja 24 miespuolista haastateltavaa.

Poimintatavan ansiosta osallistujat edustivat laajaa ammattien ja elämäntilanteiden kirjoja, ja toteutuneet ryhmät olivat riittävän suuria sisältääkseen erilaisia masennustarinoita. Naisten ja miesten haastatellut muodostivat kumpikin riittävän suuren aineiston erillistä analyysiä ajatellen (Turunen & Hänninen 2014, Valkonen & Hänninen 2014) ja aineisto sallii myös sukupuolten ja asuinpaikkojen vertailun.

Voimme arvioida aineistojen heijastavan melko hyvin nelikymppisten suomalaisten masennuskokemusten kirjoja. Kuitenkin on syytä muistaa, että todennäköisesti vain noin kymmenesosa ”haaviimme” joutuneista masennuksen kokeneista ilmoittautui haastatteluun. Tämä on voinut johtaa siihen, että kulttuurisesti vahvimmat masennustarinat ovat erityisen vahvasti edustettuina. Siksi emme uskaltaisi väittää, että naisten keskuudessa lähes joka toisen kertoma ”voimaantumistarina” olisi kaikkien masennuksen kokeneiden keskuudessa tavallisin tarina.

Laadullisessa tutkimuksessa on tavallista, että näytteen poiminta tapahtuu ”ota kiinni mistä saat” -menetelmällä (*convenience sampling*), eli tarvittava osallistujajoukko saadaan kokoon käyttämällä useita rekrytointikanavia. Lopputulos ei ole yhtä elegantti ja vakuuttava kuin vaikkapa satunnaismenetelmällä poimittu näyte, mutta paremman puutteessa tämäkin on aivan kunniallinen poiminnan tapa (tätä mieltä on myös Maxwell 2012, 95).

Voi käydä myös niin, ettei haluttua määrää osallistujia ponnisteluista huolimatta saada rekrytoitua

kohtuullisessa ajassa. Tärkeän asian tutkimista ei kuitenkaan ole osallistujien vähyiden vuoksi syytä keskeyttää. Rekrytoinnin vaikeudet voivat kertoa juuri siitä, että aihe on jollain tavalla kipeä tai salaisuuksien täyttämä, jolloin sen tutkiminen on erityisen tärkeää (esimerkiksi Hanna Pohjola, 2012, sai uransa loukkaantumiseen päättäneitä tanssijoita koskevaan väitöskirjatutkimukseensa vain kolme haastateltavaa). Jos osallistujien määrä jää vähäiseksi, tämän voi kääntää voitoksi panostamalla erityisesti aineiston sisäiseen laatuun ja analyysin perusteellisuuteen.

SISÄINEN VALIDITEETTI

Aineiston sisäisen validiteetin käsitteellä voi viitata siihen, kuinka hyvin aineisto sisällöltään ilmentää tarkastelun kohteena olevaa ilmiötä valitusta näkökulmasta (vrt. kvantitatiivisen tutkimuksen ”mittari mittaa sitä mitä on tarkoitus”).

Laadullisen tutkimuksen sisäisen validiteetin kannalta aineiston määrä osallistujaa kohti on tärkeä seikka. Onwuegbuzie ja Leech (2007) puhuvat tässäkin yhteydessä otannasta (*sampling*): miten hyvän ”otoksen” haastateltavan sanat muodostavat hänen potentiaalisesta aiheesta koskevasta puheestaan? Jos haastattelu on kovin lyhyt, tämä otos on sattumanvarainen ja niukka. Tuntumanomainen arvio on, että paljon alle puolentoista tunnin mittainen haastattelu voi jäädä liian suppeaksi.

Kuitenkin puheen laatu on vielä tärkeämpää kuin määrä. Jos haastattelussa ei saavuteta luottamusta, haastattelun ”sanojen otos” on vinoutunut. Haastattelun jatkaminen toisella kerralla voi paitsi lisätä kerrottujen asioiden määrää myös korjata vinoumia. Esimerkiksi Matti Kortteinen kirjoitti Lähiö-tutkimuksensa haastattelujen yhteydessä ”onnellisuusmuurista”, joka mureni toisella haastattelukerralla (Kortteinen 1982). Haastattelujen laatu on yhteydessä myös tarvittavaan määrään: kuten Mason (2010) toteaa, kokenut haastattelija saa pieneltä ihmisjoukolta rikkaamman haastatteluaineiston kuin aloittelija suurelta.

Pitkittäisasetelma, eli haastattelujen toistaminen pitemmän ajan kuluttua, vahvistaa laadullista aineistoa huomattavasti. Laadullisia seurantatutkimuksia on

jonkin verran tehty jo kauan (esim. Ruisniemi 2006), mutta erityisesti viime vuosina niiden suosio on lisääntynyt (esim. Nikander 2013). Pitkittäisasetelma tuottaa paljon aineistoa osallistujaa kohti, ja on luonnollista, että kovin suurta joukkoa ei ole mahdollista kuljettaa läpi aineistonkeruu- ja analyysivaiheiden.

Laadullista aineistoa on mahdollista täydentää myös ”tringulaatiotyypisesti” keräämällä primaariaineiston lisäksi samoilta henkilöiltä tietoja jollakin muulla menetelmällä, kuten haastatteluja täydentävillä elämäkertakirjoituksilla ja omaisten haastatteluilla (Koski-Jännes & al. 1998) osallistujilta kootulla mittariaineistolla (esim. Valkonen 2007) tai osallistujajoukkoa laajemmalla joukolta kerätyllä kyselyaineistolla (esim. Kotovirta 2009).

ANALYYSIN LAATU

Aineiston analyysin laatua voi pitää aineiston ulkoisen ja sisäisen validiteetin ohella kolmantena tutkimuksen validiteetin ulottuvuutena: kuinka hyvin tutkija kykenee hyödyntämään aineistoa tutkimuskysymyksiinsä vastaamisessa? Määrän kielelle muutettuna: paljonko analyysiä on tarpeeksi? Tässäkin riippuu tutkimuksen kysymyksenasettelusta, millainen analyysi on tarkoituksenmukainen.

Kun halutaan tuottaa kiteytetty ja teoreettisesti käsitteellistetty kokonaiskuvaus laajasta aineistosta, on pätevyyskriteerinä se, kuinka hyvin tämä kiteytys kykenee tiivistämään aineiston pääsisällön, kuinka systemaattisesti ja kattavasti aineistoa on hyödynnetty sekä kuinka herkästi myös aineiston poikkeamat ja säröt on otettu kokonaiskuvan rakentamisessa huomioon.

Kokemuksen sisältöön tai sen kielelliseen rakentamiseen kohdistuvat tutkimukset puolestaan pureutuvat aineiston yksityiskohtiin tehden niistä tulintoja ja päätelmiä. Rivien väleihin pureutuvalla, lähilukevalla analyysillä voi pienestäkin aineistosta saada puristettua paljon irti.

Esimerkkinä tästä voisi olla Eero Suonisen diskursiivinen analyysi yhden perheenäidin haastattelun sisältämistä diskursseista ja niiden käytöstä (Suoninen 1998). Vaikka Suonisen aineistona on vain yksi haastattelu, analyysi valottaa yleispätevästi sitä, miten

ihmiset voivat puheessaan hyödyntää erilaisia kulttuurisia puhetapoja.

Myös useiden analyysitapojen yhdistäminen samassa tutkimuksessa auttaa ottamaan aineistosta mahdollisimman paljon irti. Tästä esimerkkinä voisi olla Tarja Aaltosen (2002) väitöskirjatutkimus, jossa afaatikkojen haastatteluja analysoitiin narratiivisin ja keskusteluanalyttisin menetelmin. Analyysin merkitys huomioon ottaen aineistoa voi olla myös liikaa. Näin on silloin, kun sen riittävä analysointi ei tutkimuksen resurssien puitteissa ole mahdollista.

Määrällisessä tutkimuksessa havaintoyksiköiden määrä määrittää keskeisesti sitä, kuinka luotettavia tilastollisia yleistyksiä aineistosta voidaan tehdä. Laadullisessa tutkimuksessa ei tehdä tilastollisia yleistyksiä, mutta tarkoituksena on kuitenkin sanoa jotakin aineiston ulkopuolisesta todellisuudesta – puhutaan esimerkiksi päätelmien (*inferences*) tekemisestä tai teoreettisesta yleistämisestä. Tässä aineiston määrällä

ei ole yksiselitteistä merkitystä. Päätelmien tekemisen perusteet pitää arvioida joka tutkimuksessa erikseen.

Tutkimuksen empiiristä laatua voi siis arvioida kolmen ulottuvuuden (osallistujien määrä, aineiston runsaus osallistujaa kohti sekä analyysin syvyys) yhteistuloksena. Mikä tahansa osallistujien määrä voi olla tarpeeksi, kun vain vastaavasti kasvatetaan muita ulottuvuuksia ja otetaan aineiston laatu huomioon päätelmiä tehtäessä. Lisäksi yksittäistä laadullista tutkimusta voi vankentaa suhteuttamalla sitä muihin niin laadullisiin kuin määrällisiinkin aiheita koskeviin aiempiin tutkimuksiin.

Vilma Hänninen
YTT
professori, sosiaalipsykologia
yhteiskuntatieteiden laitos
Itä-Suomen yliopisto

LÄHTEET

- Aaltonen, Tarja (2002). *Sanatonko tarinaton? Afaatikon intersubjektiiivinen maailma ja kertoen rakentuva identiteetti*. Miina Sillanpään säätöön tutkimuksia A:5.
- Kortteinen, Matti (1982). *Lähiö. Tutkimus elämäntapojen muutoksesta*. Helsinki: WSOY.
- Kotovirta, Elina (2009). *Huumeriippuvuudesta toipuminen Nimettömien Narkomaanien toveriseurassa*. Terveyden ja hyvinvoinnin laitoksen Tutkimuksia -sarja.
- Koski-Jännes, Anja, Jussila, Antti & Hänninen, Vilma (1998). *Miten riippuvuus voitetaan*. Helsinki: Otava.
- Mason, Mark (2010). Sample size and saturation in PhD studies using qualitative interviews. *Forum Qualitative Social Research* 11(3), Art. 8.
- Maxwell, Joseph A. (2012). *A realist approach for qualitative research*. Sage, Thousand Oaks.
- Nikander, Pirjo (2013). Sukupuoli, muutos ja ikääntymistutkimuksen metodologiset haasteet. Teoksessa H. Ojala & I. Pietilä: *Miehistä puhetta. Miehet, ikääntyminen ja vanhenemisen kulttuuriset mallit*. Tampere University Press, Tampere.
- Onwuegbuzie, Anthony J. & Leech, Nancy L. (2007). A call for qualitative power analysis. *Quality & Quantity* 41, 105–121.
- Pohjola, Hanna (2012). *Toinen iho. Uransa loukkaantumiseen päättäneen nykytanssijan identiteetti*. Acta Scenica 29, Helsinki.
- Ruisniemi, Arja (2006). *Minäkuvan muutos päihderiippuvuudesta toipumisessa. Tutkimus yhteisöllisestä päihdekuntoutuksesta*. Acta Electronica Universitatis Tamperensis 525.
- Suoninen, Eero (1998). *Miten tutkia moniäänistä ihmistä? Diskurssianalyttisen tutkimusotteen kehittälyä*. Acta Universitatis Tamperensis, Tampere.
- Turunen, Anne-Maarit & Hänninen, Vilma (2014). Keski-ikäisten naisten masennuspolkuja. *Psykologia* 49(2), 212–136.
- Valkonen, Jukka & Hänninen, Vilma (2015). Narratives of masculinity and depression. *Men & Masculinities* 16(2), 160–180.
- Valkonen, Jukka (2007). *Psykoterapia, masennus ja sisäinen tarina*. Kuntoutussäätöön tutkimuksia 77, Helsinki.