

Oliver Saal ja työryhmä:

Kiitos Vaiheen 2 käsikirjoituksesta. Joululomien ja Suomen Akatemian alkuvuoden hakujen vuoksi arviointiprosessi kesti tällä kertaa vähän kauemmin, mutta nyt kaikki arviot ovat saapuneet. Arviot ovat huolellisia ja niissä on arvokkaita huomioita. Yleisesti ottaen kaikki ovat tyytyväisiä, mutta kaipaavat ennen kaikkea pohdinnan syventämistä. Olen samaa mieltä: etenkin tilanteessa, jossa tulokset ovat ristiriitaisia ja viittaavat menetelmällisiin ja/tai käsitteellisiin haasteisiin tutkimuskentällä, olisi tärkeää käydä läpi pohdinnassa erilaisia selityksiä ja tuottaa mahdollisesti uusia hypoteeseja siitä, miten tutkimuskenttä voisi oppia tuloksista ja tehdä tulevaisuudessa asioita toisin. Yksi hyvä lähde ns. negatiivisten tulosten raportointiin on (ks. Kuva/Taulukko 1):

Peltonen, E., Mohan, N., Zdankin, P., Shreedhar, T., Nguyen, T., Bayhan, S., ... & Nicklas, D. (2023). Perspectives on Negative Research Results in Pervasive Computing. IEEE Pervasive Computing, 22(3), 63-72. DOI: 10.1109/MPRV.2023.3273718

On varmasti totta, että epäasiallisen ja väkivaltaisen kohtelun käsitteiden monitulkintaisuus on keskeinen haaste. Olen arvioijien kanssa samaa mieltä, että tätä olisi syytä purkaa sekä menetelmällisesti että teoreettisesti.

Menetelmien osalta asiaa voisi lähestyä esim. ilmivaliditeetin (face validity) kannalta haastatteleamalla erilaista kohtelua kokeneita ihmisiä ja tutkimalla miten sisältö vastaa näitä kokemuksia. Vaikka aiemmassa vaiheessa mukana oli ammattilaisten työryhmä, voivat ihmisten arkiset kokemukset ja tulkinnat poiketa. Toinen mahdollisuus olisi yhdistää potentiaalisen mittarin itemeihin avoimia kysymyksiä ("cognitive pretesting"), joiden avulla vastaajat kertovat minkälaisia kokemuksia he yhdistävät kysymykseen. Teorian kannalta voisi vuorostaan olla tärkeää käydä läpi esim. onko tutkimuskentällä tarve peruskäsitteiden ja niiden suhteiden päivittämiselle. Mieleen tulee etenkin Haslamin "concept creep" – paljon käytetyillä käsitteillä on taipumus laajentua, jolloin niiden teoreettinen ja käytännöllinen teho heikkenee:

Haslam, N. (2016). Concept creep: Psychology's expanding concepts of harm and pathology. Psychological inquiry, 27(1), 1-17. DOI: 10.1080/1047840X.2016.1082418

Etenkin kolmas arvioija esittää useita menetelmällisiä huomioita, jotka on hyvä ottaa vakavasti. Teknisenä huomiona: erilaisten ratkaisujen kokeileminen on tosiaan sallittua ja toivottua, mutta suosittelen hyödyntämään liitteitä niin, että artikkeli pysyy luettavana ja ylimääräisten testien yksityiskohtia siirretään OSF:lle. Arvioijan ehdottama jako suunniteltuihin ja suunnittelemttomisiin analyysien yleensä selkeyttää rakennetta huomattavasti. Silloin on helppoa pudottaa vähemmän oleelliset suunnittelemttomat analyysit OSF:lle ja viitata pelkällä linkillä, kun lukija ymmärtää että tieto on ylimääräistä.

Mainitsen vielä erikseen, että arvioijan esittämä ratkaisu H1 raportoinnista on mielestäni toimiva. On ok todeta, ettei suunnitelman mukaista testiä ollut mahdollista toteuttaa, minkä vuoksi sen sijaan suoritetaan kokeileva testi. Tämä on jokseenkin tavallinen osa tieteellistä prosessia, eikä siinä ole mitään väärää.

Toivottavasti koette arviot hyödyllisiksi, ja kuten aiemmin, minuun voi ottaa yhteyttä jos/kun uusia kysymyksiä nousee matkalla.

Veli-Matti Karhulahti

Arvio 1: 28.12.2024

Arvioitavan tutkimuksen tarkoituksena oli selvittää kehitetyn epäasiallisen ja väkivaltaisen kohtelun mittarin suomen- ja ruotsinkielisten versioiden käytettävyys. Artikkelissa muodostettiin aikaisemmasta tutkimuksesta epäasiallisen ja väkivaltaisen kohtelun 28-osainen suomen- ja ruotsinkielisen mittari. Kehitettyä mittaria testattiin valtakunnallisessa vuonna 2024 kerätyssä kyselyaineistossa. Käsikirjoitus on osa Palosuojelurahaston rahoittamaa tutkimushanketta "Pelastustoimen nollatoleranssi syrjinnälle ja häirinnälle – faktaa vai fiktiota?" Muutamia ajatuksia ja kommentteja 2 vaiheen tekstiin.

1. Tutkimuksen aineisto käsitti 960 kyselyyn vastaajaa, joista miehiä oli noin 65 % ja kohderyhmästä noin 60 % oli pelastuslaitoksista ja ensihoidosta. Aineiston koko ja yksipuoleisuus vaikuttaa tuloksiin ja tätä tulee käsitellä pohdinnassa vielä tarkemmin.
2. Kuviossa 2 esitetään kohtelun ja työuupumuksen muuttujien välisiä korrelaatioita. Kuviossa muuttujien nimet ovat esitetty muodossa k12_5, k3_2 jne. Nämä lyhenteet olisi hyvä lisätä Taulukon 6 ensimmäiseen sarakkeeseen selkeyden vuoksi.
3. Tutkimus jakautui kahteen osatutkimukseen, joista ensimmäinen osuus on eksploraatiivinen ja siinä selvitettiin kehitetyn mittariston faktorirakennetta. Ensimmäisessä vaiheessa tähän oli suunniteltu tutkimusjoukoksi 20 % aineistosta, vähintään 200 vastaajaa. Tämä osuus toteutettiin 30 % satunnaisotoksella aineistosta, 280 vastaajaa. Sopivuus testaukset osoittivat ongelmia datan sopivuuden kanssa ja sen vuoksi kirjoittajat suorittivat lisäksi testauksen 14 omakohtaisesti koetun kohtelun muuttujalle. Tulokset olivat ristiriitaisia ja kirjoittajat päättivät suorittaa 4 eksploraatiivista faktorianalyysimallia käyttäen kaikkia muuttujia ja toisessa tarkastelussa omakohtaisesti koetun kohtelun muuttujia. Nämä molemmat mallit sovitettiin yhden ja kahden faktorin malleina. Näistä malliksi valikoitui 2 faktorin malli, jossa mukana olivat omakohtaisesti koetun kohtelun muuttujat, verrannollisesti paremman sopivuuden perusteella. Tätä ratkaisua on perusteltu tuloksissa ja johtopäätöksissä riittävästi.
4. Toinen osuus oli konfirmatorinen ja siinä tarkasteltiin ensimmäisen osatutkimuksen faktorirakenteen valideettia ja invarianssia, $n=680$. Konfirmatorisen faktorianalyysin muuttujat rajattiin lataamaan ensimmäisessä osuudessa valitun mallin mukaisesti ja modifikaatioindeksit antoivat viitteitä mahdollisesta bifaktorirakenteesta, jossa epäasialliselle ja väkivaltaiselle kohtelulle oli ympäristökohtaiset syyt ja yksi yleisfaktori. Ulkoisen valideetin tarkastelut työuupumuksen mukaan osoittivat vaihtelevia tuloksia. Tätä tulisi tarkastella lisää pohdinnassa.
5. Mittaristo ei ollut täysin yhdenmukainen kaikissa konteksteissa (ryhmissä, suureissa ja malleissa). Tämä saattaa vaikuttaa johtopäätöksiin ryhmien välisistä eroista tai mittauksen luotettavuudesta. Tämä vaihtelevuus tulee huomioida vielä tarkemmin pohdinnassa.
6. Pohdintaosuudessa tuodaan esiin mittariston toimivuuteen vaikuttavia tekijöitä kuten väkivallan määrittelyn ongelman. Pohdintaa mittarin soveltuvuuden osalta tulisi lisätä. Pohdinnan loppuosa on yleisellä tasolla ja siinä tulisi ottaa selkeämmin kantaa kehitetyn mittarin käytettävyyteen.
7. Käsikirjoitus tulee tarkistaa vielä aikamuotojen ja oikeinkirjoituksen osalta.

Arvio 2: 10.1.2025

Kiitos mahdollisuudesta arvioida käsikirjoitus "Epäasiallisen ja väkivaltaisen kohtelun suomen- ja ruotsinkielisen mittarin validointi pelastus- ja ensihoitoalalla". Tutkimuksessa testataan epäasiallisen ja väkivaltaisen kohtelun mittaria. Arvioni mukaan käsikirjoitus vaatii vielä joitain muutoksia ennen julkaisua.

Abstraktissa esitetään selkeä yhteenveto tutkimuksen sisällöstä. Harkitsisin kuitenkin ensimmäisen lauseen muokkausta. Johdannossa viitataan myös tutkimuksiin epäasiallisen ja väkivallan yleisyydestä, mutta abstraktissa viitataan vain uutisoinnin perusteella syntyneeseen mielikuvaan.

Siltä osin kuin pystyn arvioimaan menetelmiä, ne näyttävät myös toteutuneessa muodossaan vastaavan tutkimuskysymyksiin. En huomannut sellaisia poikkeamia suunnitelmasta, joita ei olisi perusteltu. Hienoa, että tutkimuksessa on laskettu tilastollinen voima ja huomioitu kato! Tämä aina mahdollistaa tutkimuskysymykseen vastaamisen luotettavammin. Faktorianalyysi ei kuitenkaan ole oman osaamiseni ydintä, joten en pystynyt arvioimaan analyysien asianmukaisuutta tarkasti. Muut arvioijat ovat toivottavasti pystyneet ottamaan tähän kantaa paremmin.

Lopputulosten tekstin oikeinkirjoitus ja tekstin sujuvuus olisi hyvä vielä tarkistaa. Tämä on pieni asia, mutta parantaa tekstin luettavuutta. Esimerkiksi lauseessa "Kohtelun tekoympäristö näytti vaikuttavat kohtelun esiintyvyyteen eri tavoin" on kirjoitusvirhe (vaikuttavat pitäisi olla vaikuttavan). Lause "Ensimmäiseen, eksploratiiviseen tutkimuskysymykseen voimme todeta" olisi sujuvampi esim. "Vastauksena ensimmäiseen, eksploratiiviseen tutkimuskysymykseen voimme todeta".

Lopputuloksissa oli useita kohtia, joita olisi hyvä vielä selkeyttää. Tärkeimpänä tekstissä voisi pohtia enemmän sitä, että mitä tulokset tarkoittavat mittarin validiteetin kannalta nyt ja jatkotutkimuksissa. Lisäksi mm. seuraavia kohtia voisi selkeyttää:

- Ensimmäisessä kappaleessa ei käsittääkseni pitäisi käyttää sanaa "vaikuttaa" (asetelma ei mahdollista kausaalisuhteiden arviointia)?
- Tekstissä voisi avata tarkemmin, että mitä tässä tarkoitetaan kohtelun tekoympäristöllä. Tämä selviää tietysti aiemmasta tekstistä, mutta lukijaa voisi muistuttaa.
- Tekstissä todetaan, että "Pitkittäistutkimus voisi tukea kausaaliiteetin selvittämisessä, mutta mahdollisia välillisiä tekijöitä tulisi silloin selvittää tarkemmin." Pitkittäistutkimus ei kuitenkaan itsessään mahdollista kausaaliiteetin selvittämistä (ja toisaalta poikittaistutkimuksesta voidaan joissain tapauksissa päätellä kausaalisuhteita). Ehkä tässä voisi viitata yleisemmin kausaalipäätelyn mahdollistaviin asetelmiin? Tai vaihtoehtoisesti antaa tarkemman ehdotuksen asetelmasta.
- "Sovittamamme mallit eivät sopineet aikaisempaan tutkimukseen pohjautuvan teorian mukaisesti, jolloin käsitteistöä olisi syytä tarkistaa uudelleen." Mitä tämä tarkoittaa? Voisi kirjoittaa auki.
- "Väkivalta voi tällä hetkellä olla niin monitulkinnallinen käsite mitattavaksi, että vastaajat tulkitsevat käsitettä hyvin vaihtelevasti." Tämä on varmasti totta ja tärkeä huomio, mutta millä perusteella vastaajat tulkitsevat käsitettä vaihtelevasti?

Yleisesti tutkimus vaikuttaa asiallisesti toteutetulta ja käsikirjoitus vaatii vain joitain tarkennuksia ja korjauksia.

Arvio 3: 27.1.2025

Tulokset raportoiva tutkimus on varsin kiinnostava ja arvokas lisä epäasiallisen ja väkivaltaisen kohtelun tutkimuskirjallisuuteen Suomessa ja kansainvälisesti. Tulosten raportoinnissa on kuitenkin useita puutteita ja rajoitteita, jotka vaikeuttavat tutkimustulosten ja niiden vaikuttavuuden tulkintaa.

Käsikirjoituksen liitteeksi tulisi laittaa kyselylomake vähintään faktorianalyysiin kuuluneiden kysymysten ja kysymyksenasetteluiden osalta. Vaikka tämä lomake on nyt jaossa OSF-palvelussa, lukukokemusta parantaisi huomattavasti, jos mittariston kysymykset olisivat arvioitavassa osana käsikirjoitusta tai sen välittömässä yhteydessä erillisenä liitteenä. Kysymysasettelu vaikuttaa faktoreihin ja muuttujien faktorianalyysiin, ja siksi käsikirjoituksessa tulisi olla nähtävissä ja arvioitavissa myös kysymisen tavat sekä käytössä olleet skaalat. Viite tällaiseen liitteeseen sopisi esimerkiksi luvun 2.4.1 päätteeksi eli kohtaan, jossa kuvataan mittarin kehitysvaiheita. Toinen ja suosittelemani vaihtoehto on raportoida kysymykset ja kysymyksenasettelut suoraan käsikirjoituksessa, osana EFA:n raportointia.

Ehdotan, että esirekisteröinnissä määritellyn ja käsikirjoituksessa raportoidun Hullin menetelmän ja CFI, CAF sekä RMSEA indeksien lisäksi kirjoittajat raportoisivat eksploratiivisesti myös EFA-tutkimuksessa useammin käytettyjen Minimum Average Partial (MAP) sekä erityisesti Parallel Analysis (PA) -testien tulokset, mitä tulee faktoreiden lukumäärän määrittämiseen kaikkien muuttujien mallissa sekä omakohtaisesti koetun kohtelun malleissa. Näin kahdesta syystä: ensinnäkin Hullin menetelmän mukaiset indeksit ehdottavat keskenään varsin erilaista faktoreiden lukumäärää kaikkien muuttujien mallissa, ja myös omakohtaisesti koetun kohtelun muuttujien mallissa oli eroja. Toiseksi, Hullin menetelmässä käytetyt indeksit (CFI, CAF ja RMSEA) ovat luonteeltaan enemmän teorialähtöisiä menetelmiä, koska ne perustuvat mallin yleiseen sopivuuteen suhteessa aineistoon ja tiettyyn rakenteeseen - toisin sanoen tiettyihin teoreettisiin odotuksiin. Esimerkiksi RMSEA soveltuu parhaiten analyysihin, joissa on jokin teoreettinen hypoteesi mallin rakenteesta.

Vaikka Hullin menetelmää on argumentoitu MAP- ja PA-testejä tarkemmaksi nimenomaan EFAn kontekstissa, jälkimmäiset menetelmät ovat luonteeltaan Hullin menetelmästä poikkeavia, sillä ne keskittyvät faktoreiden lukumäärän määrittämiseen aineiston rakenteesta käsin ilman ennako-oletuksia tietystä mallin sovituksesta. Näillä menetelmillä pyritään paljastamaan datan sisäinen rakenne ja minimoimaan satunnaisuuden vaikutus valittuun ratkaisumalliin. Ehdottaisin itse ainakin parallel analysis -testin käyttöä Hullin menetelmän rinnalla, sillä satunnaisdatasta johdetut vertailut estävät ylisovituksen ja tuovat faktoreiden lukumäärän määrittämiseen tilastollisen kontrollin, mikä tekee tästä testistä eksploratiivisemmän ja vähemmän riippuvaisen teoreettisista ennako-oletuksista.

Koskien edellä kuvattua ehdotusta täydentävien testien raportoinnista, ja myös koskien konfirmatorisen faktorianalyysin lukua, kirjoittajien tulisi tekstissään selvästi erotella, mitkä testit ja analyysit ovat suoraan esirekisteröinnin suunnitelmaan perustuvia ja mitkä ovat laajennuksia ja täydennyksiä esirekisteröidyssä suunnitelmassa kuvattuun malliin nähden. Tämä on olennaista, jotta lukija kykenee tunnistamaan ne kohdat, joissa tutkimus liikkuu konfirmatorisesta eksploratiivisemmän tutkimusotteen suuntaan. Tämä palaute koskee erityisesti konfirmatorisen faktorianalyysin lukua sekä

toisaalta myös Lopputulokset-lukua. Lopputulokset-luku tulisi siten jäsentää myös tavalla, joka selvästi kuvaa lukijalle, 1) mitkä analyysit perustuivat esirekisteröityyn suunnitelmaan, ja mitkä ovat näihin tuloksiin liittyvät päätelmät sekä 2) mitkä analyysit ovat laajennuksia suunnitelmasta, ja minkälaisia päätelmiä näihin liittyen voidaan tehdä.

Käsikirjoituksesta puuttuu EFA-tutkimuksen motivointi ja rationale. Tarkoitan, että käsikirjoituksessa olisi syytä kuvata ne syyt, miksi analyysi edellyttää eksploraatiivista analyysia, ja miten valitut indeksit/testit tukevat juuri analyysin eksploraatiivisuutta sekä näiden analyysien tavoitetta. Implisiittisesti esirekisteröidystä suunnitelmasta tämä käy ilmi, mutta kaipaen analyysin tulokset esittävän EFA-luvun alkuun menetelmän käyttöä tukevaa tiivistelmää, joka selkeästi perustelee myös valitut menetelmät, mitä tulee faktoreiden lukumäärän tunnistamiseen.

Taulukon 8 luentaa vaikeuttaa se, että näiden indeksien ja testien viitearvoja EFA-kontekstissa ei ole kerrottu; kirjoittajat toteavat Mallin 2B sopivuusarvojen olevan muita malleja paremmat, mutta käsikirjoituksessa tulisi raportoida myös, miten 2B-mallin ja muiden mallien arvot vertautuvat kirjallisuudessa määriteltäviin tyydyttäviin/hyviin indeksi- ja testiarvoihin EFAn viitekehityksessä. Toisin sanoen lukijan on vaikeaa arvioida käsikirjoituksen nykytilassa sitä, minkälaisen lähtökohdan Mallin 2B arvot antavat konfirmatorisella faktorianalyysille. Lukija tietää ainoastaan, että tämän mallivaihtoehdon sopivuusarvot ovat muita malleja paremmat.

Jos kirjoittajat raportoivat eksploraatiivisen faktorianalyysin mukaiset standardoidut lataukset Kuvion 3 mukaisesti, heidän tulisi myös keskustella tässä yhteydessä siitä, minkälaisen lähtökohdan instrumentti/item-kohtaiset lataukset antavat mittarin validoinnille ja konfirmatoriselle faktorianalyysille. Pelkästään latausten perusteella Kuviossa 3 kuvatun kahden faktorin mallin instrumenteista olisi perusteet poistaa molemmat ristiinlatautuneet muuttujat sekä "Fyysinen väkivalta, sisäinen", "Omaisuus-väkivalta, sisäinen" ja mahdollisesti myös "Omaisuus-väkivalta, ulkoinen" -muuttujat ennen konfirmatoriseen vaiheeseen siirtymistä – jos tähän on olemassa teorialähtöiset perusteet. En myöskään täysin ymmärrä, miten tähän malliin päädyttiin, sillä mallissa raportoidaan kolme muuttujaa, joiden faktorilataukset ovat alle 0.4, vaikka kirjoittajat ovat määritelleet: "Sisällyttämme muuttujan faktoriin, jos sen faktorilataus on yli .40". Faktorilatausten osalta kehoitan myös yhtenäistämään kirjoitusasua, kirjoittajat käyttävät välillä .XX-mallia ja välillä 0,xx-mallia, sama koskee myös indeksien raja-arvojen raportointia.

En itse pidä myöskään EFAn tulosten esittämistä Kuvion 3 -muotoisena parhaana ratkaisuna. Tällaista kuviota käytetään usein CFA-mallien raportoinnissa, mutta EFA-raportoinnissa se on harvinaista. Yksi tärkeä syy tähän on, että EFAssa on tärkeää osoittaa kaikkien muuttujien kaikki lataukset faktoreille, ja kuvio ei mahdollista tätä järkevällä tavalla. Pyydän siis kirjoittajia raportoimaan täyden faktorilataustaulukon 2B-mallille, jonka kanssa he etenevät CFA-vaiheeseen. Täyden faktorilataustaulukon tulisi sisältää myös uniqueness- tai communality-arvo jokaiselle instrumentille/itemille. Communality-arvo kuvaa, kuinka suuri osa yksittäisen itemin tai koko mittarin varianssista selittyy yhteisten tekijöiden avulla. Yksittäisen itemin osalta korkea communality-arvo osoittaa, että item jakaa merkittävän määrän varianssia muiden mittarin itemien kanssa ja on siten hyvin yhteneväinen yleisen faktorirakenteen kanssa. Koko mittarin osalta korkeat communality-arvot kautta linjan viittaavat siihen, että mittari

on kokonaisuudessaan tehokas ja kattaa hyvin sen tarkoittamat konstruktit, mikä puolestaan viestii mittarin luotettavuudesta ja validiteetista. Mittarinkehityksen tavoitteena on saavuttaa mahdollisimman korkeat communality-arvot sekä yksittäisen itemin että koko mittarin tasolla, jotta instrumentti vaikuttaisi toimivalta ja kykenevältä mittaamaan tarkoitettuja ilmiöitä.

CFA:n osalta (Kuvio 4) kirjoittajien tulisi raportoida instrumentti/item-tasolla faktorilatausten lisäksi myös residuaali/virhevarianssi sekä sisällyttää näihin liittyvä keskustelu osaksi analyysia. Toiseksi yleisten sopivuusmittojen χ^2 , RMSEA, CFI, SRMR lisäksi faktoreille tulisi laskea CR (composite reliability / komposiittireliabiliteetti) sekä AVE (Average Variance Explained / keskimääräinen selitetty varianssi) -arvot, ja suhteuttaa jälkimmäinen faktoreiden välisen korrelaation neliöön (Fornellin ja Larckerin kriteeri). CR-testissä hyväksyttävä arvo on 0.70 ja AVE-testissä 0.50 sekä tulos, jossa faktorikohtainen yli 0.50 AVE-arvo ylittää faktoreiden välisen korrelaation arvon. CR- ja AVE-testit ovat osa construct validity -testaamista ja AVE:n vertaaminen faktoreiden välisen korrelaation neliöön on osa instrumentin discriminant validity -testausta. Nämä testit tulisi raportoida, vaikka niitä ei olisi alkuperäisessä suunnitelmassa suoraan määriteltäkään tehtäväksi. Lisäksi mallissa (Kuvio 4) on useita heikosti latautuneita instrumentteja/itemeitä.

CFA-raportoinnin osalta luku olisi syytä selkeästi jakaa edellä mainitsemallani tavalla eli esirekisteröinnissä esitetyn suunnitelman mukaisiin ja sitä täydentäviin eksploratiivisiin kokeiluihin, joihin tulkintani mukaan osittain siirrytään käsikirjoituksen rivillä 423, vaihtoehtoisten mallien vertailun takia. Tällaisen jaon voisi tehdä hyvin jo otsikkotasolla, tämä auttaisi lukijaa. Vaihtoehtoisten mallien luenta CFA:n yhteydessä on periaatteellisella tasolla myös hieman horjuvaa, koska CFA:n tulisi teorialähtöisenä mallina olla riippumaton eksploratiivisesta otteesta ja perustua empiria- ja teorialähtöisen yhden mallin testaamisen. Toki jos kahteen kilpailevaan malliin on teoria- ja empirialähtöiset perusteet, tämä tulisi selvästi ilmaista ja perustella jo ennen CFA-analyysihin ryhtymistä. Lisäksi CFA-mallin hyväksymiskriteeristänne: "Hyväksyimme mallin, jos vähintään kaksi sopivuusindeksiä neljästä olivat raja-arvojen sisällä" tulee todeta, että julkaisuilla on käytössä mallien hyväksyntäkriteerien suhteen erilaisia malleja. Toisissa julkaisuissa mallit ovat hyväksyttäviä vain, jos kaikki sopivuusindeksit pysyvät raja-arvojen sisällä. Tähän liittyen olisi hyvä lisätä lähteitä, joissa todetaan 2/4 täyttyvien kriteerien oleva riittävää mallin hyväksynnän kannalta.

Riviltä 423 alkaen esitelty Malli 2, joka kuvataan graafisesti rivillä 446, on mielestäni hämmäntävä. Jos tässä eksploratiivisessa mallissa kaikki itemit katsotaan latautuviksi Yleinen-faktorille että Organisaatio-sisäinen/Organisaatio-ulkoinen-faktoreille, eikö tällöin ole kyseessä II-tason faktorianalyysi? Toisen tason faktorianalyysissä Organisaatio-sisäinen ja Organisaatio-yleinen katsottaisiin molemmat Yleinen-faktorille latautuviksi, kun taas kyselyn indikaattorit latautuisivat Malli 1:n mukaisesti joko Organisaatio-sisäinen tai Organisaatio-ulkoinen faktoreille. Kuitenkin huomautan, että Mallissa 2 on varsin matalia latauksia, jotka eivät tue tätä eksploratiivista mallia. Kutsuisin tätä mallia todellakin toisen asteen faktorianalyysiksi bifaktorirakenteen sijasta – siksikin, että jälkimmäisen voi helposti sekoittaa mittariston havaittuun kaksiulotteisuuteen. Kuitenkaan eksploratiivinen tutkimus ei sinällään tue tämän mallin validointia, joten kyseessä on todellakin eksploratiivinen konstruktio.

Tutkimustulokset-luvussa raportoitu H1-testaus, joka perustuu riveillä 434-442 korrelaatioihin kehitteillä olevan mittarin ja työuupumuksen välillä on sikäli ongelmallinen, että tällainen rakenneyhtälömalli on mahdollista toteuttaa vain validoiduilla instrumenteilla, eikä kehitteillä olevan mittarin validaatiota voi pitää tutkimuksen perusteella pitävänä. Siksi myöskään instrumentin korrelointi aiemman tutkimuksen validoiman mittarin kanssa ei ole täysin mielekästä. Minusta tässä yhteydessä pitäisi vähintään todeta tämä seikka ja samalla todeta, että suunnitelman mukaista H1-testausta ei tästä syystä voida toteuttaa. Kuitenkin jos kirjoittajat haluavat tällaisen testin toteuttaa, se lienee mahdollista eksploratiivisena – mutta en ole kovin varma sellaisenkaan mielekkyydestä validoimattomalla mittarilla. Yksi vaihtoehto olisi keskustella sen sijaan yleisellä tasolla rivien 348-349 raportoidun korrelaatiotaulun tuloksista pohtien yksittäisten muuttujien korrelaatioita työuupumuksen vastaavien muuttujien kanssa, ja todeta, ettei H1 tarkastelu ollut suunnitellulla tavalla mahdollista. Tietyssä mielessä tämä huomio koskee myös muita konfirmatoriselle faktorianalyysille asetettuja hypoteeseja. Yksittäisten itemien kohdalla invarianssitestaukset toki ovat silti mahdollisia.

Tutkimustulokset-luku on hyvä, mutta mahdollisen jatkotutkimuksen osalta hieman suppea. Jos tutkijat toteuttaisivat ja raportoisivat täyden EFA-tilin, he voisivat tässä luvussa keskustella myös eri muuttujien latauksista sekä ristiinlatauksista tarkemmin, sekä pohtia minkälaisia implikaatioita tällä voisi olla mittarin jatkokehitykselle tai keskeisten käsitteiden kuten väkivallan tulkinnoilla.

Näkinsin tuloksiin perustuen, että käsillä oleva tutkimus voisi haasteista huolimatta toimia ensimmäisenä vaiheena epäasiallisen ja väkivaltaisen kohtelun mittarinkehitykselle. Mittarin mahdollisen jatkokehityksen kannalta olisi kuitenkin ensiarvoisen tärkeää tunnistaa myös datalähtöisesti niitä syitä, jotka vaikeuttivat mittariston validointia. Tähän tarkoitukseen EFAn tulosten nykyistä laajempi analyysi voisi tuoda merkittävää lisäarvoa.

Jukka Vahlo