

Rasch analyysin käyttömahdollisuudet hoitotieteessä – esimerkkinä jalkojen omahoidon tietotesti

MINNA STOLT

TtT, dosentti, jalkaterapeutti (AMK),
yliopistonlehtori, professori

Turun yliopisto
Hoitotieteen laitos
Itä-Suomen yliopisto
Hoitotieteen laitos

MIKO PASANEN

FM, statistikko

Turun yliopisto
Hoitotieteen laitos

RIITTA SUHONEN

TtT, sairaanhoitaja, professori,
sivutoiminen ylihoitaja

Turun yliopisto
Hoitotieteen laitos
Turun yliopistollinen keskussairaala ja Turun
kaupunki, hyvinvoinnin palvelukokonaisuus

TIIVISTELMÄ

Rasch analyysi, osiovaste-teorian (Item Response Theory, IRT) menettely, on mittarin luotettavuuden arvioinnissa käytetty tilastollinen menetelmä. Sillä tarkastellaan osioiden (kysymysten/väittämiä) vaikeustasoa ja vastaajia mitattavana olevaa kykyä tai ominaisuutta. Hoitotieteessä Rasch analyysin käyttö on vähäistä. Tutkimuksessa kuvataan Rasch analyysiä menetelmänä, sen käyttömahdollisuuksia ja sen tarjoamaa informaatiota mittarin psykometristen ominaisuuksien arvioinnissa hoitotieteen alalla. Käytämme esimerkkinä poikkileikkaustutkimusta, jossa aineisto kerättiin nivelrikkoa sairastavilta henkilöiltä (n=129) jalkojen omahoitotiedot -tietotestillä (20 väittämää, 3-portainen vastausasteikko: oikein, väärin, en tiedä). Rasch analyysi kohdistui mittarin yksiuotteisuuteen, osiokohtaisiin kiinnittymisarvoihin, erotteluindeksiin ja erilaiseen osiotoimintaan. Jalkojen omahoitotiedot -tietotesti oli vaikea suhteessa vastaajien kykyihin. Tietotestin yksiuotteisuus oli matala, mutta kaikkien osioiden kiinnittymisarvot hyvät. Erottelun laajuus vastaajissa oli heikko, mutta osioissa korkea. Jatkossa mittarin ominaisuuksia, erityisesti vaikeustasoa ja sensitiivisyyttä, on tarpeen tarkastella muissa vastaajaryhmissä. Rasch analyysillä on monia käyttömahdollisuuksia hoitotieteessä uusien ja jo olemassa olevien mittareiden arvioinnissa. Klassisen testiteorian menetelmät ovat edelleen tarpeellisia ja käyttökelpoisia, mutta usein laajat aineistot vaativat osiovasteteorian mallien (kuten Rasch analyysi) hyödyntämistä.

ABSTRACT

Possibilities to use Rasch analysis in nursing science – foot care knowledge test as an example

*Minna Stolt, PhD, Docent, Professor
Miko Pasanen, MSc, Statistician
Riitta Subonen, PhD, Professor,
Director of Nursing (part-time)*

Rasch analysis, a part of Item Response Theory (IRT), is a statistical method that can be used to evaluate psychometric properties of instruments. Rasch analysis is based on assumption of integrating item difficulty and respondents' ability. The use of Rasch analysis in nursing science is scarce. The aim was to describe Rasch analysis as a method and its potential to evaluate psychometric properties of instruments developed in the field of nursing science. We used a descriptive cross-sectional study as an example where data were collected from persons with arthritis (n=129) with the Foot Care Knowledge Test (20 items, 3-point response options: correct, incorrect, I do not know). The data were analysed with Rasch analysis including unidimensionality, item goodness-of-fit, separation and differential item functioning. The unidimensionality and person separation of the Foot Care Knowledge Test were low. However, all other goodness-of-fit values were acceptable. In the future, the psychometric properties of the test could be beneficial to evaluate in different populations and reconsider the level of item difficulty and item separation. In

Avainsanat: Rasch analyysi, tilastollinen analyysi, mittarin kehittäminen, mittarin validointi, luotettavuus

nursing science, Rasch analysis has many advantages to be used while developing new instruments or validating existing ones. Classical test theory is still relevant, but analysis of large datasets could benefit from utilizing item response theory models, such as Rasch analysis.

Key words: Rasch analysis, statistical analysis, instrument development, instrument validation, reliability, validity

Mitä tutkimusaiheesta jo tiedetään?

- Rasch analyysi on osiovasteteoriaan kuuluva mittarin ominaisuuksien tilastollinen analyysimenetelmä ja se on vaihtoehtoinen klassiseen testiteoriaan (CTT) sisällyville ja paljon käytetyille tarkastelutavoille.
- Rasch analyysillä selvitetään vastaajien vastausten ja osioiden vaikeustason tarkastelulla mittarin kykyä tuottaa luotettavia tuloksia.
- Rasch analyysiä on hyödynnetty psykologian ja kasvatustieteen aloilla laajasti, mutta sen käyttö hoitotieteessä on toistaiseksi vielä vähäisempää.

Mitä uutta tietoa artikkeli tuo?

- Rasch analyysillä voidaan arvioida mittarin luotettavuutta sen kehitysvaiheessa tai testata jo olemassa olevan mittarin ominaisuuksia.
- Hoitotieteessä on kehitetty ja kehitetään mittareita hoitotieteellisten kohteiden tutkimiseksi, ja uudet menetelmät antavat vaihtoehtoisia tarkastelumahdollisuuksia.
- Mittarin yksiulotteisuus, osioiden ja vastaajien sopivuus Rasch malliin ja osioiden ja vastaajien erottelevuus ovat keskeisiä suureita Rasch analyysissä.

Mikä merkitys tutkimuksella on hoitotyölle, hoitotyön koulutukselle ja johtamiselle?

- Artikkelin sisältöä ja esitetyn esimerkkitutkimuksen tuloksia voidaan hyödyntää mittarin kehittämisessä ja olemassa olevien mittareiden testauksessa sekä hoitotieteen menetelmäopintojen laajentamisessa.
- Rasch analyysi täydentää klassiselle testiteorialle tyypillistä mittarin rakenteen arviointia tuottamalla tietoa yksittäisten osioiden keskinäisistä yhteyksistä tai kyvystä muodostaa laajempi kokonaisuus.

Tutkimuksen lähtökohdat

Rasch analyysi on yksi mittarin luotettavuuden arvioinnissa käytetty tilastollinen menetelmä, joka on osiovaste-teoriaan (*Item Response Theory, IRT*) kuuluva menettely. Rasch analyysin taustalla on tanskalaisen matemaatikon George Raschin vuonna 1960 kehittämä Raschin mittausteoria. Mittausteoria perustuu ajatukseen mittarin osioiden (kysymysten/väittämien) vaikeustasosta ja vastaajien kyvystä. Oletuksena on: Mitä korkeampi tai parempi vastaajan mitattava ominaisuus on, sitä paremmin hän menestyy

testissä ja vastaa tai ratkaisee yksittäiset osiot oikein (Rasch (1960/1980). Hoitotieteessä klassinen testiteoria (*Classical Test Theory*) on usein käytetty lähestymistapa, jossa hyödynnetään mm. pääkomponenttianalyysiä, faktorianalyysiä tai osioanalyysiä. Klassinen osioanalyysi (*Item analysis*, esim. Ferketich 1991) kohdistuu vastausten summamuuttujaan, joka on järjestyslukuasteikollinen muuttuja (Bortolotti ym. 2013). Rasch analyysissä puolestaan vastaukset voidaan logaritimuunnoksilla esittää intervalliasteikollisina (Bond & Fox 2015). Huolimatta Rasch analyysin erilaisista käyttömahdollis-

suuksista hoitotieteessä, sitä on toistaiseksi hyödynnetty melko vähän. Artikkelin tavoite on esitellä vielä vähän käytettyä menetelmää, joka erilaisella lähestymistavalla toisi instrumenttien luotettavuuden tarkasteluun uusia mahdollisuuksia. Suomenkielisiä lähteitä aiheesta on vähän, ja tilastollisia analyysejä käsiteltäessä puhutaan yleisesti menetelmän englanninkielisillä nimillä.

Mittari tarkoittaa usean osion muodostama kokonaisuutta. Mittarilla voidaan mitata mitattavan kohteen luonteen mukaisesti ilmiötä tai kohdetta suoraan tai epäsuorasti (Waltz ym. 2017). Suora mittaaminen tuottaa yleensä tarkan numeerisen arvon, kuten pituus, paino, verenpaine. Epäsuora mittaaminen kohdistuu jonkin abstraktin tai latentin piirteen mittaamiseen, kuten taito, tieto, osaaminen. Latenttia piirrettä ei voida luotettavasti mitata yhdellä osiolla, vaan sen mittaamiseen tarvitaan useita osioita. Mittaamisessa abstrakti käsite ensin määritellään ja määritelmän keskeiset määritteet muodostavat mittarin rakenteen ja sisällön (DeVellis 2017, Waltz ym. 2017). Rasch-analyysi tuottaa yhden mahdollisuuden tämänkaltaisten latenttien ominaisuuksien tarkasteluun (Bond & Fox 2015).

Rasch-analyysiä on käytetty hoitotieteessä instrumenttien luotettavuuden ominaisuuksien tarkasteluun (Stolt ym. 2022). Se on käyttökelpoinen uuden mittarin kehittämisessä tarkasteltaessa yksittäisten osioiden keskinäistä yhteyttä tai kykyä muodostaa laajempi kokonaisuus (Tennant ym. 2004, Spurlock & Wonder 2015, Boone 2016). Menetelmä on käyttökelpoinen myös jo olemassa olevan mittarin tarkasteluun (Prieto ym. 2003), esimerkiksi pyrittäessä vähentämään mahdollisia osioita (Boone 2016), pyrkimyksenä lyhentää mittaria tai tarkasteltaessa mittarin erikielisten versioiden yhdenmukaisuutta (*equivalence*) kielen (Suhonen ym. 2013, Stolt ym. 2016) tai eri kohderyhmän kyseessä ollessa.

Rasch-analyysi sisältää myös osiovaste-teorian mukaisen osioiden tarkastelun (*item ana-*

lysis). Osiovaste-teoria pohjautuu logistisiin funktioihin, joiden avulla saadaan tietoa osioiden vaikeudesta tai helppoudesta ja osioiden toimivuudesta. Osiovaste-teoriassa on kolmentasoisia malleja: yksi-, kaksi- ja kolmeparametriset mallit. Näistä yksiparametriset mallit ovat ns. Raschin malleja (*Rasch models*) ja niissä pääpaino on osion vaikeutta kuvaavassa (*item difficulty*) parametrissa. Kaksiparametriset mallit hyödyntävät osion vaikeutta (*difficulty*) kuvaavan parametrin lisäksi osion erottelukykystä (*discrimination*), osion ominaisuus erotella vastaajat eri kykytasolla kuvaavaa parametria. Kolmeparametriset mallit sisältävät osion vaikeus (*difficulty*)- ja erottelukykyparametrien (*discrimination*) lisäksi myös arvauskorjausparametrien (*guessing*), jolla huomioidaan arvauksen osuus vastauksissa (Bond & Fox 2015). Näitä kaikkia malleja voidaan analysoida Rasch-analyysin erilaisia testejä hyödyntämällä.

Raschin mittausteoria ja keskeiset arviointikohteet

Raschin mittausteoria perustuu vastaajien kykytason (*abilities*) ja heidän tuottamiensa vastausten samanaikaiseen arviointiin (Bond & Fox 2015). Rasch-analyysi soveltuu niin kaksi- kuin useampiluokkaisia muuttujia sisältävien aineistojen tarkasteluun. Rasch-analyysissä kohteena on aineiston perusteella tarkastella vastaajan todennäköisyyttä vastata oikein kuhunkin osioon. Näin ollen mittarin osioille ja vastaajille lasketaan ratkaisutodennäköisyyksien logaritimuunnos. Vastaajien kykyarvot ja osioiden vaikeustasot määritellään logaritmoituilla odds (*log odds*)-arvoilla (*calibrations*). Tämän jälkeen vastaajien kykyarvot ja osioiden vaikeustasot estimoidaan iterointimenetelmällä yhteen ja voidaan siten esittää samalla logistisella mitta-asteikolla (intervalliaasteikko). Toisin kuin klassinen osioanalyysi (*Item analysis*, esim. Ferketich 1991), Raschin mittausteoriassa testin ja vastaajien ominaisuudet voidaan erottaa toisistaan.

Raschin mallin valinta määrittyy mittarin vastausvaihtoehtojen määrän perusteella. Dikotominen Raschin malli soveltuu kaksiluokkaisille mittareille (dikotominen) ja moniluokkainen malli kolme tai useampi luokkaisille mittareille (*polytomous model*, Lerdal ym. 2016). Rasch analyysille on kehitetty useita ohjelmistoja, joista Winsteps, RUMM ja ConQuest ovat yleisimmin käytettyjä (Leung ym. 2014). Kukin ohjelmisto on omanlaisensa ja tutkija päättää mitä analyysijä suoritetaan. Koska yksittäisiä analyysijä on paljon, tässä artikkelissa kuvataan Rasch analyysin vähimmäisvaatimukset (ks. Tennant & Conaghan 2007) ja osoitetaan joitain esimerkkejä hyödyntäen Winsteps-ohjelman tulosteita. Rasch analyysi voidaan toteuttaa melko pienelläkin aineistolla. Yleinen tavoite aineiston koon suhteen on 100 vastaajaa mittariin, jossa on vähintään 10 osiota (Törmäkangas & Törmäkangas 2009, s.42).

Raschin mallin keskeinen lähtökohta on selvittää, sopiiko malli aineistoon. Mallissa on yksiulotteisuuden (*unidimensionality*, Kline 2005) oletus, jolloin vastaajien tuottamat vastaukset perustuvat ainoastaan kykyyn, jota mittari mittaa (esim. tieto). Kyseessä on epäsuora mittaus, joka kohdistuu latentin piirteen arviointiin. Mittarissa on useita osioita mitattavan ominaisuuden indikaattorina. Tämänkaltaisessa testauksessa peruslähtökohta on, että mitä parempi vastaajan mitattava ominaisuus on, sitä parempi on hänen menestymisensä testissä (Bond & Fox 2015). Tämän myötä voidaan olettaa, että vastaaja vastaa helpoimpiin osioihin oikein ja kykytasoaan vaikeampiin osioihin väärin. Yksiparametrinen Raschin malli ei sovellu tilanteisiin, jossa vastausten tuottamisessa tarvitaan useita samanaikaisia kykyjä, tarkastellaan useampi luokkaiseen vastausasteikkoihin perustuvaa aineistoa tai aineistoa kokonaisuudessaan (sisältäen oikeat ja väärät vastausvaihtoehdot). Näissä tilanteissa voi hyödyntää moniulotteisen osiovasteteorian (multidimensional item response theory, MIRT) malleja (esim. Multidimensional generalized

partial credit model tai Multidimensional Continuous Response model) tai koko aineistoa hyödyntävää IRT analyysiin pohjautuvaa tarkastelua (complete option choice information, Ramsay ym. 2020).

Rasch analyysissä tarkastellaan useita vastauksiin, vastaajiin ja mittariin liittyviä elementtejä, jotka kokonaisuudessaan muodostavat käsityksen mittarin luotettavuudesta. Kuvailevalla tiedolla osioiden jakaumahierarkiasta (*rating scale functioning*) voidaan päätellä, ovatko vastaajat käyttäneet kaikkia mittarin vastausvaihtoehtoja. Vastausten määrä jokaisessa vastausvaihtoehdossa kuvaa käytetyn asteikon laajuutta. Yleensä tasaisesti jakautuneet vastausmäärät ovat toivottavia. Jos vastausmäärät yksittäisissä vastausvaihtoehdoissa ovat matalia, laskee vaihtoehdon voimaa tuottaa luotettavia tuloksia. Yleensä vähintään 10 vastausta joka vastausvaihtoehdossa on suositeltavaa (Linacre 2002).

Yksiulotteisuus tarkoittaa mittarin mittaavan yhtä piirrettä kerrallaan, esimerkiksi tietoa tai taitoa. Yksiulotteisuutta tarkastellaan kokonaisuutena ja tarkkaa raja-arvoa ei voida asettaa. Yksiulotteisuus (*unidimensionality*) arvioidaan Raschin mallin residuaalien pääkomponenttianalyysillä (*principal component analysis of residuals*). Suositeltavaa on, että ensimmäinen komponentti selittäisi yli 50 prosenttia vaihtelusta ja seuraava komponentti enintään 5 prosenttia tai sen ominisarvo (*eigenvalue*) on alle 2 (Smith & Miao 1994, Raiche 2005). Jos seuraavan komponentin selitysosa on yli 5% ja ominisarvo on yli 2, on oletettavaa, että testi ei ole yksiulotteinen, vaan mittaamisessa on havaittavissa moniulotteisia piirteitä. Tässä tapauksessa oleellista on tarkempi osioiden tarkastelu ja ongelmallisten osioiden tunnistaminen. Voidaan selvittää, mikä kyseisissä osioissa aiheuttaa sopimattomuuden yksiulotteisuuden oletukseen. Joskus kyse voi olla osion terminologiasta, epäselvästä ilmaisusta tai kahden asian kysymisestä yhdessä kysymyksessä.

Yksiulotteisuuden tarkastelussa on tärkeää arvioida myös yksittäisten osioiden sopi-

vuus malliin. Rasch analyysissä sopivuuden tarkastelu kohdistuu osioiden kiinnitysparametriarvoihin (*item fit*). Kiinnitysparametrien arvot ilmoitetaan keskineliövirheen (MNSQ) ja standardoidun keskineliövirheen (ZSTD) lähipainotus (INFIT)- ja etäpainotus (OUTFIT) periaatteella. Oletusarvo näille suureille on 1 ja ne voivat vaihdella nollan ja äärettömän välillä. Tavanomaisesti keskineliövirheen lähipainotusperiaatteella arvioitujen arvojen suositeltava vaihteluväli on 0.6–1.4. Vaihteluväli voi olla myös esimerkiksi 0.7–1.3 tai 0.6–1.4 riippuen mittaustarkoituksesta (Bond & Fox 2015). Näiden arvojen lisäksi standardoitu normaalijakauman testiarvo (ZSTD) tulisi olla -2.0 ja +2.0 välillä (Wright & Linacre 1994). Arvoja tarkasteltaessa on tärkeää tunnistaa osiot, jotka eivät sovi malliin, toisin sanoen osiot, jotka ovat tasoltaan alle 0.6 tai yli 1.4. Osiot, joilla on matalia arvoja (alle 0.6) eivät toimi itsenäisesti suhteessa toisiinsa. Tässä tapauksessa osiot sopivat malliin liian hyvin, jolloin aineisto on ennustettavampi kuin malli odottaisi sen olevan. Toisin sanoen väittämät tuottavat yksimielisen vastauksen ja variaatio aineistossa on vähäistä. Tämä saattaa johtua siitä, että mittarissa on paljon samanlaisia osioita (kysymyksiä/väittämiä). Vastaavasti korkeat yli 1.4 arvot osoittavat heikkoa kiinnittymistä malliin, jolloin aineisto on vähemmän ennustettavissa kuin mitä malli odottaa sen olevan. (Tennant ym. 2004, Bond & Fox 2015.) Tavoitearvot alittavat tai ylittävät osiot on hyvä arvioida ja pyrkiä tunnistamaan, miksi ne ovat muista poikkeavia. Rasch analyysissä tavoitearvoista poikkeava osio voidaan jättää analyysin ulkopuolelle ja tarkastella mittarin toimivuutta ilman kyseistä osiota. Jos arvot kohenevat poiston myötä, voidaan kyseisen osion päätellä olevan huonosti toimiva ja tarvitsevan ilmaisullista tai teoreettista muokkautusta olettaen, että osia halutaan silti säilyttää mittarissa.

Rasch analyysillä voidaan selvittää myös mittarin kykyä erotella vastaajia mitattavan asian suhteen (*separation*). Yksilöiden erot-

televuusindeksillä (*person separation index*, PSI) tarkastellaan mittarin kykyä erotella vastaajat ryhmiin vastauksiensa perusteella. Erottelevuusindeksin arvon tulee olla yli 2 ja yksilöiden reliabiliteettiä yli 0.80, jotta mittarin voidaan sanoa erottelevan vastaajat kahteen luokkaan: hyvin testissä menestyvät ja heikosti testissä menestyvät (Fisher 1992). Osioiden erottelu (*item separation*) kuvaa, miten hyvin vastaajat vastauksillaan erottelevat testin osioita (Wright & Stone 1999). Osioiden erotteluindeksin suositeltava arvo on yli kolme; mitä suurempi arvo sen parempi osioiden erottelu ja mittaamisen tarkkuus ovat (Linacre 2012). Osioiden erottelun reliabiliteettiä suositellaan yli 0.90 (Smith ym. 2003). Rasch analyysin avulla voidaan laskea koko aineistolle reliabiliteettikerroin keskimääräisen virhevarianssin neliöjuuren avulla (*Root Mean Square Error*, RMSE). Rasch analyysissä on kaksi tapaa määrittää koko aineiston reliabiliteettia, joista ensimmäinen kohdistuu kokonaisvirhevarianssiin (muodostaa reliabiliteetin ns. alarajan) ja toinen mallinnuksen virhevarianssiin (reliabiliteetin yläraja). Kertoimet vaihtelevat 0 ja 1 välillä, jossa vähintään 0.70 on hyväksyttävä arvo (Smith 2001).

Rasch analyysissä voidaan myös visuaalisesti Wrightin kuvion (Wright map; Boone 2014) avulla tarkastella osioiden jakautumista suhteessa osioiden vaikeustasoon ja henkilöiden vastauksiin. Kuvioista nähdään testin vaikeus ja osio- ja vaikeustasojen keskittyneisyys ja tunnistetaan lattia- (*floor effect*) ja kattovaikutus (*ceiling effect*). Osioiden vaikeustasokuviossa helpoimmat osiot sijoittuvat kuvion alapäähän ja vaikeimmat sen yläpähän. Kuvioista tarkastellaan, ovatko osiot jakautuneet tasaisesti helpoimmasta vaikeimpaan ja onko useita samantasoisia (yhtä vaikeita) osioita tunnistettavissa. Tavoitetasona on, että vastaajien keskiarvo (M) ja osioiden vaikeustason keskiarvo (M) ovat samalla tasolla. Jos vastaukset kohdistuvat kuvion yläpähän, testin kyky erotella hyvät vastaajat on heikko. Vastaavasti, jos

vastaukset sijoittuvat kuvion alapäähän, testi ei erottele heikompia vastaajia hyvistä vastaajista (Bond & Fox 2015). Kuvioista voidaan tarkastella myös vastausten hierarkiaa ja tunnistaa vaikeusasteeltaan samantasoiset osiot (*item clustering*). Samantasoiset osiot ovat kuviossa samalla tasolla vierekkäin. Jos samantasoisia osioita on, yhden osion poistamista tai sen ilmaisua tulisi harkita suhteessa mittauksen tarkoitukseen (Törmäkangas & Törmäkangas 2009). Esimerkiksi, jos tavoite on mitata tiettyä ominaisuutta (esim. tieto) kysymysten vaikeustasolle olisi hyödyksi olla tasaisesti nouseva (Boone 2016).

Erilaisen osiotoiminnan tarkastelussa (*differential item functioning*, DIF) selvitetään osioparametrien estimaattien erilaisuutta perusjoukon eri ryhmissä, toisin sanoen analysoidaan toimivatko osiot samalla tavalla eri ryhmissä, esimerkiksi eri-ikäisten tai eri koulutustason omaavien välillä. Tarkastelu perustuu osioiden vaikeustason arviointiin esimerkiksi kahden ryhmän (miehet/naiset) välillä. Jos vaikeustaso ryhmien välillä poikkeaa yli 0.5 log odds -arvoa ja sen tilastollisen merkitsevyyden taso on ≤ 0.05 , kyseessä on erilainen osiotoiminta (osioharha, *item bias*). Osioden toiminta yksilotteisessa mittarissa tulisi olla samanlaista vastaajien taustoista huolimatta. (Bond & Fox 2015, Boone ym. 2016.) Jos esimerkiksi miehet ja naiset vastaavat tiettyihin osioihin systemaattisesti eri tavoin, tämä vaikuttaa mittarin kokonaistulosten arviointiin sukupuolten välillä. Näin ollen osaryhmätarkastelun avulla voidaan vahvistaa mittarin tulosten yleistettävyyttä.

Rasch analyysin avulla on mahdollista tarkastella sekä aineistoa että vastaajia monipuolisesti. Sen optimaalinen käyttö vaatii riittävän aineiston ja tutkijan kykyä ymmärtää ja soveltaa modernin testiteorian oletuksia terveystieteellisessä tutkimuksessa. Rasch analyysiä on hyödynnetty pitkään ja monipuolisesti erityisesti psykologian ja kasvatustieteiden tutkimuksessa (Bortolotti ym.

2013). Myös hoito- ja lääketieteiden alueella Raschin erilaisia käyttömahdollisuuksia on käytetty uusien tai olemassa olevien mittareiden luotettavuuden arvioinnissa esimerkiksi reumatologian (Tennant & Conaghan 2007), tuki- ja liikuntaelämäntien sairauksien (Leung ym. 2014), liikkumisen (Belvedere & de Morton 2010) tai hoitajien osaamisen arvioinnissa (Müller 2013). Hoitotieteessä on kehitetty paljon erilaisia mittareita. Pitkään käytettyjen tai juuri kehitettyjen mittareiden luotettavuuden arviointi on oleellista tarkkojen ja luotettavien tulosten saamiseksi, johon Rasch analyysi antaa uudenlaisia mahdollisuuksia.

Tutkimuksen tarkoitus ja tutkimusongelmat

Tutkimuksen tarkoituksena on kuvata Rasch analyysiä menetelmänä, sen käyttömahdollisuuksia ja sen tarjoamaa informaatiota mittarin psykometristen ominaisuuksien arvioinnissa hoitotieteen alalla. Tarkastelussa käytetään esimerkkinä Jalkojen omahoitotiedot -tietotestiä. Tavoitteena on laajentaa mittareiden kehittämisessä ja arvioinnissa hyödynnettävien uusien menetelmien käyttöönottoa hoitotieteessä.

Menetelmät

Tutkimus oli kuvaileva poikkileikkaustutkimus. Aineisto kerättiin paperikyselynä keväällä 2019 yhden potilasjärjestön jäseniltä. Kyselylomake lähetettiin kaikille (n=1318) järjestön aikuisille mitä tahansa reumatautia sairastavalle jäsenille. Tätä artikkelia varten aineistosta poimittiin nivelrikkoa sairastavien henkilöiden osa-aineisto. Kaikkiaan kyselyyn vastasi 504 henkilöä, joista 129 oli nivelrikkoa sairastavaa henkilöä.

Instrumentti

Aineisto kerättiin Jalkojen omahoitotiedot -tietotestillä (Stolt ym. 2015). Se koostuu 20 jalkojen omahoitoon liittyvästä väittämästä, jossa vastausasteikko on kolmiporainen (oikein, väärin, en tiedä). Väittämät jakautuivat viidelle osa-alueelle: ihon hoito (4 osiota), kynsien hoito (4 osiota), jalkojen rakenteellisten muutosten hoito (4 osiota), pitkäaikaissairauksiin liittyvien jalkavaivojen hoito (4 osiota) ja sopivien kenkien valinta (4 osiota). Mittarin sisäinen johdonmukaisuus on todettu hyväksi aikaisemmissa tutkimuksissa (CTT, Kuder Richardsson 0.81, Stolt ym. 2015). Taustamuuttujina kysyttiin ikä, sukupuoli ja koulutustaso.

Aineiston analyysi

Aineisto tarkastettiin, kuvailtiin (frekvenssi, prosenttiosuus, keskiarvo, keskihajonta) ja luokiteltiin uudelleen IBM SPSS Statistics for Windows, Version 25.0 -ohjelmalla (Armonk, NY: IBM Corp.). Kyselylomakkeen kolmiporainen vastausasteikko muutettiin tietotestin analyysille tyypillisesti kaksiluokkaiseksi luokittelemalla oikeat vastaukset ja väärät vastaukset. En osaa sanoa -vaihtoehto tulkittiin vääräksi vastaukseksi. Rasch analyysi tehtiin Winsteps 4.8.1.0 -ohjelmalla. Ensinnäkin tarkastettiin vastausasteikon käyttö. Yksiulotteisuus tarkasteltiin ominaisarvojen ja prosenttiosuusten perusteella. Sen jälkeen laskettiin osiokohdittaiset lähipainotus- ja etäpainotus arvot. Erotteluindeksi (*separation*) laskettiin yksilötasolla. Osoiden vaikeustason ja vastaajien vastausten sijoittumista tarkasteltiin Wrightin kuviolla. Lopuksi tarkasteltiin osaryhmätarkastelulla osioiden mahdollisesti erilaista osiotoimintaa taustamuuttujien (sukupuoli, ikä, koulutus) suhteen. Tätä varten taustamuuttujista ikä luokiteltiin kahteen luokkaan: alle 65-vuotiaat (46–64) ja 65-vuotiaat tai vanhemmat (65–86). Erilaista osiotoimintaa tarkasteltiin Mantel-Hänselin testillä, jossa p-arvo ≤ 0.05 tarkoittaa osiotoiminnan erilaisuutta ryhmien välillä (Bond & Fox 2015).

Tulokset

Vastaajien keski-ikä oli 69 vuotta (keskihajonta 8.8, vaihteluväli 46–86). Heistä valtaosa oli naisia (93%). Vastaajista noin kolmannes oli käynyt peruskoulun (n=49, 38%), kansakoulun (n=42, 33%) ja lukion (n=37, 29%).

Aineistossa uudelleen luokiteltua kaksiluokkaista vastausasteikkoa oli käytetty tasaisesti (*rating scale functioning*). Vastausvaihtoehtoja 1 (oikea vastaus) oli vähän yli puolet vastauksista (n=1510, 59%) ja vastausvaihtoehtoja 2 (väärin/ei tiedä, en osaa sanoa) vastaavasti hieman alle puolet (n=1048, 41%). Tasainen vastausvaihtoehtojen käyttö vahvistaa aineiston voimaa tuottaa luotettavia tuloksia. Jos vastausvaihtoehtoja olisi käytetty epätasaisesti, se vähentäisi aineiston vaihtelua ja siten heikentäisi analyysin luotettavuutta. Vastaajien kykyjakauman (*measure*) keskiarvo oli -0.46, mikä tarkoittaa, että vastaajien todellinen kyky vastata kysymyksiin oli alhaisempi kuin mitä osiot vaativat. Toisin sanoen testi oli vaikea suhteessa vastaajien kykyihin.

Yksiulotteisuus (*unidimensionality*) jalkojen omahoitotiedot -tietotestissä oli matalahko. Ensimmäinen komponentti selitti 40.0% vaihtelusta ja seuraava komponentti 6.1% (ominaisarvo [*eigenvalue*] 2.0364). Näiden perusteella voitaisiin tulkita, että tietotesti on moniulotteinen mittari. Tietotesti muodostuu viiden osa-alueen kokonaisuuksista (summuuttajat), joiden on teoreettisesti ajateltu kattavan jalkojen omahoidon kokonaisuuden. Mittarin yksiulotteisuuden (*unidimensionality*) tarkastelussa tulee vaihtelun lisäksi tarkastella myös yksittäisten osioiden sopivuutta malliin kiinnitysparemetrien (*item fit*) tarkastelun avulla. Osoiden kiinnitysparemetritarkastelussa keskineliövirheen (*MNSQ*) arvot asettuivat välille 0.85–1.16 (Taulukko 1). Keskineliövirheen (*MNSQ*) arvot olivat suositellulla tasolla (Taulukko 2). Yhtä osiota (kys9) lukuun ottamatta kaikkien osioiden standardoidut keskineliövirhearvot (*ZSTD*) olivat suositellulla tasolla (<2), välillä -2.26–1.56. Osoiden keskineliöpoik-

Taulukko 1. Yhteenvedo Rasch analyysin arviointikohteista, niiden tulokset ja suositellut tasot

Arviointikohte	Tulos	Suosittelut taso
Yksiuotteisuus (unidimensionality)		
ensimmäinen komponentti (1 st component)	40.0%	<50%
seuraava komponentti (2 nd component)	6.1%	<5% ja
	(ominaisarvo/eigenvalue 2.0364)	(ominaisarvo/eigenvalue <2)
Vastaaajien kykyjakauma (measure)		
Keskiarvo	-0.46	
Osioiden kiinnitysparemetrit (item fit)		
Osioiden kiinnitysparemetrien keskineliövirheiden (MNSQ) vaihteluväli	0.85–1.16	0.6–1.4
Osioiden standardoitujen keskineliövirhearvojen (ZSTD) vaihteluväli	-2.26–1.56	-2.0 ja +2.0 välillä
Osioiden keskineliöpoikkeamien (MNSQ) keskiarvo	1.00 (lähi/infit), 0.96 (etä/outfit)	lähellä arvoa 1
Osioiden standardoidut keskineliöpoikkeamat (ZSTD)	0.01 (lähi/infit), 0.09 (etä/outfit)	lähellä arvoa 0
Erottelu (separation)		
Vastaaajat (persons)	1.50	>2
Osiot (items)	6.04	>3
Reliabiliteetti (reliability)		
Vastaaajien reliabiliteetti (person reliability)	0.69	>0.80
Osioiden reliabiliteetti (item reliability)	0.97	0.90
Koko aineisto	0.69–0.73	>0.70
Erilainen osiotoiminta (differential item functioning, DIF)		
Sukupuoli	kys3, p=0.0231	>p=0.05
Ikä	kys12, p=0.0424 kys20, p=0.0322	

Taulukko 2. Osioiden kiinnitysparemetriarvot lähipainotusperiaatteella.

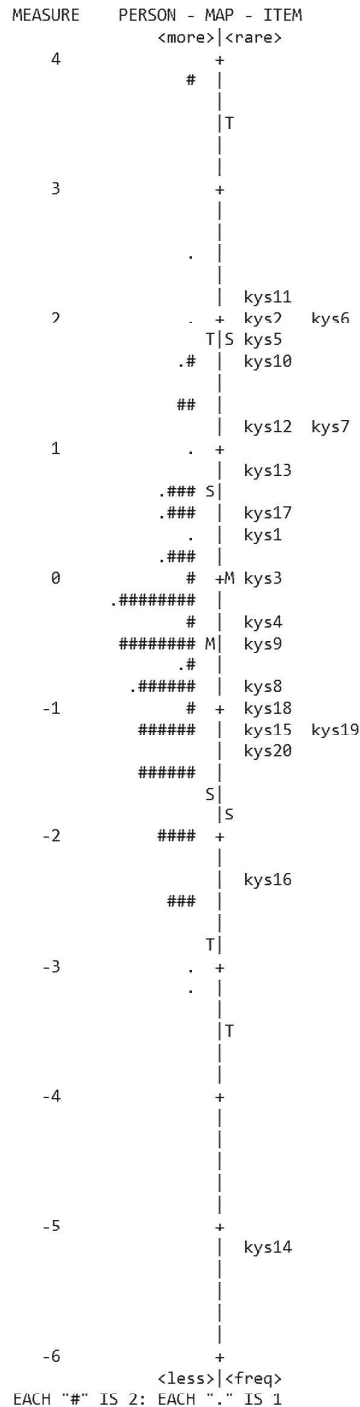
Osion numero (item)	Vaikeustason paremetriarvo (measure)	Mallin keskivirhe (standard error)	Lähipainotusperiaate (infit) Keskineliövirhe (MNSQ) ¹	Stadardoitu keskineliövirhe (ZSTD) ²
3	0.02	0.20	1.04	0.55
10	1.67	0.28	1.16	0.90
8	-0.76	0.20	1.11	1.56
12	1.21	0.24	0.95	-0.28
17	0.54	0.21	1.04	0.45
1	0.38	0.22	1.05	0.56
16	-2.39	0.25	1.04	0.29
5	1.76	0.28	1.02	0.19
19	-1.22	0.20	1.02	0.25
13	0.87	0.23	1.01	0.12
18	-1.08	0.20	1.01	0.12
7	1.21	0.24	1.00	0.04
14	-5.17	0.72	1.00	0.23
15	-1.13	0.20	1.00	-0.03
4	-0.33	0.20	0.99	-0.18
6	2.00	0.31	0.95	-0.15
2	2.01	0.31	0.93	-0.25
11	2.11	0.31	0.93	-0.22
20	-1.26	0.20	0.91	-1.07
9	-0.46	0.20	0.85	-2.26

¹ suositeltu vaihteluväli 0.60–1.40² suositeltu vaihteluväli -2.0–2.0

keamien keskiarvot (*MNSQ/ZSTD mean*) olivat lähellä arvoa 1, joten tämän aineiston arvot 1.00 (lähipainotus, *infit*) ja 0.96 (etäpainotus, *outfit*) olivat erinomaiset. Lähi- ja etäpainotettujen standardoitujen keskineliöpoikkeamien keskiarvot olivat oletuksen mukaisesti lähellä nollaa (0.01/lähipainotus ja 0.09/etäpainotus). Osioiden osalta malli sopii aineistoon erinomaisesti.

Wrightin kuviossa (Kuvio 1) osiot ovat vasemmalla ja vastaajat pystykatkoviivan oikealla puolella. Osiot ovat vaikeusjärjestyksessä vaikein ylhäällä ja helpoin alhaalla. Kuvioista selviää, että osio "kys14" on helpoin ja "kys11" vaikein. Muut osiot ovat järjestyneet jatkumon keskivaiheille. Kolme osiota ("kys6", "kys7" ja "kys19") ovat samalla tasolla jonkin toisen osion kanssa, ollen vaikeustasoltaan samankaltaisia. Näitä kolmea osiota on jatkossa syytä tarkastella niiden sisällön näkökulmista ja harkita onko niissä jotain mahdollisuutta muuttaa tai poistaa päällekkäistä vaikeutta suhteessa muihin osioihin. Kun tavoitellaan mahdollisimman lyhyttä, tarkkaa ja vastaajaystävällistä mittaria, joi-sain tapauksissa osioiden poisto voisi olla tarpeen. Poistot tulee kuitenkin tehdä huolellisesti harkiten suhteessa mittarin tarkoitukseen ja sen taustalla oleviin teoreettisiin lähtökohtiin (*theory vrt.* Boone 2016).

Vastaajien erottelun laajuus (*person separation*) kuvaa vastaajien kykyjakauman laajuutta. Tässä aineistossa erottelun luku oli 1.50 (Kuvio 2). Vastaajien reliabiliteetti (*person reliability*) oli matala (0.69, tavoite 0.80). Erottelun perusteella vastaajat ovat jakautuneet melko suppealle alueelle, jolloin testin erottelun laajuus ei ole kovin hyvä. Saman tuloksen voi tulkita myös Wright kuvioista (Kuvio 1). Kuviossa vastaajien keskiarvo (M) ja osioiden vaikeustason keskiarvo (M) ovat melko lähellä toisiaan. Vastauksista valtaosa sijoittuu kuvion keski- tai alaosaan, kun taas valtaosa kysymyksistä on kuvion yläosassa, tämä indikoi kysymysten vaikeutta suhteessa vastaajien kykyihin. Matala vastaajien erottelu (<2, *person reliability index* < 0.8) kuvastaa mitta-



Kuvio 1. Wrightin kuvio, jossa tietotestiin vastaajat (merkitty symboleilla "#" = 2 vastaajaa, "." = yksi vastaaja) ovat kykyjensä mukaisessa järjestyksessä keskiviivan vasemmalla puolella ja osiot vaikeusjärjestyksessä oikealla.

SUMMARY OF 129 MEASURED PERSON

	TOTAL	COUNT	MEASURE	MODEL	INFIT		OUTFIT	
	SCORE			S. E.	MNSQ	ZSTD	MNSQ	ZSTD
MEAN	28.0	19.8	-.46	.60	1.00	.01	.96	.09
SEM	.3	.0	.10	.01	.03	.09	.06	.06
P. SD	3.4	.5	1.17	.09	.30	1.02	.65	.70
S. SD	3.4	.5	1.17	.09	.30	1.03	.66	.70
MAX.	39.0	20.0	3.82	1.06	1.91	2.59	6.20	4.12
MIN.	20.0	17.0	-3.20	.54	.48	-2.39	.16	-1.09
REAL RMSE	.65	TRUE SD	.97	SEPARATION	1.50	PERSON RELIABILITY	.69	
MODEL RMSE	.61	TRUE SD	.99	SEPARATION	1.63	PERSON RELIABILITY	.73	
S. E. OF PERSON MEAN = .10								

SUMMARY OF 20 MEASURED ITEM

	TOTAL	COUNT	MEASURE	MODEL	INFIT		OUTFIT	
	SCORE			S. E.	MNSQ	ZSTD	MNSQ	ZSTD
MEAN	180.3	127.9	.00	.26	1.00	.04	.96	-.01
SEM	7.5	.4	.40	.03	.02	.17	.03	.16
P. SD	32.8	1.6	1.75	.11	.07	.73	.14	.69
S. SD	33.6	1.6	1.80	.12	.07	.75	.14	.71
MAX.	256.0	129.0	2.11	.72	1.16	1.56	1.29	1.89
MIN.	142.0	122.0	-5.17	.20	.85	-2.26	.72	-1.57
REAL RMSE	.29	TRUE SD	1.73	SEPARATION	6.04	ITEM RELIABILITY	.97	
MODEL RMSE	.28	TRUE SD	1.73	SEPARATION	6.09	ITEM RELIABILITY	.97	
S. E. OF ITEM MEAN = .40								

Kuvio 2. Näyttökuva Winsteps -ohjelman tulosteesta, josta näkyy tietotestin vaikeustaso (MEASURE), kiinnitysparametrien keskiarvot (Infit: MNSQ, ZSTD), vastaajien ja osioiden erottelu (PERSON/ITEM SEPARATION) ja reliabiliteetti (PERSON/ITEM RELIABILITY).

rin heikkoa sensitiivisyyttä ja jatkossa osioiden lisääminen mittariin voisi olla tarpeen.

Osioiden erottelun (*item separation*) laajuus oli 6.04 (Kuvio 2). Suositeltavaa on, että arvo on mahdollisimman korkea (vaihtelee nollan ja äärettömän välillä). Tässä aineistossa erottelu oli korkea, jolloin vastaajat pystyivät erottelemaan osiot toisistaan hyvin. Osioiden reliabiliteetin hyväksymisraja on 0.90 (Smith 1992), joka tässä aineistossa ylittyi selvästi (arvo 0.97).

Tässä aineistossa erilaista osiotoimintaa (*differential item functioning, DIF*) tarkasteltiin kolmen taustamuuttujan (sukupuoli, ikä ja koulutus) näkökulmista. Erilaista osiotoimintaa oli naisten ja miesten vastauksis-

sa osiossa ”kys3” ($p=0.0231$). Iän osalta erilaista osiotoimintaa oli alle 65-vuotiaiden ja yli 65-vuotiaiden vastauksissa osioissa ”kys12” ($p=0.0424$) ja ”kys20” ($p=0.0322$). Koulutuksen osalta erilaista osiotoimintaa ei havaittu. Havaittu erilainen osiotoiminta on tärkeä analysoida ja selvittää onko osioissa sisällöllistä tai ilmaisullista virhettä, mikä olisi tarpeen korjata. Sukupuolen osalta on huomattava mies -vastaajien vähäinen määrä (7%) aineistossa, joka osaltaan voi haitata luotettavaa osiotoiminnan tarkastelua.

Koko aineiston reliabiliteetikertoimet tässä aineistossa vaihtelivat 0.69 ja 0.73 välillä. Varianssi on alhainen, mutta hyväksyttävä (<0.70 , Smith 2001).

Pohdinta

Tutkimuksessa hyödynnettiin monipuolisesti Rasch analyysiä Jalkojen omahoitotiedot -tietotestin psykometrinen ominaisuuksien arviointiin tuottaen tietoa tietotestin vastausasteikon toimivuudesta, aineiston sopivuudesta Rasch malliin, tietotestin kyvystä erotella henkilöitä ja sen toimivuudesta perusjoukon eri ryhmissä. Menetelmää on käytetty suhteellisen vähän esimerkiksi hoitotieteessä (Stolt ym. 2022), ja menetelmän hyödyt erityisesti mittareiden ominaisuuksien tarkasteluun ovat laajasti hyödyntämättä.

Jalkojen omahoitotiedot -tietotestin yksilutteisuus oli 40% mikä on alle suositellun 50 prosentin alarajan (Smith & Miao 1994, Raiche 2005). Tietotesti voi todennäköisesti muodostua useammasta kuin yhdestä kokonaisuudesta ja siksi sen osioiden yksityiskohtainen tarkastelu voi olla tarpeen. Toisaalta osioiden kiinnitysparametriaarvot lähipainotusperiaatteella olivat kaikki tavoitellulla tasolla, joka puolestaan osoitti aineiston sopivan hyvin Raschin malliin. Osiotarkastelussa kolme osiota sijoittui vaikeustasoltaan samalla tasolle muiden osioiden kanssa. Kaikkiaan osioiden sisällöllinen ja ilmaisullinen tarkastelu tulevaisuudessa on tärkeää, jotta mahdollisesti päällekkäiset osiot voidaan tunnistaa ja poistaa tai uudelleen nimetä. Aineiston koko (n=129) tässä tutkimuksessa jäi teoreettista tavoitetta (10x20=200) pienemmäksi, joka osaltaan heikentää tutkimuksen luotettavuutta, mutta siitä huolimatta aineistosta oli mahdollista analysoida mittarin ominaisuuksia.

Tietotesti kykeni erottelamaan vastaajia, mutta toisaalta erottelun laajuus oli suppeaa. Tietotesti erotteli vastaajat kahteen ryhmään: oikein vastaajat ja väärin vastaajat. Jalkojen omahoidon asiat olivat vastaajille pääosin tuttuja, mutta jotkin yksittäiset asiat olivat paikoin vastaajille vaikeita. Vaikka erottelun laajuus oli kriteereihin nähden suppea, on syytä muistaa, että tietotestin vastausasteikko oli alun perin kolmiluokkainen (oi-

kein, väärin, en tiedä), joka tietotestin analyysille tyypillisesti luokiteltiin kaksiluokkaiseksi (oikein, väärin). Vastausasteikkoon suhteutettuna tietotesti kykeni tunnistamaan vastaajat, jotka vastasivat osioihin joko oikein tai väärin, mikä tietotestityyppisessä mittarissa on tavoitteen mukaista. Jatkossa voisi harkita laajennetun IRT analyysin käyttöä (Ramsay ym. 2020), jolloin aineistoa voisi hyödyntää kolmiluokkaisena ja siten arvioida sen psykometrisia ominaisuuksia.

Tutkimus toteutettiin hyvää tieteellistä käytäntöä (TENK 2012) noudattaen. Tutkimukselle saatiin eettisen toimikunnan puoltava lausunto (koodi: 8/2021, 29.1.2018) ja tutkimuslupa organisaation käytänteiden mukaisesti. Tutkimukseen osallistuneet saivat kirjallisesti tietoa tutkimuksen tarkoituksesta, nimettömyydestä, aineistonkeruun toteutuksesta, vastausten luottamuksellisesta käsittelystä ja raportoinnista. Vastaaminen kyselylomakkeeseen tulkittiin tietoon perustuvaksi suostumukseksi. Kysely toteutettiin paperikyselynä. Aineisto säilytettiin yliopiston tietoturvallisessa Seafile -pilvipalvelussa ja pääsy alkuperäiseen aineistoon oli vain vastuututkijalla (MS).

Johtopäätökset

Tässä esimerkinomaisesti tarkastellussa Jalkojen omahoitotiedot -tietotestin yksilutteisuus oli matalahko, mutta kaikki mittarin osiot kuitenkin sopivat lähipainotusperiaatteen tukemana Raschin malliin. Tietotestin kysymyksistä osa oli vaikeampia, jolloin vain osa vastaajista kykeni vastaamaan osioihin oikein. Tämä on tavoiteltava mittarin ominaisuus, koska se paljastaa ne vastaajat, jotka todella tietävät.

Rasch analyysillä on monia käyttömahdollisuuksia hoitotieteessä. Klassisen testi-teorian menetelmät ovat edelleen tarpeellisia ja käyttökelpoisia, mutta usein hoitotieteessäkin kerätyt esimerkiksi kansainväliset aineistot, vaativat osiovasteteorian mallien

(kuten Rasch analyysi) hyödyntämistä. Rasch analyysi soveltuu hoitotieteen alalle erittäin hyvin, koska usein instrumentaation tavoitteena on mitata esimerkiksi osaamista, käyttäytymistä tai asenteita. Rasch analyysin, kuten vastaavien muidenkin edistyneiden analyysimenetelmien, toteuttaminen vaatii tutkijalta tilastotieteen osaamista, vahvaa teoreettista ymmärrystä Raschin mitausteoriasta ja käytännön taitoja hallita analyysiohjelmistoa. Osaamisen tueksi tämä artikkeli tuottaa suomenkielistä menetelmä-

kuvausta kansainvälisen kirjallisuuden rinnalle, jota voidaan hyödyntää hoitotieteen analyysimenetelmien laajentamisessa ja menetelmäopinnoissa.

VASTUUALUEET

Tutkimuksen suunnittelu MS, MP, RS, aineistonkeruu MS, aineiston analysointi MS, MP, käsikirjoituksen kirjoittaminen MS, MP, RS ja kommentointi MP, RS.

LÄHTEET

- Belvedere SL. & de Morton NS. (2010) Application of Rasch analysis in health care is increasing and is applied for variable reasons in mobility instruments. *Journal of Clinical Epidemiology* **63**(12), 1287–1297.
- Bond TG. & Fox CM. (2015) *Applying the Rasch model. Fundamental measurement in the human sciences*. 3rd ed. New York, NY: Routledge.
- Boone WJ. (2016) Rasch analysis for instrument development: Why, when, and how? *CBE Life Sciences Education* **15**, rm4.
- Boone WJ., Staver JR. & Yale, MS. (2014) Wright maps: First steps. Teoksessa: *Rasch Analysis in the Human Sciences*. Dordrecht: Springer.
- Bortolotti SLV., Tezza R., de Andrade DF., Bornia AC. & de Sousa Jr. AF. (2013) Relevance and advantages of using the item response theory. *Quality and Quantity* **47**, 2341–2360.
- DeVellis RF. (2017) *Scale development: theory and applications*. Los Angeles: Sage.
- Ferketich S. (1991) Focus on psychometrics. Aspects of item analysis. *Research in Nursing and Health* **14**(2), 165–168.
- Fisher W. (1992) Reliability, separation, strata statistics. *Rasch Measurement Transactions* **6**(3), 238.
- Hong I. & Bonilha HS. (2017) Psychometric properties of upper extremity outcome measures validated by Rasch analysis: a systematic review. *International Journal of Rehabilitation Research* **40**(1), 1–10.
- Kline TJ. (2005) Modern test theory: assumptions, equations, limitations, and item analyses. Teoksessa: Kline, TJ. *Psychological testing: A practical approach to design and evaluation*. Thousand Oaks, CA: SAGE Publications Ltd.
- Linacre JM. (2002) Optimizing Rating Scale Category Effectiveness. *Journal of Applied Measurement* **3**(1), 85–106.
- Linacre JM. (2004) Rasch model estimation: Further topics. *Journal of Applied Measurement* **5**(1), 95–110.
- Linacre JM. (2011) Rasch measures and unidimensionality. *Rasch Measurement Transactions* **24**(4), 1310.
- Linacre JM. (2012) A user guide to Winsteps Ministep Rasch model computer programs: Program manual 3.75.0. PDF-julkaisu. <http://www.winsteps.com/a/winstepsmanual.pdf> (29.3.2022)
- Leung YY., Png ME., Conaghan P. & Tennant A. (2014) A systematic literature review on the application of Rasch analysis in musculoskeletal disease – A special interest group report of OMERACT 11. *Journal of Rheumatology* **41**(1), 159–164.
- Müller M. (2013) Nursing competence: psychometric evaluation using Rasch modelling. *Journal of Advanced Nursing* **69**(6), 1410–1417.
- Prieto L., Alonso J. & Lamarca R. (2003) Classical Test Theory versus Rasch analysis for quality of life questionnaire reduction. *Health and Quality of Life Outcomes* **1**, 27.
- Raiche G. (2005) Critical eigenvalue sizes in standardized residual principal component analysis. *Rasch Measurement Transactions* **19**(1), 1012.
- Ramsay J., Wiberg M. & Li J. (2020) Full information optimal scoring. *Journal of Educational and Behavioral Statistics* **45**, 297–315.
- Rasch G. (1960/1980). *Probabilistic models for some intelligence and attainment tests*. Chicago: The University of Chicago Press.
- Smith Jr., EV. (2001) Reliability of measures and validity of measure interpretation: A Rasch measurement perspective. *Journal of Applied Measurement* **2**(3), 281–311.
- Smith RM., Linacre JM. & Smith, Jr. EV. (2003) Guidelines for manuscripts. *Journal of Applied Measurement* **4**, 198–204.
- Smith RM. & Miao CY. (1994) Assessing unidimensionality for Rasch measurement. Teoksessa: Wilson M. *Objective measurement: Theory into practice*. Greenwich: Ablex.
- Spurlock D. Jr & Wonder AH. (2015) Validity and reliability evidence for a new measure: The evidence-based practice knowledge assessment in nursing. *Journal of Nursing Education* **54**(11), 605–613.
- Stolt M., Suhonen R., Puukka P., Viitanen M., Voutilainen P. & Leino-Kilpi H. (2015) Nurses' knowledge of foot care in the context of home care: a cross-sectional correlational survey study. *Journal of Clinical Nursing* **24**(19–20), 2916–2925.

- Stolt M., Charalambous A., Radwin L., Adam C., Katajisto J., Lemonidou C., Patiraki E., Sjövall K., Suhonen R. (2016) Measuring trust in nurses – Psychometric properties of the Trust in Nurses Scale in four countries. *European Journal of Oncology Nursing* **25**(1), 46–54.
- Stolt, M., Kottorp, A., & Suhonen, R. (2022). The use and quality of reporting of Rasch analysis in nursing research: A methodological scoping review. *International Journal of Nursing Studies* **132**, 104–244.
- Suhonen R., Schmidt L., Katajisto J., Berg A., Idvall E., Kalafati M., Land L., Lemonidou C., Välimäki M. & Leino-Kilpi H. (2013) Cross-cultural validity of the Individualised Care Scale - A Rasch model analysis. *Journal of Clinical Nursing* **22**(5–6), 648–660.
- Tennant A., McKenna SP. & Hagell P. (2004) Application of Rasch analysis in the development and application of quality of life instruments. *Value Health* **7**(Suppl 1), S22–S26.
- Tennant A. & Conaghan PG. (2007) The Rasch measurement model in rheumatology: What is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis & Rheumatism* **57**(8), 1358–1362.
- TENK. (2012) Hyvä tietellinen käytäntö ja sen loukausepäilyjen käsitteleminen Suomessa. PDF-julkaisu. <https://tenk.fi/fi/ohjeet-ja-aineistot/HTK-ohje-2012> (21.6.2022)
- Törmäkangas K. & Törmäkangas T. (2009) *Osoanalyysi testien arvioinnissa*. Jyväskylän yliopisto. Koulutuksen tutkimuslaitos. Jyväskylä.
- Waltz CF., Strickland OL & Lenz ER. (2017) *Measurement in nursing and health research*. 5th ed. New York: Springer Publishing Company.
- Wright BD. & Linacre JM. (1994) Reasonable mean-square fit values. *Rasch Measurement Transactions* **8**(3), 370.
- Wright BD. & Stone MH. 1999. *Measurement essentials*. 2nd ed. Wilmington: Wide Range Inc.

Minna Stolt, TtT, dosentti, yliopistonlehtori, Hoitotieteen laitos, 20014 Turun yliopisto, minna.stolt@utu.fi

Miko Pasanen, FM, statistikko, Hoitotieteen laitos, 20014 Turun yliopisto, misapas@utu.fi

Riitta Subonen, TtT, professori, Hoitotieteen laitos, 20014 Turun yliopisto, riitta.subonen@utu.fi