



**TURUN
YLIOPISTO**

Mercuri Urvalin kognitiivisen kykytestistön validiteetti

Psykologian
pro gradu -tutkielma

Laatija:
Katri Lavonius

Ohjaajat:
Apulaisprofessori Jukka Leppänen
Johtava psykologi, Mercuri Urval, Annukka Tervo

12.01.2023
Turku

Turun yliopiston laatujärjestelmän mukaisesti tämän julkaisun alkuperäisyys on tarkastettu
Turnitin OriginalityCheck -järjestelmällä.

Pro gradu -tutkielma

Oppiaine: Psykologia

Tekijä: Katri Lavonius

Otsikko: Mercuri Urvalin kognitiivisen kykytestistön validiteetti

Ohjaajat: Apulaisprofessori Jukka Leppänen ja johtava psykologi, Mercuri Urval, Annukka Tervo

Sivumäärä: 37 sivua

Päivämäärä: 12.01.2023

Mercuri Urval on suorahakuun, rekrytointeihin, soveltuvuusarviointeihin ja monenlaiseen HR-konsultointiin erikoistunut yritys, joka on kehittänyt omat menetelmänsä persoonallisuuden ja kognitiivisen kyvykkyyden mittaamiseen. Yrityksen omia menetelmiä käytetään heidän soveltuvuusarviointiprosessinsa osana. Lain yksityisyyden suojasta työelämässä 759/2004 mukaan henkilöarvioinneissa käytettyjen menetelmien ja niistä saatujen tuloksien tulee olla luotettavia. Arviointimenetelmien luotettavuutta ja pätevyyttä tulee siis tutkia, jotta soveltuvuusarviointiprosessi voidaan toteuttaa laadukkaasti ja eettisesti.

Kognitiivisen kyvykkyyden on todettu olevan vahvasti yhteydessä työssä suoriutumiseen ja sen onkin arvioitu olevan paras yksittäinen tulevan työssä suoriutumisen ennustaja. Tämän vuoksi soveltuvuusarvioinneissa monesti hyödynnetään kognitiivista kyvykkyyttä mittaavia menetelmiä. Tässä tutkielmassa tarkastellaan Mercuri Urvalin kognitiivisen kykytestistön (MU CT) käsitevaliditeettia rinnakkaisvaliditeetin kautta vertaamalla testistöstä saatuja tuloksia validiteetiltaan hyväksi todetun Intelligence Structure Test (IST) -kognitiivisen kykytestistön tuloksiin. Molempien testistöjen taustalla vallitsevat ajatukset g-faktorista sekä kognitiivisen kyvykkyyden hierarkkisesta rakenteesta, ja molemmat testistöt kartoittavat kognitiivista kyvykkyyttä kolmen alaskaalan (abstrakti/visuaalisspatiaalinen päättelykyky, numeerinen päättelykyky ja kielellinen päättelykyky) kautta. Tutkielmassa vertaillaan myös tässä tutkimuksessa havaittuja MU CT -testistöstä saatuja tuloksia aiemmissa tutkimuksissa havaittuihin tuloksiin. Vertailussa tarkastellaan yleistä suoriutumista testistöstä sekä testistön osa-alueiden välisiä korrelaatioita. Tämän tutkimuksen ja aiempien tutkimuksien tulosten vertailun tarkoituksena on kartoittaa Mercuri Urvalin testistöjen uusien normien ja uuden testialustan toimivuutta.

Tutkimuksen aineisto koostuu 187 yliopisto-opiskelijasta (83.4 % naisia). Aineisto kerättiin lähettämällä tutkimuskutsu Turun, Jyväskylän ja Itä-Suomen yliopistojen ainejärjestöjen sekä kaikkien suomalaisten psykologian oppiaineen ainejärjestöjen sähköpostilistoille. Tutkimukseen mukaan ilmoittautuneet opiskelijat saivat suorittaa IST- ja MU CT -testistön omilla tietokoneillaan itselleen sopivana ajankohtana ja sopivassa paikassa erillisinä päivinä testiviikon aikana. Testistöjen tulosten välistä yhteyttä arvioitiin Pearsonin korrelaatiokertoimilla ja testitulosten keskiarvojen eroja tarkasteltiin parittaisen t-testin avulla. MU CT -testistöstä tässä ja aiemmissa tutkimuksissa saatujen tulosten samankaltaisuutta arvioitiin yleisen suoriutumisen osalta riippumattomien otosten t-testillä ja testistön osa-alueiden välisten korrelaatioiden osalta tässä tutkimuksessa havaittujen korrelaatioiden luottamusvälejä tarkastelemalla.

Testistöjen yleisen kognitiivisen kyvykkyyden tulosten välillä havaittiin voimakas yhteys, mikä tukee MU CT -testistön rinnakkaisvaliditeettia ja täten myös käsitevaliditeettia. Myös numeeristen alaskaalojen välinen yhteys viittaa siihen, että alaskaalat kartoittavat samantyyppistä prosessointia. Muiden alaskaalojen väliset yhteydet eivät tue MU CT -testistön rinnakkaisvaliditeettia. Testistöjen numeerisen ja kielellisen päättelykyvyn keskiarvot eivät eronneet toisistaan, mutta yleisen kognitiivisen kyvykkyyden tulokset sekä abstraktin ja visuaalisspatiaalisen päättelykyvyn tulokset erosivat. MU CT -testistön osa-alueiden väliset korrelaatiot vastasivat pääosin aiemmissa tutkimuksissa havaittuja korrelaatioita, vaikka suoriutuminen tässä tutkimuksessa oli korkeampaa kuin aiemmissa tutkimuksissa. MU CT -testistön validiteetti sai siis tukea yleisen kognitiivisen kyvykkyyden mittarina, ja testistön uudet normit sekä testialusta vaikuttavat toimivan.

Avainsanat: älykyys, kognitiivinen kyvykyys, ongelmanratkaisutaidot, psykologiset testit, psykometriset testit, kognitiiviset testit, validiteetti

Sisällysluettelo

1	Johdanto	4
1.1	Yleinen kognitiivinen kyvykkyys	5
1.1.1	Yleisen kognitiivisen kyvykkyuden teoriataustaa	5
1.1.2	Yleisen kognitiivisen kyvykkyuden arviointi	8
1.1.3	Yleisen kognitiivisen kyvykkyuden ennustevaliditeetti henkilöarvioinnissa	9
1.2	Mittarin psykometriset ominaisuudet: validiteetti ja reliabiliteetti	10
1.3	Mercuri Urval Cognitive Tests (MU CT)	12
1.3.1	Osatestit	12
1.3.2	Testistölle aiemmin tehdyt tutkimukset	13
1.4	Intelligence Structure Test (IST)	14
1.4.1	Osatestit	15
1.4.2	Testistölle aiemmin tehdyt tutkimukset	15
1.5	Tutkimuksen tavoitteet ja tutkimuskysymykset	17
2	Menetelmät	19
2.1	Aineisto ja tutkimuksen kulku	19
2.2	Käytetyt mittarit	19
2.2.1	Mercuri Urval Cognitive Tests (MU CT)	19
2.2.2	Intelligence Structure Test (IST)	20
2.3	Tilastolliset analyysit	21
3	Tulokset	22
3.1	Kuvailevat tulokset	22
3.2	Korrelaatiot	23
3.3	Tulosjakaumien vertailu parittaisella t-testillä	25
3.4	MU CT -testistön tulokset suhteessa aiempaan tutkimukseen	27
4	Pohdinta	29
4.1	Keskeiset tulokset	29
4.1.1	Validiteetti	29
4.1.2	MU CT -testistön tulokset suhteessa aiempaan tutkimukseen	33
4.2	Tutkimuksen vahvuudet ja rajoitukset	34
4.3	Jatkotutkimusehdotukset	36
	Lähteet	38

1 Johdanto

Henkilöarvioinnin tarkoituksena on kartoittaa työntekijän ominaisuuksia ja kyvykkyyttä, joiden pohjalta voidaan arvioida hänen työkäyttäytymistään sekä soveltuvuuttaan tiettyyn työtehtävään tai työympäristöön (Psykologiliitto, 2019). Henkilöarvioinnissa pyritään siis ennustamaan työntekijän tulevaa työssä suoriutumista ja työssä oppimista (Schmidt & Hunter, 1998) suhteessa arvioinnin pohjana toimiviin työtehtävästä ja -ympäristöstä ammennettuihin kriteereihin (Psykologiliitto, 2019).

Hyvä henkilöarviointi koostuu aina usean luotettavan arviointimenetelmän monipuolisesta käytöstä (Psykologiliitto, 2019). Näin varmistetaan, että arvioinnista saadaan tarpeeksi kattava kuva työntekijän profiilista. Henkilön ominaisuuksia ja mahdollista työssä pärjäämistä voidaan kartoittaa esimerkiksi haastattelulla, työnäytteillä ja -simulaatioilla tai arvioimalla hänen saamia suosituksia (Schmidt & Hunter, 1998). Psykologisessa henkilöarvioinnissa henkilön tietojen, taitojen ja osaamisalueiden lisäksi arvioidaan myös hänen psyykkisiä ominaisuuksiaan ja persoonallisuuden piirteitään (Psykologiliitto, 2019). Tällöin henkilöarviointiprosessissa hyödynnetään psykologisia testimenetelmiä, kuten kognitiivisia kykyjä mittaavia testejä tai persoonallisuustestejä.

Laki yksityisyyden suojasta työelämässä (759/2004, 4 luku 13 §) määrittelee, että henkilöarvioinnissa käytettävien testausmenetelmien ja niistä saatujen tietojen tulee olla luotettavia. Jotta lain asettama vaatimus toteutuu, on erityisen tärkeää, että henkilöarvioinneissa käytettävien menetelmien ominaisuuksia ja luotettavuutta tutkitaan. Psykologisten menetelmien luotettavuutta voidaan tutkia tarkastelemalla menetelmien psykometrisia ominaisuuksia, kuten validiteettia (Kuuskorpi & Heikkinen, 2014). Validiteetti ilmaisee menetelmän kykyä mitata haluttua ilmiötä, esimerkiksi kognitiivisen testin kykyä mitata kognitiota.

Mercuri Urval on suorahakuun, rekrytointeihin, soveltuvuusarviointeihin ja monenlaiseen HR-konsultointiin erikoistunut maailmanlaajuinen yritys, jonka henkilöarviointiprosessi perustuu tieteelliseen näyttöön ja ajan myötä karttuneeseen kokemukseen (Mercuri Urval, n.d.). Mercuri Urval on kehittänyt omat soveltuvuusarviointiin sopivat testimenetelmät sekä kognitiivisten kykyjen että persoonallisuuden kartoittamiseen (Carlstedt ym., 2020). Nämä testimenetelmät pyrkivät mittaamaan hakijan persoonallisuutta ja kognitiota kattavasti suhteessa työympäristöön.

Tässä tutkielmassa kartoitetaan Mercuri Urvalin kognitiivisen testistön validiteettia vertaamalla testipatteristoa validiteetiltaan hyväksi todettuun (Hogrefe, 2017) Intelligence Structure Test -kognitiiviseen testistöön. Mercuri Urvalin kognitiivisen testistön validiteettia on arvioitu muilla tavoin aiemmin (Carlstedt ym., 2020), mutta tämänkaltaista toiseen testistöön vertaamista ei ole toteutettu ennen. Validiteettitarkastelun lisäksi tässä tutkimuksessa saatuja tuloksia verrataan aiemmissa tutkimuksissa saatuihin tuloksiin Mercuri Urvalin kognitiivisen testistön osalta. Mercuri Urval on siirtynyt uuteen testijärjestelmäalustaan sekä päivittänyt testistöjensä normiston viime vuonna, ja tulosten vertailun tavoitteena on kartoittaa uusien normien sekä uuden testijärjestelmän toimivuutta.

1.1 Yleinen kognitiivinen kyvykkyys

Rakkaalla lapsella on monta nimeä – siksi yleisestä kognitiivisesta kyvykkyydestäkin voidaan puhua monin eri termein. Yleisellä kognitiivisella kyvykkyydellä, älykkyydellä tai g-faktorilla viitataan samaan asiaan eli henkilön tiedonkäsittelytoimintoihin, päättelykykyyn, ongelmanratkaisukykyyn ja oppimiskykyyn (Latvala ym., 2020). Yleisellä kognitiivisella kyvykkyydellä on vahva geneettinen pohja, mutta sen kehittymiseen vaikuttavat myös monet muut tekijät, kuten lapsuuden kasvuympäristö, riittävä ravinnon saanti ja kouluympäristö.

Yleinen kognitiivinen kyvykkyys on suhteellisen pysyvä ominaisuus, sillä se pysyy hyvin samanlaisena 20-vuotiaasta 70-vuotiaaksi asti (Vuoksimaa, 2019). On todettu, että lapsuudessa mitattu älykkyys ennustaa kognitiivista kyvykkyyttä jopa yli 90-vuotiaaksi. Vanhuudessa yleinen kognitiivinen kyvykkyys usein heikkenee hieman, mutta sitä ennen heikentymistä esiintyy lähinnä yksittäisissä kognitiivisissa toiminnoissa, kuten prosessointinopeudessa. Normaaliin vanhenemiseen ei siis kuulu kokonaisvaltainen huomattava kognitiivinen taantuminen.

1.1.1 Yleisen kognitiivisen kyvykkyuden teoriataustaa

Ajatus yleisestä kognitiivisesta kyvykkyydestä perustuu psykologi ja tilastotieteilijä Charles Spearmanin 1900-luvun alussa esittämään g-faktoriin (Latvala ym., 2020). Spearman (1904) havaitsi, että kognition eri osa-alueet korreloivat toistensa kanssa positiivisesti, ja tulkitsti sen johtuvan siitä, että kognition eri osa-alueiden taustalla on yhteinen vaikuttava tekijä eli g-faktori. Toisin sanoen hän huomasi, että ne yksilöt, jotka menestyvät hyvin jossakin kognitiivisessa tehtävässä, menestyvät todennäköisesti hyvin myös muissa kognitiivisissa

tehtävissä. Myöhemmin onkin todettu, että g-faktori selittää miltei puolet yksilöiden välisistä eroista kognitiivisten testien suorituksissa (Latvala ym., 2020).

Toinen merkittävä älykkyyttä tutkinut psykologi on Raymond Cattell. Hänen (1967) mukaansa älykkyyttä voidaan tarkastella kahden ulottuvuuden eli joustavan ja kiteytyneen älykkyyden avulla. Kiteytyneellä älykkyydellä viitataan ajan myötä kasautuneeseen tietoon, kuten yleistietoon tai sanavarastoon, kun taas joustava älykkyyys tarkoittaa yleistä ongelmanratkaisua- ja päätöksentekokykyä, jota tarvitaan uusien asioiden käsittelyssä (Schneider & Newman, 2015). Joustava älykkyyys ei siis edellytä aiempaa tietopohjaa niin kuin kiteytynyt älykkyyys edellyttää. Tutkimuksissa on havaittu, että Spearmanin g-faktori ja Cattellin joustava älykkyyys korreloivat hyvin vahvasti keskenään (Carlsted ym., 2020). Joidenkin tutkimusten mukaan joustava älykkyyys ja g-faktori itseasiassa tarkoittavatkin samaa asiaa (ks. esim. Gustafsson, 1984). Täten yleistä kognitiivista kyvykkyyttä voidaan luonnehtia myös joustavan älykkyyden käsitteen avulla.

Kognitiivista kyvykkyyttä kuvataan usein myös hierarkkisten mallien avulla (Bryan & Mayer, 2020). Hierarkkiset mallit koostuvat kahdesta tai useammasta tasosta ja ne pyrkivät kuvaamaan g-faktorin rakennetta. Kolmen tason malleissa alimmalla eli ensimmäisellä tasolla sijaitsee spesifejä kapea-alaisia kognitiivisia kykyjä, joita ovat esimerkiksi visuaalinen muisti ja mentaalinen rotaatio. Keskimmäisellä eli toisella tasolla sijaitsee laaja-alaisia kognitiivisia kykyjä, jotka koostuvat ensimmäisen tason spesifeistä kognitiivisista kyvyistä. Esimerkiksi visuaalinen havaitseminen voidaan nähdä laaja-alaisena kognitiivisena kykynä, jonka osia visuaalinen muisti ja mentaalinen rotaatio ovat. Ylimmällä eli kolmannella tasolla sijaitsee g-faktori, joka taas koostuu toisen tason laaja-alaisista kognitiivisista kyvyistä. Visuaalinen havaitseminen voidaan nähdä siis g-faktorin yhtenä osa-alueena.

Yleinen kognitiivinen kyvykkyyys jaotellaan usein kolmitasoiseen malliin, jossa g-faktorin ajatellaan koostuvan kolmesta laaja-alaisesta kognitiivisesta kyvystä (Carlsted ym., 2020). Nämä laaja-alaiset kyvyt ovat kiteytynyt älykkyyys, joustava älykkyyys ja visuaalinen havaitseminen. Toinen tunnettu ja laajaa empiiristä tukea saanut älykkyyden hierarkkinen malli on Cattell-Horn-Carroll (CHC) -malli, joka sisältää peräti 16 laaja-alaista kognitiivista kykyä (Bryan & Mayer, 2020). CHC-mallin ensimmäinen ja toinen taso eli g-faktori ja mallin 16 laaja-alaista kognitiivista kykyä on esitelty Kuviossa 1.

Tyypillisesti toisen tason kognitiivisia kykyjä ei luokitella kategorioihin, mutta CHC-mallin laaja-alaiset kyvyt voidaan kuitenkin luokitella vielä kolmeen kategoriaan: geneerisiin

kykyihin, kiteytyneeseen tietouteen liittyviin kykyihin sekä sensoriseen ja motoriseen toimintaan liittyviin kykyihin (van Aken ym., 2019). Kategoriat on koodattu Kuvioon 1. niin, että kyvyt 1–6 ovat geneerisiä kykyjä, kyvyt 7–10 ovat kiteytyneeseen tietouteen liittyviä kykyjä ja kyvyt 11–16 ovat sensoriseen sekä motoriseen toimintaan liittyviä kykyjä.

Kuvio 1

Cattell-Horn-Carroll-mallin kolmas taso eli g-faktori ja toinen taso eli laaja-alaiset kognitiiviset kyvyt (van Aken ym., 2019). Laaja-alaiset kognitiiviset kyvyt on luokiteltu kolmeen kategoriaan. Kuva ei sisällä mallin ensimmäistä tasoa eli spesifejä kapea-alaisia kognitiivisia kykyjä.



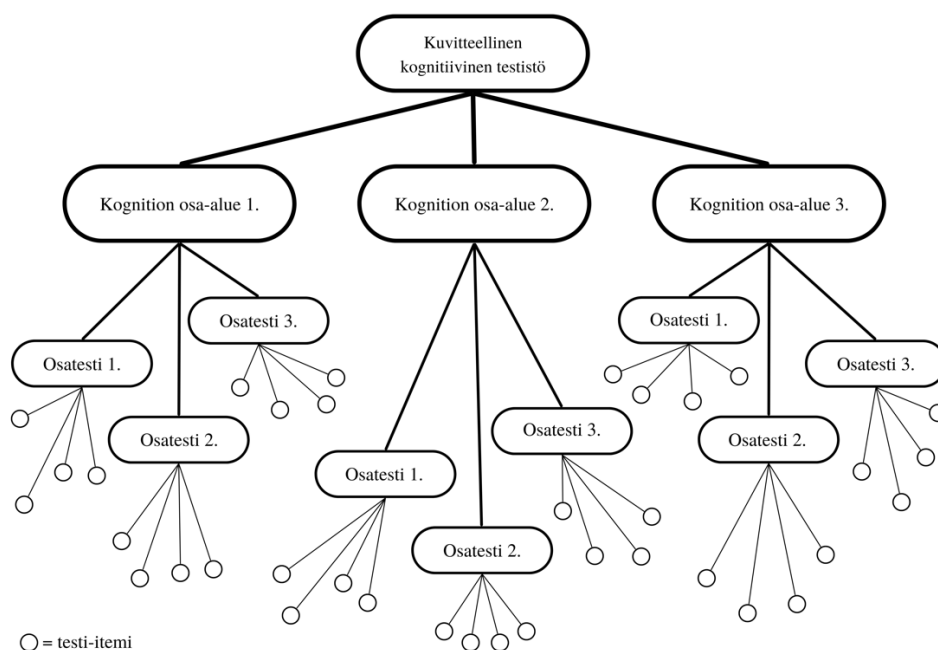
1.1.2 Yleisen kognitiivisen kyvykkyyden arviointi

Monien nykypäivänä käytössä olevien kognitiivisten kykytestistöjen taustalla vallitsee ajatukset yleisestä älykkyystekijästä ja g-faktorista (Latvala ym., 2020). Täten yleistä kognitiivista kyvykkyyttä mitataan tyypillisesti standardoiduilla kognitiivisilla testisarjoilla, jotka sisältävät tehtäviä kognition eri osa-alueilta. Tehtäviä voi olla esimerkiksi kielellisestä ymmärtämisestä, työmuistista, näönvaraisesta päättelykyvystä, prosessointinopeudesta tai matemaattisista taidoista. Kognitiivista arviointia hyödynnetään monilla sovellusaloilla, kuten tutkimustyössä, psykologien työkaluina ja soveltuvuusarvioinneissa.

Tyypillistä kognitiivisen testistön rakennetta voidaan havainnollistaa kuvitteellisen kognitiivisen testistön avulla (Kuvio 2.). Tämä kuvitteellinen kognitiivinen testistö koostuu kolmesta kognition osa-alueesta. Jokaista kognition osa-aluetta mitataan kolmella osatestillä, jotka ovat erilaisia, mutta jotka kaikki kuitenkin kuvastavat edustamansa kognition osa-alueen piirteitä. Jokainen osatesti koostuu vielä useasta yksittäisestä tehtävästä eli testi-itemistä. Esimerkissä yksi kognition osa-alue voisi olla numeerinen päättelykyky. Numeerinen päättelykyky muodostuu testistössä kolmesta osatestistä, joista yksi voisi olla laskutoimitustehtävä. Laskutoimitustehtävässä testattavalle esitetään useita päässäälaskuja, jotka hänen tulee ratkaista. Yksittäinen päässäälasku edustaa aina yhtä tehtävää eli testi-itemiä.

Kuvio 2

Kuvitteellisen kognitiivisen testistön rakenne



Yleistä kognitiivista kyvykkyyttä kartoittavilla testistöillä tyypillisesti arvioidaan testattavan kykyprofiilia sekä yleistä kognitiivista suoriutumista. Kykyprofiilista ilmenee testattavan suoriutuminen niillä kognition osa-alueilla, mitä kyseinen kognitiivinen testistö kartoittaa. Jokaisen kognitiivisen osa-alueen suoriutuminen muodostuu kyseisen osa-alueen osatestien pistemääristä. Yleisen kognitiivisen suoriutumisen tulos taas muodostuu kognition osa-alueiden tuloksista. Osa-alueiden tulosten ja yleisen kognitiivisen kyvykkyyden tuloksen muodostamisessa käytettävä laskukaava riippuu testistöstä ja se selviää aina kunkin testistön testimanuaalista sekä pisteytysohjeista.

1.1.3 Yleisen kognitiivisen kyvykkyyden ennustevaliditeetti henkilöarvioinnissa

Kognitiivisten kykytestistöjen käyttöä henkilöarvioinnissa perustellaan kognitiivisen kyvykkyyden vahvalla yhteydellä työssä suoriutumiseen. Lukuisten tutkimusten mukaan yleinen kognitiivinen kyvykkyys on yksi parhaimpia tulevan työssä suoriutumisen ennustajia (ks. esim. Salgado, 2017; Schmidt & Hunter, 1998). Jos otetaan huomioon arviointimenetelmien kustannustehokkuus, yleisen kognitiivisen kyvykkyyden kartoittaminen on paras tulevan työssä suoriutumisen ennustaja. Salgaden (2017) meta-analyyseja kokoavan katsauksen mukaan yleisen kognitiivisen kyvykkyyden ennustevaliditeetti eli korrelaatio työssä suoriutumisen kanssa on voimakas ($r = .62$). Yleinen kognitiivinen kyvykkyys selittää siis peräti 38 % työsuoriutumisen vaihtelusta.

Myös spesifit kognitiiviset kyvyt ennustavat työssä suoriutumista, mutta niiden ennustevaliditeetit jäävät pienemmiksi verrattuna yleisen kognitiivisen kyvykkyyden ennustevaliditeettiin (Salgado, 2017). Esimerkiksi numeerisen kyvykkyyden ennustevaliditeetti jää vain jonkin verran alemmaksi ($r = .52$), kun taas verbaalisen kyvykkyyden ennustevaliditeetti on jo huomattavasti matalampi ($r = .38$). Täten henkilöarvioinnissa kannattaa keskittyä yleisen kognitiivisen kyvykkyyden kartoittamiseen sen sijaan, että kartoitettaisiin vain yksittäisiä kognition osa-alueita.

Yleinen kognitiivinen kyvykkyys on myös positiivisesti yhteydessä oppimiseen, mikä näkyy työkontekstissa työhön liittyvänä oppimisena (Schmidt & Hunter, 2004). Koska ihmisten oppimiskyvyssä on yksilöllisiä eroja, sama määrä työkokemusta ei vielä takaa samaa määrää opittuja taitoja. Yleinen kognitiivinen kyvykkyys on oppimisen kautta positiivisesti yhteydessä siihen, miten työkokemuksen karttuminen eli työssä oppiminen muuntuu työssä suoriutumiseksi. Toisin sanoen, mitä kognitiivisesti kyvykkäämpi henkilö on, sitä

tehokkaammin hän todennäköisesti oppii kokemuksistaan työelämässä ja osaa käyttää oppimaansa myös hyödyksi.

Vaikka yleinen kognitiivinen kyvykkyys on jo itsessään hyvä työssä suoriutumisen ennustaja, ennustevaliditeettia voidaan lisätä yhdistämällä kognitiivisen kyvykkyuden kartoittaminen muihin arviointimenetelmiin (Salgado, 2017). Kahden selittävän muuttujan regressiossa kognitiivinen kyvykkyys ja tarkasti strukturoitu haastattelu yhdessä olivat parhaat työssä suoriutumisen ennustajat ($r = .77$), kun taas kolmen selittävän muuttujan regressiossa paras trio työssä suoriutumisen ennustamiseen oli kognitiivinen kyvykkyys, tarkasti strukturoitu haastattelu ja hakijan tunnollisuus ($r = .84$). Tarkasti strukturoitu haastattelu tässä yhteydessä tarkoittaa strukturoitua haastattelumenetelmää, jossa haastattelukysymykset on johdettu työnkuvasta ja sen kannalta keskeisistä elementeistä, jossa kaikilta haastateltavilta kysytään samat kysymykset, jossa kaikkien haastateltavien vastaukset analysoidaan tietyin kriteerein ja jossa haastattelija on koulutettu toteuttamaan tällainen työhaastattelu ja sen analysointi.

Yleinen kognitiivinen kyvykkyys ennustaa myös ammatillista tasoa eli sitä, kuinka vaativiin tehtäviin henkilö todennäköisesti päätyy (Schmidt & Hunter, 2004). Ylemmän ammatillisen tason työpaikoissa työntekijöiden kognitiivisen kyvykkyuden keskiarvo on korkeampi kuin alemman ammatillisen tason työpaikoissa. Pitkittäistutkimuksissa on myös havaittu, että jos henkilön työ on vaativampaa kuin hänen yleinen kognitiivinen kyvykkyytensä, hän todennäköisesti siirtyy myöhemmin vähemmän vaativiin tehtäviin. Sen sijaan, jos henkilön työ on vähemmän vaativaa kuin hänen kognitiivinen kyvykkyytensä, hän siirtyy todennäköisesti myöhemmin haastavampiin tehtäviin.

Tutkimustuloksien perusteella yleisen kognitiivisen kyvykkyuden kartoittaminen on mielekästä henkilöarviointiprosessin aikana. Tämän vuoksi monet soveltuvuusarviointeja tekevät yritykset, kuten Mercuri Urval, käyttävät hakijan kognitiivisia ominaisuuksia kartoittavia mittareita osana arviointiprosessiaan.

1.2 Mittarin psykometriset ominaisuudet: validiteetti ja reliabiliteetti

On tärkeää, että käytössä olevat mittarit, kuten kognitiiviset testistöt, ovat toimivia, laadukkaita, päteviä ja luotettavia (Kuuskorpi & Heikkinen, 2014). Mittarin toimivuutta voidaan arvioida tutkimalla mittarin psykometrisia ominaisuuksia. Kaksi tärkeintä mittarin psykometristä ominaisuutta ovat mittarin validiteetti ja reliabiliteetti.

Mittarin reliabiliteetti eli luotattavuus viittaa mittarin toistettavuuteen ja virheettömyyteen (Kuuskorpi & Heikkinen, 2014). Mittarin toistettavuus tarkoittaa sitä, saako sama testattava mittarista samanlaisen tuloksen eri mittauskerroilla. Esimerkiksi jos testattava tekee saman persoonallisuustestin kuukauden välein ja testin mukaan hänen persoonallisuutensa rakentuu joka testikerralla hyvin eri tavalla, persoonallisuustestin toistettavuus on heikko. Mittarin virheettömyys taas viittaa siihen, mittaako mittari haluttua ominaisuutta vai satunnaisvaihtelua. Satunnaisvaihtelua voi aiheuttaa esimerkiksi testattavan vireystila, motivaatio tai testiahdistus. Jos mittarin antama tulos koostuu suurelta osin satunnaisvaihtelusta, tuloksella ei ole juurikaan painoarvoa halutun ominaisuuden luonteen ymmärtämisessä.

Mittarin validiteetti eli pätevyys tarkoittaa sitä, mittaako mittari juuri sitä ilmiötä, mitä sen on tarkoituskin mitata (Valkeinen ym., 2014). Validiteetti tarkoittaa esimerkiksi sitä, mittaako masennustesti nimenomaan masennusta eikä vain ahdistuneisuutta, tai kartoittaako luetun ymmärtämisen testi nimenomaan sitä, kuinka hyvin henkilö on ymmärtänyt lukemansa, eikä vain lukunopeutta tai lukutarkkuutta. Ilman validiteettia mittari ja sen antamat tulokset ovat arvottomia, sillä tällöin mittarin tuottama informaatio ei ole relevanttia halutun ilmiön luonteen ymmärtämisen kannalta (Hiltunen, 2009).

Validiteetti voidaan jakaa moniin alalajeihin (Hiltunen, 2009), mutta tyypillisesti se jaetaan sisältö-, käsite- ja kriteerivaliditeettiin (Scroggins ym., 2008). Sisältövaliditeetti ilmaisee, missä määrin mittarin sisältö kattaa tutkittavan ilmiön sisällöt eli kuinka tarkka vastike mittari itsessään on tutkittavasta ilmiöstä (Valkeinen ym., 2014). Henkilöarvioinneissa sisältövaliditeetilla usein viitataan siihen, vastaako käytetty mittari työtehtävässä suoriutumista ja siinä tarvittavia tietoja sekä taitoja (Scroggins ym., 2008). Sen sijaan käsitevaliditeetti tarkoittaa sitä, missä määrin mittarin tulokset kattavat tutkittavan ilmiön konstruktion (Valkeinen ym., 2014). Käsitevaliditeettia voidaankin kutsua myös rakenne- tai konstruktiovaliditeetiksi (Hiltunen, 2009). Henkilöarvioinnin kontekstissa käsitevaliditeetti usein ilmaisee sitä, miten hyvin mittarin tulokset kattavat työssä suoriutumisen kannalta keskeisiä piirteitä (Scroggins ym., 2008).

Kriteerivaliditeetti kuvaa jonkin ennustajan ja kriteerin välistä yhteyttä (Scroggins ym., 2008). Ennustaja voi olla esimerkiksi henkilöarviointitilanteessa käytetyn mittarin pistemäärä ja kriteeri henkilön suoriutuminen työssä. Jos toinen henkilö on saanut mittarista matalat pisteet ja hän myös suoriutuu työtehtävissään välttävästi, kun taas toinen on saanut korkeat pisteet ja

hän suoriutuu työtehtävistään kiitettävästi, voidaan ajatella, että mittarin kriteerivaliditeetti on korkea. Kriteerivaliditeetti voidaan jakaa edelleen ennuste- ja rinnakkaisvaliditeettiin (Hiltunen, 2009). Ennustevaliditeetti ilmaisee, kuinka hyvin käytetyn mittarin tulokset ennustavat haluttua asiaa, esimerkiksi kuinka hyvin pääsykoe ennustaa opinnoissa menestymistä. Rinnakkaisvaliditeetti taas tarkoittaa sitä, kuinka yhteneväiset tutkittavan mittarin tulokset ovat toisen samaa asiaa kartoittavan mittarin tulosten kanssa.

Voidaan myös ajatella, että käsitevaliditeetti pitää sisällään kaikki validiteetin alalajit, sillä käsitevaliditeettia voidaan arvioida sisältö- ja kriteerivaliditeetin kautta (Hiltunen, 2009). Tässä tutkimuksessa kartoitetaan Mercuri Urvalin kognitiivisen testistön käsitevaliditeettia rinnakkaisvaliditeetin avulla niin, että testistöstä saatuja pistemääriä verrataan tunnettuun ja validiteetiltaan hyväksi todettuun (Hogrefe, 2017) Intelligence Structure Test -kognitiivisesta testistöstä saatuihin pistemääriin. Mikäli testistöistä saadut pistemäärät korreloivat keskenään voimakkaasti, voidaan olettaa, että testistöjen taustalla olevat konstruktiot korreloivat keskenään ja täten testistöt mittaavat samaa asiaa (Cronbach & Meehl, 1955). Tarkemmat rajat halutuille korrelaatioille on raportoitu johdannossa tutkimuskysymyksien ja hypoteesien kohdalla.

1.3 Mercuri Urval Cognitive Tests (MU CT)

Mercuri Urval on kehittänyt oman soveltuvuusarviointiin sopivan menetelmän kognitiivisen kyvykkyyden mittaamiseen ja sitä käytetään heidän soveltuvuusarviointiprosessinsa osana (Carlsted ym., 2020). Mercuri Urvalin kognitiivinen kykytestistö (MU CT) pohjautuu aiemmin esiteltyihin ajatuksiin g-faktorista ja älykkyyden hierarkiasta, mutta erityisesti Cattellin ajatuksiin joustavasta älykkyydestä. Testistö on kehitetty mittaamaan joustavaa älykkyyttä kolmen osatestin avulla, mitkä kartoittavat hakijan kielellistä, numeerista ja abstraktia päättelykykyä.

1.3.1 Osatestit

Kielellistä päättelykykyä kartoitetaan sanaparitestillä (Mercuri Urval, 2020). Sanaparitestissä testattavalle esitetään jokin sanapari sekä joukko irrallisia sanoja. Testattavan tulee päätellä irrallisten sanojen seasta vastaavanlainen sanapari kuin hänelle esitetty sanapari on. Esimerkiksi tutkittavalle voitaisiin esittää sanapari ”lääkäri–stetoskooppi” ja sanajoukko ”kapteeni, verstaas, kissa, sivellin, soutuvene, maalari”. Testattavan tulisi keksiä haetun

sanaparin olevan ”maalari–sivellin”, sillä tällöin molemmat sanaparit sisältävät sekä ammatin että siihen sopivan työvälineen.

Numeerista päättelyä mitataan numerosarjatestillä, jossa testattavalle esitetään tietyllä kaavalla muodostettu numerojono (Mercuri Urval, 2020). Testattavan tulee päätellä, mikä on numerojonon loogisin seuraava numero eli minkä kaavan mukaan numerot ovat asettuneet jonoon. Esimerkiksi numerojonon ”1, 9, 2, 8, 3, 7” seuraava loogisin numero on ”4”.

Abstraktia päättelykykyä kartoitetaan BIG-testillä eli visuaalisspatiaalisella matriisitestillä (Mercuri Urval, 2020). Testattavalle näytetään samaan aikaan kaksi kuviosarjaa sekä yksittäisiä irrallisia kuvioita. Testattavan tulee tunnistaa, sopivatko yksittäiset kuviot jompaankumpaan hänelle esitettyyn kuviosarjaan ja niissä esiintyviin lainalaisuuksiin. Vaihtoehtoina on, että yksittäinen kuvio sopii molempiin, vain toiseen tai ei kumpaankaan kuviosarjaan.

1.3.2 Testistölle aiemmin tehdyt tutkimukset

MU CT -testistön psykometrisiä ominaisuuksia on viimeisimmässä tutkimuksessa kartoitettu kymmenessä Euroopan maassa, jokaisessa yli tuhannen osallistujan otosjoukolla ($N = 20381$) (Carlstedt ym., 2020). Tutkimuksissa on mitattu testistön validiteettia eli pätevyyttä ja reliabiliteettia eli luotettavuutta. Tutkimusten mukaan testistön reliabiliteetti ja validiteetti ovat hyvät.

Korrelaatiotutkimuksissa on todettu, että MU CT -testistön osatestit kartoittavat kognition eri osa-alueita (Carlstedt ym., 2020). Osatestien välisistä korrelaatioista BIG-matriisien ja numerosarjatehtävän välinen korrelaatio on korkein ($r = .45$). Seuraavaksi korkein korrelaatio on BIG-matriisien ja sanapari-tehtävän välillä ($r = .40$) ja alhaisin korrelaatio on numerosarjatehtävän ja sanapari-tehtävän välillä ($r = .39$). Jos osatestien väliset korrelaatiot olisivat yli .80, osatestit mittaisivat pitkälti samaa asiaa, eivätkä ne silloin onnistuisi vangitsemaan kognition eri osia. Osatestien väliset kohtalaiset korrelaatiot kuitenkin viittaavat siihen, että osatestit pystyvät kartoittamaan kognition eri osa-alueiden lisäksi myös kaikkien osa-alueiden taustalla olevaa yleistä kognitiivista kyvykkyyttä.

Osatestien välisten korrelaatioiden lisäksi on tutkittu numerosarjatehtävän ja BIG-matriisitehtävän välistä yhteyttä testistön kartoittamaan yleiseen päättelykykyyn (Mercuri Urval, 2022). Tutkimuksien mukaan MU CT -testistön numerosarjatehtävä ($r = .84$) ja BIG-

tehtävä ($r = .87$) korreloivat erittäin voimakkaasti testistön yleisen päättelykyvyn tuloksen kanssa.

Myös eri maiden testiversioita on verrattu alkuperäiseen ruotsalaiseen testiversioon korrelaatioiden avulla (Carlstedt ym., 2020). Korrelaatiot vaihtelivat välillä .83–1.00 niin, että korkeimmat korrelaatiot saatiin ei-kielellisistä osatesteistä ja muutama alle .90 korrelaatio havaittiin kielellisestä osatestistä. Suomalaisen testiversion ja ruotsalaisen testiversion väliset korrelaatiot olivat erittäin voimakkaita jokaisen osatestin eli BIG-matriisien ($r = .97$), numerosarjojen ($r = .99$) ja sanaparien ($r = .83$) osalta. MU CT -osatestien tulokset ovat siis yhteneviä yli maantieteellisten rajojen.

Testidatalle tehdyn pääkomponenttianalyysin mukaan datasta on löydettävissä neljä komponenttia: yksi pääkomponentti, joka kuvastaa g-faktoria, sekä kolme muuta komponenttia, jotka kuvastavat kolmea osatestiä (Carlstedt ym., 2020). Kaikki neljä komponenttia yhdessä selittivät 89 % datan varianssista ja pääkomponentti yksinään selitti 50 %. Kolmen muun komponentin korrelaatio pääkomponentin kanssa oli vähintään .63 ja jokaisen komponentin korrelaatio oli yli .40 edustamansa osatestin kanssa. Pääkomponenttianalyysi tukee sitä, että MU CT -testistön rakenne vastaa teoriataustansa mukaista kognitiivisen kyvykkyyden rakennetta.

Osatestien sisäistä konsistenssia on tutkittu Cronbachin alfan avulla (Carlstedt ym., 2020). Kun osatesteille lasketaan yhdistetty Cronbachin alfa kaikkien kymmenen Euroopan maan testiversioista, osatesteistä korkein reliabiliteetti on numerosarjoilla ($\alpha = .82$). Sen sijaan BIG-matriisin ($\alpha = .71$) sekä sanaparien ($\alpha = .70$) reliabiliteetti on hieman matalampi. Suomalaisen testiversion sisäinen konsistenssi mukailee näitä tuloksia, sillä myös suomalaisessa testiversiossa korkein reliabiliteetti on numerosarjoilla ($\alpha = .81$), kun taas BIG-matriisin ($\alpha = .68$) ja sanaparien ($\alpha = .69$) reliabiliteetti jää hieman alemmaksi. Kaiken kaikkiaan MU CT -testistön reliabiliteetti on korkea.

1.4 Intelligence Structure Test (IST)

Intelligence Structure Test -testistö on kognitiivinen kykytestistö, jolla voidaan arvioida henkilön joustavaa kognitiivista päättelykykyä (Hogrefe, n.d. -a). Testistön taustalla vallitsee ajatukset yleisestä kognitiivisesta kyvykkyydestä sekä älykkyyden hierarkkisesta rakenteesta, ja se pohjautuu erityisesti johdannossa aiemmin esiteltyyn Cattell-Horn-Carroll -malliin.

Testistö on kehitetty mittaamaan kognitiivista kyvykkyyttä kolmen alaskaalan eli kielellisen,

numeerisen ja visuaalisspatiaalisen päättelykyvyn avulla. IST-testistön on julkaissut Hogrefe Psychologien Kustannus Oy, ja sen on kehittänyt Detlev Liepmann, André Beuducel, Burkhard Brocke sekä Rudolf Amthauer (Hogrefe, n.d. -b).

1.4.1 Osatestit

IST-testistössä jokaista päättelykyvyn alaskaalaa mitataan kolmella erilaisella osatestillä (Hogrefe, n.d. -a). Kielellistä päättelykykyä kartoitetaan lauseentäydennys-, sanapari- ja samankaltaisuustestillä. Lauseentäydennystestissä testattavalle näytetään lause, josta puuttuu yksi sana, ja hänen tulee valita viiden sanan joukosta lauseesta puuttuva sana.

Sanasuhdetestissä testattavalle näytetään sanapari, yksi yksittäinen sana ja viiden sanan sanajoukko. Testattavan tulee tunnistaa, miten esitetyn sanaparin sanat liittyvät toisiinsa, ja valita sanajoukosta sana, joka sopii yksittäisen sanan pariin soveltaen esitetystä sanaparista ilmenevää sääntöä. Samankaltaisuustestissä testattavan tulee valita kuuden sanan joukosta kaksi sanaa, joita yhdistää jokin yleinen käsite.

Numeerista päättelykykyä mitataan numerosarjoilla, laskutoimitusmerkkitehtävällä sekä perinteisillä laskutoimituksilla (Hogrefe, n.d. -a). Numerosarjoissa testattavan tulee keksiä näkemänsä numerosarjan loogisin seuraava numero. Laskutoimitusmerkkitehtävässä testattavalle näytetään laskutoimitus ilman laskutoimitusmerkkejä (plus-, miinus-, jako- ja kertomerkki) ja hänen tulee lisätä laskuun ne merkit, joilla laskutoimitus menee oikein. Perinteisissä laskutoimituksissa tutkittavalle esitetään laskutehtäviä numeerisessa muodossa ja hänen tulee ratkaista ne.

Visuaalisspatiaalista päättelykykyä kartoitetaan kuvion valinta -tehtävällä, kuutiotehtävällä ja matriisitehtävällä (Hogrefe, n.d. -a). Kuvion valinta -tehtävässä testattavalle näytetään joukko geometrisiä kuvioita sekä irrallisia kuviopaloja. Testattavan tulee tunnistaa, mikä kuvio paloista muodostuu. Kuutiotehtävässä testattava näkee esimerkkikuution kolme sivua ja hänen tulee tunnistaa, mikä annetuista vaihtoehdoista on esimerkkikuutio, kun sitä on käännetty. Matriisitehtävässä tutkittavan tulee valita annetuista vaihtoehdoista kuvio, joka sopii hänelle näytettyyn kuviosarjaan ja siinä esiintyvään kaavaan.

1.4.2 Testistölle aiemmin tehdyt tutkimukset

Intelligence Structure Test -testistön suomenkielisen version validiteettia ja reliabiliteettia kartoitettiin samalla, kun testistölle kerättiin uusi normiaineisto henkilöstöarvioinneissa

vuosina 2014–2016 (Hogrefe, 2017). Normiaineiston otos koostuu kokonaisuudessaan 665 tutkittavasta, mutta normiaineiston koko vaihtelee osatesteittäin, sillä vain 626 tutkittavaa suorittivat kaikki osatestit. Tutkimuksen mukaan IST-testistön suomenkielisen version reliabiliteetti ja validiteetti ovat hyvät.

IST-testistön validiteettia arvioitiin konfirmatorisella faktorianalyysillä (Hogrefe, 2017). Faktorianalyysin mukaan kolmen alaskaalan väliset korrelaatiot olivat korkeita niin, että korkein korrelaatio oli numeerisen päättelykyvyn ja visuaalisspatiaalisen päättelykyvyn alaskaalan välillä ($r = .69$) ja matalin korrelaatio kielellisen päättelykyvyn ja visuaalisspatiaalisen päättelykyvyn välillä ($r = .52$). Korkeista korrelaatioista huolimatta faktorianalyysin perusteella normiaineistoon sopi parhaiten kolmen faktorin malli yhden faktorin sijaan. Tämä tukee sitä, että IST-testistön rakenne koostuu kolmesta päättelykyvyn alaskaalasta.

Faktorianalyysistä käy myös ilmi, että osatestien korrelaatiot edustamiensa alaskaalojen kanssa vaihtelevat kohtalaisesta voimakkaaseen (Hogrefe, 2017). Kielellisen alaskaalan ja sen osatestien väliset korrelaatiot vaihtelevat välillä .56–.76, numeerisen alaskaalan ja sen osatestien väliset korrelaatiot vaihtelevat välillä .73–.78 ja visuaalisspatiaalisen alaskaalan ja sen osatestien väliset korrelaatiot vaihtelevat välillä .51–.62.

Lisäksi IST-testistön kielellisen alaskaalan validiteettia on tutkittu vertaamalla IST-testistön tuloksia AVO-9-testistön R3-testiin, joka arvioi sanasuhteita (Hogrefe, 2017). IST-testistön ja R3-testin sanasuhdetehtävät ovat erilaiset ja niiden välinen korrelaatio on kohtalainen ($r = .42$). R3-testi oli yhteydessä myös muihin IST-testistön kielellisiin osatesteihin ($r = .38$). Tämän mukaan IST-testistön sanasuhdetehtävällä ja R3-testillä arvioidaan saman tyyppistä kielellistä prosessointia, vaikka tehtävät ovatkin erilaisia.

IST-testistön suomalaisen soveltuvuusarvioaineiston sisäistä konsistenssia on tutkittu Cronbachin alfan avulla (Hogrefe, 2017). Testistön alaskaalojen sisäinen konsistenssi on korkea: korkein reliabiliteetti on numeerisella päättelykyvyllä ($\alpha = .95$), seuraavaksi korkein visuaalisspatiaalisella päättelykyvyllä ($\alpha = .90$) ja matalin kielellisellä päättelykyvyllä ($\alpha = .83$). Myös päättelykyvyn kokonaistuloksen reliabiliteetti on korkea ($\alpha = .95$). Numeerisen päättelykyvyn osatestien alfakerroin vaihtelee välillä .84–.92, visuaalisspatiaalisen päättelykyvyn osatestien alfakerroin vaihtelee välillä .74–.82 ja kielellisen päättelykyvyn osatestien alfakerroin vaihtelee välillä .61–.83.

1.5 Tutkimuksen tavoitteet ja tutkimuskysymykset

Tämän pro gradu -tutkielman päätavoitteena on tarkastella Mercuri Urvalin kognitiivisen testistön käsitevaliditeettia. MU CT -testistön käsitevaliditeettia tarkastellaan rinnakkaisvaliditeetin avulla vertaamalla testistöstä saatuja tuloksia validiteetiltaan hyväksi todetun (Hogrefe, 2017) Intelligence Structure Test -testistöstä saatuihin tuloksiin. Molempien testistöjen teoriatausta on samankaltainen ja molemmat testistöt mittaavat yleistä kognitiivista kyvykkyyttä kolmen alaskaalan eli numeerisen, kielellisen ja abstraktin tai visuaalisspatiaalisen päättelykyvyn avulla (Carlstedt ym., 2020; Hogrefe, n.d. -a). Molempien testistöjen tulokset myös standardoidaan samalle välille. Testistöistä saatuja tuloksia verrataan sekä korrelaatioin että parittaisen t-testin avulla.

Tutkielman toisena tavoitteena on tarkastella, vastaavatko tässä tutkimuksessa MU CT -testistöstä saatu data testistölle aiemmin tehdyissä tutkimuksissa raportoituja tuloksia keskimääräisen suoriutumisen ja testistön osa-alueiden välisten korrelaatioiden osalta. Tässä tutkimuksessa ja aiemmissa tutkimuksissa havaittuja tuloksia vertaillaan, koska Mercuri Urval on siirtynyt uuteen testijärjestelmälustaan ja päivittänyt omien menetelmiensä normiston viime vuonna. Tarkastelussa halutaan selvittää, poikkeavatko uudella normistolla ja testijärjestelmälustalla suoritettujen testien tulokset merkittävästi vanhoilla normeilla ja alustalla saaduista tuloksista.

Tutkimuskysymykset ja hypoteesit ovat:

1. Ovatko Mercuri Urvalin kognitiivisesta testistöstä saadut tulokset yhteydessä Intelligence Structure Test -kognitiivisesta testistöstä saatuihin tuloksiin? Kertooko testistöjen välinen yhteys rinnakkaisvaliditeetista?

Ensimmäisen hypoteesin mukaan MU CT -testistöstä saadut tulokset (abstraktin, numeerisen ja kielellisen päättelykyvyn sekä yleisen päättelykyvyn tulokset) korreloivat positiivisesti IST-testistöstä saatujen tulosten (visuaalisspatiaalisen, numeerisen ja kielellisen päättelykyvyn sekä yleisen päättelykyvyn tulokset) kanssa.

On kuitenkin huomioitavaa, että IST-testistön visuaalisspatiaalinen alaskaala painottaa enemmän visuaalisspatiaalista kyvykkyyttä kuin MU CT -testistön abstrakti alaskaala. Täten oletetaan, että IST-testistön matriisitehtävä vastaa lähemmin MU CT -testistön abstraktin päättelykyvyn tulosta eli BIG-tehtävästä saatua tulosta kuin IST-testistön visuaalisspatiaalisen päättelykyvyn kokonaistulos. Toisen hypoteesin mukaan oletetaan siis, että BIG-tehtävän ja

IST-testistön matriisitehtävän välillä on vahvempi korrelaatio kuin MU CT -testistön abstraktin päättelykyvyn ja IST-testistön visuaalisspatiaalisen päättelykyvyn välillä.

Kolmannen hypoteesin mukaan oletetaan, että yleisen päättelykyvyn, numeerisen päättelykyvyn, kielellisen päättelykyvyn sekä BIG-tehtävän ja IST-testistön matriisitehtävän välillä vallitsee vähintään voimakas korrelaatio, mikä kertoo testistöjen välisestä rinnakkaisvaliditeetista.

2. Eroavatko Mercuri Urvalin kognitiivisen testistön tulokset Intelligence Structure Test -testistön tuloksista parittaisella t-testillä arvioituna?

Hypoteesin mukaan testistöjen tulosjakaumien ajatellaan olevan päällekkäisiä eli testistöjen tuloksissa ei ole eroa.

3. Vastaavatko tässä tutkimuksessa Mercuri Urvalin kognitiivisesta testistöstä saadut tulokset testistölle aiemmissa tehdyissä tutkimuksissa raportoituja tuloksia keskimääräisen suoriutumisen ja testistön osa-alueiden välisten korrelaatioiden osalta?

Korrelaatioiden osalta hypoteesi on, että tässä tutkimuksessa havaitut korrelaatiot vastaavat aiemmin havaittuja korrelaatioita. Keskimääräisen suoriutumisen osalta hypoteesia ei aseteta.

2 Menetelmät

2.1 Aineisto ja tutkimuksen kulku

Poikkileikkausaineisto kerättiin vuoden 2022 helmi- ja maaliskuun aikana lähettämällä tutkimuskutsut Turun, Jyväskylän ja Itä-Suomen yliopistojen ainejärjestöjen sekä kaikkien suomalaisten psykologian oppiaineen ainejärjestöjen sähköpostilistoille. Tutkimuskutsuja lähetettiin siis yhteensä yli sadan ainejärjestön sähköpostilistalle. Tutkimukseen ilmoittautui 261 yliopisto-opiskelijaa, mutta kaikki heistä eivät suorittaneet tutkimusta loppuun. Tutkimuksen todellinen otos koostui 187 opiskelijasta, jotka suorittivat molemmat tutkimukseen kuuluvat kognitiiviset kykytestistöt. Tutkittavista enemmistö eli 83.4 % ($n = 156$) oli naisia ja tutkittavien keski-ikä oli 26.21 vuotta ($kh = 6.66$). Tutkimusdataa käsiteltiin GDPR:n mukaisesti. Tutkimukselle ei haettu tutkimuseettisen lautakunnan lausuntoa.

Opiskelijoille lähetetty tutkimuskutsu sisälsi linkin, jonka kautta halukkaat osallistujat pystyivät ilmoittautumaan tutkimukseen. Ilmoittautumisen yhteydessä tutkittavilta kerättiin henkilötietoina nimi, ikä, sukupuoli ja sähköposti. Tutkimukseen mukaan ilmoittautuneet opiskelijat saivat testiviikolla sähköpostiinsa personoidut testilinkit, joiden kautta he pystyivät suorittamaan tutkimuksessa käytetyt kognitiiviset testistöt erillisinä päivinä omalla tietokoneellaan testiviikon aikana. Testiviikkoja järjestettiin yhteensä kaksi. Tutkittavia oli ohjeistettu tekemään testistöt rauhallisessa tilassa, mikä mahdollistaa niihin keskittymisen, ja varaamaan kummankin testistön tekemiseen puolitoista tuntia aikaa, sekä tekemään IST-testistön ennen MU CT -testistöä. Ennen tilastollisten analyysien aloittamista tutkimusdatasta poistettiin henkilöiden nimet ja sähköpostit.

Tutkimukseen osallistuminen ja kognitiivisten kykytestistöjen tekeminen oli vapaaehtoista. Niiden tutkittavien kesken, jotka suorittivat molemmat kognitiiviset testistöt, arvottiin 100 euron lahjakortti S-ryhmään. Tutkittavat saivat myös halutessaan itselleen omat testituloksensa molemmista kognitiivisista testistöistä.

2.2 Käytetyt mittarit

2.2.1 Mercuri Urval Cognitive Tests (MU CT)

Mercuri Urvalin kognitiivisella testistöllä kartoitetaan henkilön joustavaa päättelykykyä (Carlsted ym., 2020). MU CT -testistöä hyödynnetään Mercuri Urvalin soveltuvuusarvioinneissa ja se pyrkii mittaamaan kognitiota kattavasti suhteessa

työympäristöön. Testistö koostuu kolmesta osatestistä, jotka mittaavat kielellistä, numeerista ja abstraktia päättelykykyä. Osatestien sisällöt on esitelty johdannossa. Kielellisen osatestin maksimi suoritus aika on kuusi minuuttia, numeerisen 20 minuuttia ja abstraktin päättelykyvyn 35 minuuttia. Testistöstä saa yksittäiset testitulokset jokaiselle osatestille eli päättelykyvyn osa-alueelle sekä yleisen päättelykyvyn kokonaistuloksen. Yleisen päättelykyvyn tulos muodostetaan numeerisen ja abstraktin päättelykyvyn tuloksista.

Testistön osatesteista saadut raakapistemäärät kertovat, kuinka monta tehtävää henkilöllä on mennyt osatestistä oikein (Mercuri Urval, 2020). Raakapistemäärät voivat siis vaihdella nollasta osatestin itemien lukumäärään. Kielellinen osatesti eli sanapari tehtävä koostuu 24 tehtävästä eli testi-itemistä, ja testistölle tehtyjen tutkimuksien mukaan kielellisestä osatestistä saatu keskimääräinen tulos raakapisteissä on 15.8 ($kh = 3.46$). Numeerinen osatesti eli numerosarjatehtävä koostuu 24 tehtävästä ja osatestin keskimääräinen tulos raakapisteissä on 13.1 ($kh = 4.45$). Abstraktin päättelyn osatestissä eli BIG-matriiseissa on 22 tehtävää ja osatestin keskimääräinen raakapistetulos on 10.7 ($kh = 3.48$).

Raakapistemääristä muodostetaan standardoidut testitulokset, jotka asettuvat välille 1–9 (Carlsted ym., 2020). Korkeammat testitulokset viittaavat suurempaan kognitiiviseen kyvykkyyteen. Tulokset lasketaan käyttäen Mercuri Urvalin henkilöarvioinneissa kerättyjä normeja. Mercuri Urvalin asiakaskunta koostuu pitkälti korkeakoulutetuista henkilöistä, joten normisto sopii myös tämän tutkimuksen otokselle. Testistölle tehtyjen aiempien tutkimuksien mukaan standardoitujen testitulosten jakauma noudattaa normaalijakaumaa, ja testistöstä tyypillisin saatu standardoitu tulos (sekä päättelykyvyn kokonaistulos että osatestien tulos) on viisi.

2.2.2 Intelligence Structure Test (IST)

Intelligence Structure Test -kykytestistö kartoittaa henkilön joustavaa kognitiivista päättelykykyä (Hogrefe, n.d. -a). Tutkimuksessa käytettiin Hogrefen tarjoamaa IST-testin internetversiota. Testistö jakautuu kolmeen alaskaalaan eli kielelliseen, numeeriseen ja visuaalisspatiaaliseen päättelykykyyn, ja jokaista alaskaalaa mitataan kolmella osatestillä. Jokainen osatesti koostuu 20 tehtävästä ja osatestien maksimi vastausaika vaihtelee kuudesta kymmeneen minuuttiin (Hogrefe, n.d. -b). Testistöstä on mahdollista laskea sekä päättelykyvyn kokonaistulos että yksittäiset testitulokset jokaiselle alaskaalalle (Hogrefe, n.d. -a). Päättelykyvyn kokonaistulos muodostetaan kaikkien testistön alaskaalojen tuloksista.

Testistön raakapistemääristä muodostetut standardoidut testitulokset asettuvat välille 1–9 (Hogrefe, n.d. -a). Korkeammat testitulokset viittaavat suurempaan kognitiiviseen kyvykkyyteen. Testistön tulosten laskemiseen käytettiin suomalaisia henkilöarvioinneissa kerättyjä normeja, jotka ovat tarkoitettu ylemmän korkea-asteen ja tutkijakoulutuksen käyneille testattaville. IST-testistön tarjoamat ylemmän korkea-asteen normit kuvastavat tutkimuksen otosta ja haluttua kohdepopulaatiota eli Mercuri Urvalin asiakaskuntaa paremmin kuin toiset tarjolla olevat normit, jotka on tarkoitettu alemman korkea-asteen ja sitä alempien kouluasteiden käyneille henkilöille.

2.3 Tilastolliset analyysit

Tilastolliset analyysit tehtiin lähtökohtaisesti IBM SPSS Statistics 28 -ohjelmalla (IBM Statistical Package for the Social Sciences). Tässä tutkimuksessa saatujen tulosten ja aiemmissa tutkimuksissa saatujen tulosten vertailuun käytettiin kuitenkin GraphPadin (n.d.) tarjoamaa t-testilaskuria. Näiden t-testien yhteydessä efektikoon estimaatin laskemiseen käytettiin Psychometrica-sivuston tarjoamaa efektikoon estimaatilaskuria numero viisi (Lenhard & Lenhard, 2016). Tutkimuksen muuttujien eli testitulosten jakaumat eivät ole normaalisti jakautuneita, mutta keskeiseen raja-arvolauseeseen nojaten tutkimuksessa käytettiin kuitenkin parametrisia testejä.

Mercuri Urvalin kognitiivisen testistön käsitevaliditeettia tutkittiin rinnakkaisvaliditeetin avulla tarkastelemalla Pearsonin korrelaatiokertoimia suhteessa Intelligence Structure Test – testistöön. Tässä tutkimuksessa alle .30 korrelaatiot mielletään heikoiksi, välille .30–.59 asettuvat korrelaatiot kohtalaisiksi, välille .60–.80 asettuvat korrelaatiot voimakkaiksi ja yli .80 korrelaatiot erittäin voimakkaiksi. Jotta korrelaation voidaan katsoa kertovan testistöjen välisestä rinnakkaisvaliditeetista, testistöjen tulosten välisten korrelaatioiden on oltava voimakkaita tai erittäin voimakkaita. Nämä korrelaatorajat mukailevat muissa rinnakkaisvaliditeettitutkimuksissa esiintyviä rajoja (ks. esim. Heineman ym., 2013; Rubio-Codina ym., 2016; Weber ym., 2018)

MU CT ja IST-testistöistä saatujen tulosten samankaltaisuutta tarkasteltiin parittaisella t-testillä. Lisäksi tämän tutkimuksen tuloksia verrattiin aiemmissa tutkimuksissa saatuihin tuloksiin MU CT -testistön osalta korrelaatiokertoimien luottamusvälejä tarkastelemalla sekä riippumattomien otosten t-testillä niiltä osin, kun se oli mahdollista.

3 Tulokset

3.1 Kuvailevat tulokset

Kuvailevat tunnusluvut on esitetty Taulukossa 1. Mercuri Urvalin kognitiivisen testistön tulokset koostuvat abstraktin päättelykyvyn eli BIG-matriisitehtävän (BIG_MUCT), numeerisen päättelykyvyn eli numerosarjatehtävän (NS_MUCT) ja kielellisen päättelykyvyn eli sanaparitehtävän (SP_MUCT) sekä yleisen päättelykyvyn (G_MUCT) arvoista.

Intelligence Structure Test -testistön tulokset koostuvat visuaalisspatiaalisen päättelykyvyn (VSP_IST), numeerisen päättelykyvyn (NP_IST) ja kielellisen päättelykyvyn (KP_IST) sekä yleisen päättelykyvyn (G_IST) arvoista. Taulukossa esitetyt testitulokset ovat standardoituja tuloksia.

Taulukko 1

Muuttujien havaitut vaihteluvälit, keskiarvot (ka), keskihajonnat (kh), mediaanit (md) ja moodit (mo), (n = 187)

Muuttujat	Havaittu vaihteluväli	ka	kh	md	mo
Ikä	19–53	26.21	6.66	24	21
BIG_MUCT	1–9	5.79	1.92	6	7
NS_MUCT	1–9	6.20	1.81	6	6
SP_MUCT	2–9	6.47	0.91	7	7
G_MUCT	1–9	5.91	1.79	6	5
VSP_IST	1–9	6.37	2.11	7	9
NP_IST	1–9	6.17	2.04	6	6
KP_IST	1–9	6.64	1.66	7	8
G_IST	1–9	6.70	1.90	7	8 9

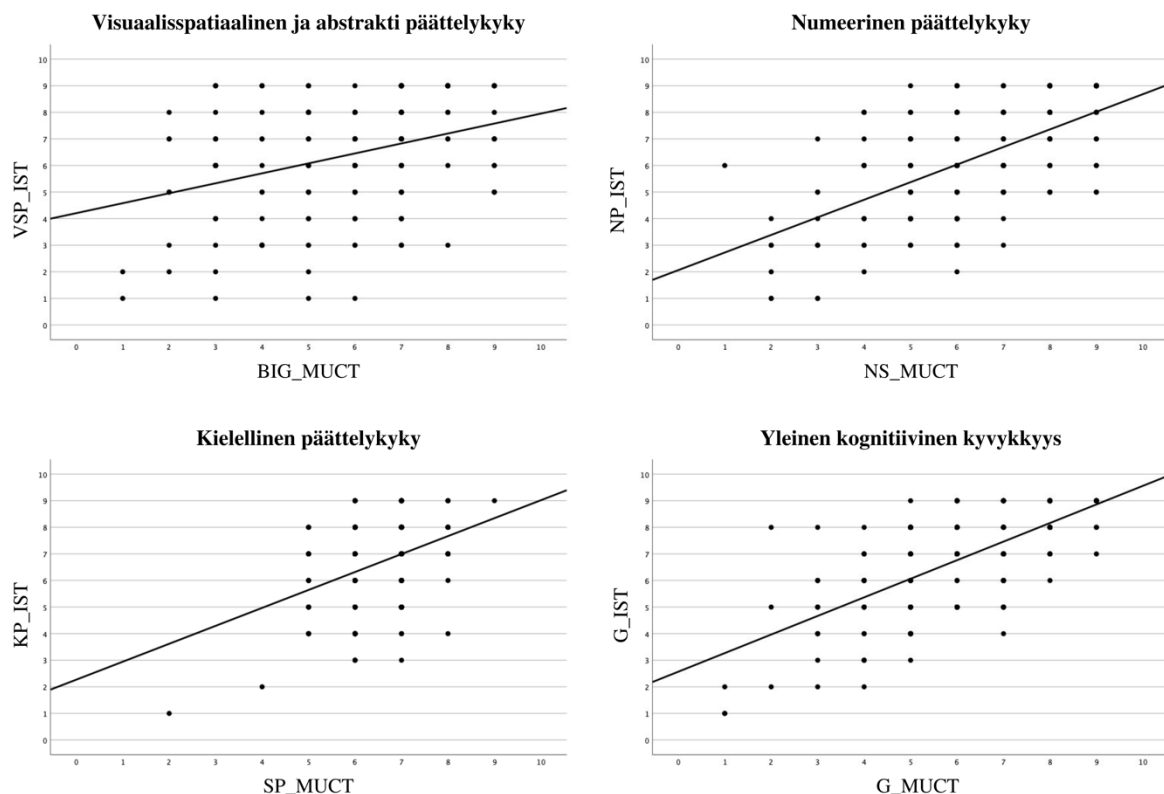
Huom. MUCT = Mercuri Urvalin kognitiivinen testistö, IST = Intelligence Structure Test -testistö, BIG = abstrakti päättelykyky, VSP = visuaalisspatiaalinen päättelykyky, NS/NP = numeerinen päättelykyky, SP/KP = kielellinen päättelykyky, G = yleinen kognitiivinen kyvykkyys

3.2 Korrelaatiot

Ennen korrelaatiokertoimien laskemista, Mercuri Urvalin kognitiivisen testistön ja Intelligence Structure Test -testistön välistä yhteyttä tarkasteltiin sirontakuvioiden avulla (Kuvio 3.). Sirontakuvioissa on näkyvissä myös muuttujien lineaarista yhteyttä kuvastava regressiosuora. Sirontakuvioissa on havaittavissa poikkeavia arvoja, mutta mitään arvoja ei kuitenkaan poistettu aineistosta. MU CT ja IST-testistön välistä yhteyttä tarkasteltiin Pearsonin korrelaatiokertoimen avulla. MU CT -testistön tulokset korreloivat positiivisesti IST-testistön tulosten kanssa. Testistöjen osatestien sekä yleisen päättelykyvyn tulosten väliset Pearsonin korrelaatiokertoimet on esitetty Taulukossa 2. Kaikki korrelaatiot olivat tilastollisesti merkitseviä ($p < .01$).

Kuvio 3

Mercuri Urvalin kognitiivisen testistön ja Intelligence Structure Test -testistön standardoitujen tulosten väliset sirontakuviot ja lineaarista yhteyttä kuvaavat regressiosuorat



Huom. MUCT = Mercuri Urvalin kognitiivinen testistö, IST = Intelligence Structure Test -testistö, BIG = abstrakti päättelykyky, VSP = visuaalisspatiaalinen päättelykyky, NS/NP = numeerinen päättelykyky, SP/KP = kielellinen päättelykyky, G = yleinen kognitiivinen kyvykkyys

Taulukko 2

Mercuri Urvalin kognitiivisen testistön ja Intelligence Structure Test -testistön standardoitujen tulosten väliset Pearsonin korrelaatiokertoimet, (n = 187)

Muuttujat	1.	2.	3.	4.	5.	6.	7.	8.
1. BIG_MUCT	–							
2. NS_MUCT	.38**	–						
3. SP_MUCT	.40**	.23**	–					
4. G_MUCT	.81**	.79**	.40**	–				
5. VSP_IST	.34**	.41**	.31**	.47**	–			
6. NP_IST	.36**	.59**	.32**	.57**	.46**	–		
7. KP_IST	.44**	.36**	.37**	.48**	.31**	.36**	–	
8. G_IST	.48**	.62**	.41**	.66**	.78**	.85**	.56**	–

Huom. MUCT = Mercuri Urvalin kognitiivinen testistö, IST = Intelligence Structure Test -testistö, BIG = abstrakti päättelykyky, VSP = visuaalisspatiaalinen päättelykyky, NS/NP = numeerinen päättelykyky, SP/KP = kielellinen päättelykyky, G = yleinen kognitiivinen kyvykkyys

** $p < .01$

MU CT ja IST-testistöjen kielelliset alaskaalat ($r = .37$) ja numeeriset alaskaalat ($r = .59$) korreloivat kohtalaisesti toistensa kanssa. Myös MU CT -testistön abstraktin päättelykyvyn alaskaala ja IST-testistön visuaalisspatiaalisen alaskaala korreloivat kohtalaisesti ($r = .34$). Testistöjen yleisen päättelykyvyn tulokset korreloivat voimakkaasti ($r = .66$).

MU CT -testistön yleisen kognitiivisen kyvykkyuden tulos sekä kaikki päättelykyvyn tulokset korreloivat IST-testistön tuloksista vahvimmin yleisen päättelykyvyn tuloksen kanssa. IST-testistön numeerisen päättelykyvyn tulos korreloi hieman voimakkaammin MU CT -testistön numeerisen päättelykyvyn kanssa kuin muiden testitulosten kanssa, mutta muut IST-testistön tulokset korreloivat myös kaikista vahvimmin MU CT -testistön tuloksista yleisen päättelykyvyn tuloksen kanssa.

Mercuri Urvalin kognitiivisen testistön alaskaalojen tulokset koostuvat yhdestä alaskaalaa kartoittavasta tehtävästä, kun taas Intelligence Structure Test -testistön jokainen alaskaala koostuu kolmesta tehtävästä. MU CT -testistön tulosten ja IST-testistön tehtävätulosten väliset Pearsonin korrelaatiokertoimet on esitetty Taulukossa 3. Kaikki korrelaatiot olivat tilastollisesti merkitseviä ($p < .05$).

Taulukko 3

Mercuri Urvalin kognitiivisen testistön standardoitujen tulosten ja Intelligence Structure Test -testistön standardoitujen tehtävätulosten väliset Pearsonin korrelaatiokertoimet, (n = 187)

IST-testistön osatestit	BIG_MUCT	NS_MUCT	SP_MUCT	G_MUCT
VSP_Kuvion_valinta	.32**	.28**	.23**	.37**
VSP_Kuutiot	.26**	.30**	.28**	.35**
VSP_Matriisit	.25**	.40**	.21**	.40**
NP_Laskutehtävät	.26**	.56**	.25**	.50**
NP_Numerosarjat	.35**	.51**	.31**	.50**
NP_Laskutoimitusmerkit	.29**	.46**	.22**	.46**
KP_Lauseentäydennys	.31**	.16*	.24**	.27**
KP_Sanasuhteet	.30**	.26**	.36**	.35**
KP_Samankaltaisuudet	.36**	.33**	.27**	.41**

Huom. MUCT = Mercuri Urvalin kognitiivinen testistö, IST = Intelligence Structure Test -testistö, BIG = abstrakti päättelykyky, VSP = visuaalisspatiaalinen päättelykyky, NS/NP = numeerinen päättelykyky, SP/KP = kielellinen päättelykyky, G = yleinen kognitiivinen kyvykkyys

** $p < .01$, * $p < .05$

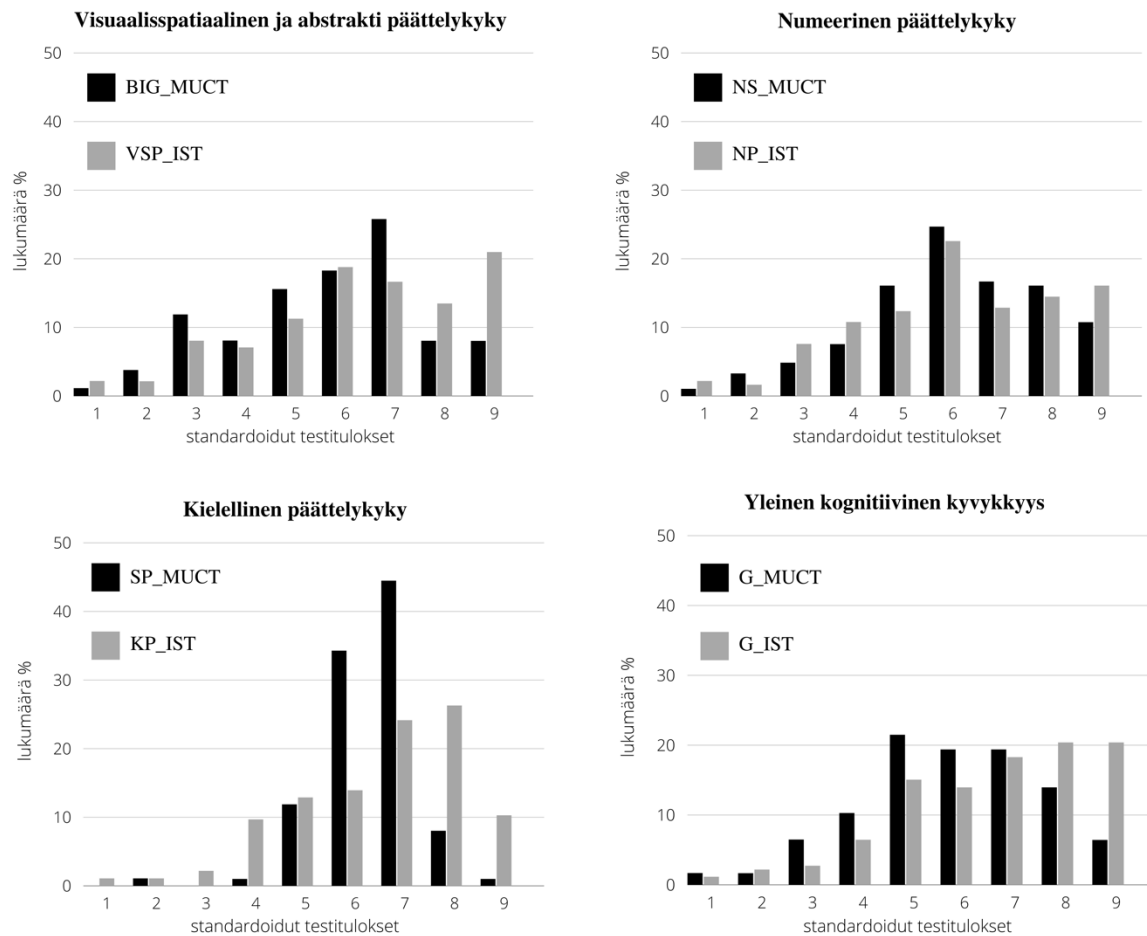
MU CT -testistön abstraktin päättelykyvyn tuloksen ja IST-testistön visuaalisspatiaalisen päättelykyvyn osatestien tulosten väliset korrelaatiot vaihtelivat välillä .25–.32. MU CT -testistön BIG-matriisitehtävän ja IST-testistön matriisitehtävän välinen korrelaatio oli heikko ($r = .25$). MU CT -testistön numeerisen päättelykyvyn tuloksen ja IST-testistön numeerisen päättelykyvyn osatestien tulosten väliset korrelaatiot vaihtelivat välillä .46–.56, kun taas MU CT -testistön kielellisen päättelykyvyn tuloksen ja IST-testistön kielellisen päättelykyvyn osatestien tulosten väliset korrelaatiot vaihtelivat välillä .24–.36. MU CT -testistön yleisen päättelykyvyn tuloksen ja IST-testistön kaikkien osatestien tulosten väliset korrelaatiot vaihtelivat välillä .27–.50. Korkeimmat korrelaatiot olivat MU CT -testistön yleisen päättelykyvyn ja IST-testistön numeerisen päättelykyvyn osatestien välillä.

3.3 Tulosjakaumien vertailu parittaisella t-testillä

Mercuri Urvalin kognitiivisen testistön ja Intelligence Structure Test -testistön standardoitujen tulosten jakaumat on esitetty Kuviossa 4. Pylväsdiagrammien x-akselilla sijaitsee testistöjen standardoitujen tulosten vaihteluväli, kun taas y-akselilla on tulosten lukumäärät prosentteina. MU CT -testistön tulokset on merkattu mustalla ja IST-testistön tulokset harmaalla. Tällä otoksella testistöistä saadut tulokset painottuvat tulosasteikon yläpäähän.

Kuvio 4

Mercuri Urvalin kognitiivisen testistön ja Intelligence Structure Test -testistön standardoitujen testitulosten jakaumat



Huom. MUCT = Mercuri Urvalin kognitiivinen testistö, IST = Intelligence Structure Test -testistö, BIG = abstrakti päättelykyky, VSP = visuaalisspatiaalinen päättelykyky, NS/NP = numeerinen päättelykyky, SP/KP = kielellinen päättelykyky, G = yleinen kognitiivinen kyvykkyys

MU CT -testistön ja IST-testistön tulostulosten samankaltaisuutta tarkasteltiin parittaisen t-testin avulla. MU CT-testistön numeerisen päättelykyvyn tuloksen ($ka = 6.20$, $kh = 1.81$) ja IST-testistön numeerisen päättelykyvyn tuloksen ($ka = 6.17$, $kh = 2.04$) välillä ei havaittu tilastollisesti merkitsevää eroa ($t(186) = .249$, $p = .803$, eron 95% $CI = [-0.22, 0.29]$, $d = .02$). Myöskään MU CT -testistön kielellisen päättelykyvyn tuloksen ($ka = 6.47$, $kh = .91$) ja IST-testistön kielellisen päättelykyvyn tuloksen ($ka = 6.64$, $kh = 1.66$) välillä ei havaittu tilastollisesti merkitsevää eroa ($t(186) = -1.448$, $p = .149$, eron 95% $CI = [-0.39, 0.06]$, $d = -.11$).

Sen sijaan MU CT -testistön abstraktin päättelykyvyn tuloksen ($ka = 5.79$, $kh = 1.92$) ja IST-testistön visuaalisspatiaalisen päättelykyvyn tuloksen ($ka = 6.37$, $kh = 2.11$) välillä havaittiin tilastollisesti merkitsevä ero ($t(186) = -3.470$, $p < .001$, eron 95% $CI = [-0.92, -0.25]$, $d = -.25$). Myös MU CT-testistön ($ka = 5.91$, $kh = 1.79$) ja IST-testistön ($ka = 6.70$, $kh = 1.90$) yleisen päättelykyvyn tulosten välillä havaittiin tilastollisesti merkitsevä ero ($t(186) = -7.078$, $p < .001$, eron 95% $CI = [-1.01, -0.57]$, $d = -.52$). Tutkittavat ovat siis saaneet keskiarvollisesti korkeampia pisteitä IST-testistön visuaalisspatiaalisesta päättelykyvystä sekä yleisestä kognitiivisesta kyvykkyydestä kuin MU CT -testistön vastaavista osa-alueista.

3.4 MU CT -testistön tulokset suhteessa aiempaan tutkimukseen

Mercuri Urvalin kognitiivisesta testistöstä tässä tutkimuksessa sekä aiemmissa tutkimuksissa havaitut abstraktin, numeerisen ja kielellisen päättelykyvyn standardoitujen tulosten keskiarvot ja keskihajonnat on esitetty Taulukossa 4. Taulukkoon on eritelty myös päättelykyvyn alaskaalojen raakapistemäärien keskiarvot ja keskihajonnat.

Taulukko 4

Mercuri Urvalin kognitiivisen testistön osatesteistä saatujen standardoitujen tulosten ja raakapistemäärien (raw) keskiarvot (ka) ja keskihajonnat (kh) tässä (n = 187) sekä aiemmissa tutkimuksissa (n = 20381)

Muuttujat	Tässä tutkimuksessa		Aiemmissa tutkimuksissa	
	ka	kh	ka	kh
BIG	5.79	1.92	5	-
NS	6.20	1.81	5	-
SP	6.47	0.91	5	-
BIG_raw	12.82	3.85	10.70	3.48
NS_raw	15.46	4.19	13.10	4.45
SP_raw	18.36	2.57	15.80	3.46

Huom. BIG = BIG-tehtävä/abstrakti päättelykyky, NS = numerosarjatehtävä/numeerinen päättelykyky, SP = sanaparitehtävä/kielellinen päättelykyky

Tässä tutkimuksessa saatuja raakapistemääriä verrattiin aiemmissa tutkimuksissa saatuihin raakapistemääriin riippumattomien otosten t-testillä. BIG-tehtävän raakapistetulosten välillä ($t(20566) = 8.284$, $p < .001$, eron 95% $CI = [1.62, 2.62]$, $d = .61$), numerosarjatehtävän raakapistetulosten välillä ($t(20566) = 7.223$, $p < .001$, eron 95% $CI = [1.72, 3.00]$, $d = .53$)

sekä sanapari tehtävän raakapistetulosten välillä ($t(20566) = 10.092, p < .001$, eron 95% $CI = [2.06, 3.06]$, $d = .74$) havaittiin tilastollisesti merkitsevä ero. MU CT -testistön osatesteista saadut raakapistemäärät ovat siis tässä tutkimuksessa keskiarvollisesti korkeampia kuin mitä testistölle tehdyissä aiemmissa tutkimuksissa on havaittu.

Taulukossa 5. on esitetty MU CT -testistön osatestien sekä päättelykyvyn väliset korrelaatiot tässä ja aiemmissa tutkimuksissa sekä tässä tutkimuksessa havaittujen korrelaatioiden 95 % luottamusvälit. Korrelaatiokertoimien samankaltaisuutta tutkittiin tarkastelemalla tässä tutkimuksessa havaittujen korrelaatioiden luottamusvälejä.

Taulukko 5

Mercuri Urvalin kognitiivisen testistön osatestien sekä yleisen päättelykyvyn väliset korrelaatiot (r) tässä ($n = 187$) ja aiemmissa ($n = 20381$) tutkimuksissa sekä tässä tutkimuksessa havaittujen korrelaatioiden 95 % luottamusvälit (95 % lv)

Muuttujaparit	Tässä tutkimuksessa		Aiemmissa tutkimuksissa
	r	95 % lv	r
BIG x NS	.38	.25–.50	.45
BIG x SP	.40	.25–.50	.40
NS x SP	.23	.09–.37	.39
BIG x G	.81	.76–.86	.87
NS x G	.79	.73–.84	.84

Huom. BIG = BIG-tehtävä/abstrakti päättelykyky, NS = numerosarjatehtävä/numeerinen päättelykyky, SP = sanapari tehtävä/kielellinen päättelykyky, G = yleinen kognitiivinen kyvykkyys

Aiemmissa tutkimuksissa havaitut BIG-tehtävän ja numerosarjatehtävän, BIG-tehtävän ja sanapari tehtävän sekä numerosarjatehtävän ja yleisen päättelykyvyn väliset korrelaatiot asettuvat tässä tutkimuksessa havaittujen vastaavien korrelaatioiden luottamusvälien sisään. Täten nämä aiemmissa tutkimuksissa ja tässä tutkimuksessa havaitut korrelaatiot eivät poikkea toisistaan. Sen sijaan aiemmissa tutkimuksissa havaitut korrelaatiot numerosarjatehtävän ja sanapari tehtävän sekä BIG-tehtävän ja yleisen päättelykyvyn tuloksen välillä eivät asetu tässä tutkimuksessa havaittujen vastaavien korrelaatioiden luottamusvälien sisään. Nämä aiemmissa tutkimuksissa ja tässä tutkimuksessa havaitut korrelaatiot poikkeavat toisistaan.

4 Pohdinta

Tämän pro gradu -tutkielman päätavoitteena oli tarkastella Mercuri Urvalin kognitiivisen kykytestistön käsitevaliditeettia. MU CT -testistön käsitevaliditeettia tarkasteltiin rinnakkaisvaliditeetin avulla vertaamalla testistöstä saatuja tuloksia validiteetiltaan hyväksi todetun (Hogrefe, 2017) Intelligence Structure Test -testistöstä saatuihin tuloksiin. Tulosten välistä yhteyttä ja samankaltaisuutta tarkasteltiin sekä korrelaatioin että parittaisen t-testin avulla.

Tutkielman toisena tavoitteena oli kartoittaa Mercuri Urvalin kognitiivisen testistön uusien normien ja uuden testijärjestelmäalustan toimivuutta. Testistöstä saatuja tuloksia tässä tutkimuksessa vertailtiin aiemmissa tutkimuksissa saatuihin tuloksiin. Vertailun tavoitteena oli selvittää, poikkeavatko uudella normistolla ja testijärjestelmäalustalla suoritettujen testien tulokset merkittävästi vanhoilla normeilla ja alustalla saaduista testituloksista.

Tämä tutkielma tuotti lisätietoa Mercuri Urvalin kognitiivisen testistön validiteetista. Testistöä käytetään Mercuri Urvalin henkilöarviointiprosessin osana, ja tutkielma vastaa Lain yksityisyyden suojasta työelämässä (759/2004, 4 luku 13 §) asettamaan vaatimukseen siitä, että henkilöarviointiprosesseissa käytettyjen testausmenetelmien ja niistä saatavien tuloksien tulee olla luotettavia. Luotettavien arviointimenetelmien käyttö lisää henkilöarviointiprosessin eettisyyttä ja mielekkyyttä.

4.1 Keskeiset tulokset

4.1.1 Validiteetti

Mercuri Urvalin kognitiivisen testistön käsitevaliditeettia tutkittiin rinnakkaisvaliditeetin avulla suhteessa Intelligence Structure Test -testistöön. Tässä tutkimuksessa voimakkaiden ja erittäin voimakkaiden korrelaatioiden ajatellaan kertovan mittarien välisestä rinnakkaisvaliditeetista eli siitä, että mittarit kartoittavat samaa asiaa. Korrelaatioiden lisäksi testistöjen tulosjakaumien samankaltaisuutta tarkasteltiin parittaisen t-testin avulla. Tämän tutkimuksen tulokset tukevat osittain, mutteivat yksiselitteisesti, testistöjen välistä rinnakkaisvaliditeettia ja sitä kautta Mercuri Urvalin kognitiivisen testistön käsitevaliditeettia.

Korrelaatioiden suhteen ensimmäinen hypoteesi sai tukea, sillä testistöjen tulokset korreloivat positiivisesti keskenään. Sen sijaan toinen hypoteesi ei saanut tukea, sillä BIG-tehtävän ja IST-testistön matriisitehtävän välinen korrelaatio oli heikko, kun taas BIG-tehtävän ja IST-

testistön visuaalisspatiaalisen päättelykyvyn tuloksen välinen korrelaatio oli kohtalainen. Kolmas hypoteesi sai osittain tukea, sillä testistöjen yleisen päättelykyvyn tulokset korreloivat keskenään voimakkaasti, mutta testistöjen päättelykyvyn osa-alueiden väliset korrelaatiot olivat kohtalaisia. Täten ainoastaan yleisen päättelykyvyn tulosten välinen korrelaatio vastaa rinnakkaisvaliditeetin edellyttämää yhteyden voimakkuutta.

Testistöjen osa-alueiden väliset kohtalaiset korrelaatiot voivat tarkoittaa sitä, että osa-alueet eivät mitatakaan samanlaista kognitiivista prosessointia. Tämä voi johtua esimerkiksi siitä, että jommankumman testistön osatestit eivät olekaan täysin valideja. Toisaalta testistöt ovat aiemmissa validiteettitarkasteluissa vaikuttaneet päteviltä (Carlstedt ym., 2020; Hogrefe, 2017) ja kaikkiin mittauksiin sisältyy aina jonkin verran mittausvirhettä, minkä vuoksi myös samaa ilmiötä mittaavien mittareiden välinen korrelaatio voi jäädä kohtalaiseksi (Valkeinen ym., 2014). Kohtalaiset korrelaatiot eivät siis automaattisesti merkitse sitä, että mittarit mittaisivat eri asiaa, vaan korrelaation voimakkuuteen voi vaikuttaa muutkin tekijät. Tämän tutkimuksen korrelaatioihin mahdollisesti vaikuttavia tekijöitä pohditaan jäljempänä. On myös huomioitavaa, että testistöjen numeeristen alaskaalojen välinen kohtalainen korrelaatio ($r = .59$) jäi vain yhden mittayksikön päähän voimakkaasta korrelaatiosta, jolloin senkin olisi katsottu kertovan testistöjen välisestä rinnakkaisvaliditeetista.

Testistöjen tulosjakaumien vertailun suhteen hypoteesina oli, että testistöistä saaduissa tuloksissa ei ole eroja. Hypoteesi sai vain osittain tukea, sillä t-testin mukaan numeeriset ja kielelliset alaskaalat eivät eronneet toisistaan, mutta yleisen päättelykyvyn tulos sekä abstrakti ja visuaalisspatiaalinen alaskaala erosivat. Abstraktin ja visuaalisspatiaalisen päättelykyvyn sekä yleisen päättelykyvyn tulosjakaumien erilaisuus näkyy myös moodeja tarkastellessa. BIG-tehtävästä yleisin saatu tulos on seitsemän, kun taas IST-testistön visuaalisspatiaalisesta alaskaalasta yleisin saatu tulos on yhdeksän. MU CT -testistöstä saatu tyypillisin yleisen päättelykyvyn tulos on viisi, kun taas IST-testistöstä saadut tyypillisimmät yleisen päättelykyvyn tulokset ovat kahdeksan ja yhdeksän. Testattavat ovat siis saaneet keskiarvollisesti korkeampia tuloksia IST-testistöstä kuin MU CT -testistöstä.

Erot testistöjen tuloksissa voivat selittyä monella eri tavalla. Esimerkiksi erot yleisen päättelykyvyn tuloksissa selittyvät osittain eroilla abstraktin ja visuaalisspatiaalisen päättelykyvyn tuloksissa. IST-testistön yleisen päättelykyvyn tulos muodostetaan kaikista päättelykyvyn alaskaaloista ja MU CT -testistön yleisen päättelykyvyn tulos muodostetaan abstraktin ja numeerisen päättelykyvyn alaskaaloista. Keskiarvollisesti korkeammat

visuaalisspatiaalisen päättelykyvyn tulokset nostattavat myös IST-testistön yleisen päättelykyvyn tulosta keskiarvollisesti korkeammaksi, sillä numeerisessa ja kielellisessä suoriutumisessa ei havaittu eroja testistöjen välillä. Tämän lisäksi matalammat abstraktin päättelykyvyn tulokset painavat enemmän MU CT -testistön yleisen päättelykyvyn tuloksessa, koska yleisen päättelykyvyn tuloksen muodostamisessa ei oteta huomioon testistön kielellistä tehtävää.

On myös huomioitava, että testistöjen abstraktin päättelykyvyn ja visuaalisspatiaalisen päättelykyvyn alaskaalat poikkeavat hieman toisistaan. IST-testistössä painotetaan enemmän visuaalista päättelykykyä, sillä alaskaalan osatesteissä vaaditaan myös sellaisia visuaalisspatiaalisia kykyjä, joita BIG-matriisitehtävässä ei vaadita. Esimerkiksi IST-testistön kuutiotehtävä vaatii spatiaalista rotaatiota ja kuvion valinta -tehtävä kappaleiden muodostamista. Sen sijaan matriisitehtävissä painottuu enemmän matriisien lainalaisuuksien ja sääntöjen ymmärtäminen. Tämän vuoksi myös oletin, että testistöjen matriisitehtävien välillä olisi voimakkaampi yhteys kuin edustamiensa päättelykyvyn alaskaalojen välillä. Painotusero saattaa osittain selittää testattavien eritasoisen suoriutumisen näissä päättelykyvyn alaskaaloissa.

BIG-matriisitehtävä on myös monimutkaisempi kuin IST-testistön matriisitehtävä. IST-testistön matriisitehtävässä testattavalle näytetään matriisi, josta puuttuu yksi osa, sekä erilaisia kuvioita, joista hänen tulee valita matriisista puuttuva osa. Testattava siis tietää, että aina yksi näytetyistä kuvioista sopii matriisiin. Sen sijaan BIG-tehtävässä testattavalle näytetään yhtä aikaa kaksi matriisia, joista ei puutu osia, ja kolme yksittäistä kuviota. Testattavan tulee nimetä jokaisen kuvion kohdalla, sopiiko se jommassakummassa tai molemmissa matriiseissa esiintyviin lainalaisuuksiin, vai eikö kuvio vastaa kummassakaan matriisisissa esiintyviä sääntöjä. BIG-matriisitehtävä on todennäköisesti monimutkaisemman luonteensa vuoksi vaikeampi, mikä voi selittää siitä saatuja matalampia pistemääriä sekä matriisitehtävien välistä heikkoa korrelaatiota.

IST-testistö mittaa jokaista päättelykyvyn osa-aluetta kolmella osatestillä, jotka ovat keskenään erilaisia. Sen sijaan MU CT -testistö mittaa päättelykyvyn osa-alueita vain yhdellä osatestillä. IST-testistössä testattava voi siis kompensoida yhden osatestin tulosta toisilla osatesteillä, jolloin päättelykyvyn osa-alueen kokonaistulos voi parantua, vaikka yksi osatesti menisikin huonommin. Erilaisten osatestien tekeminen voi myös motivoida testattavaa eri tavalla. Jos testattava esimerkiksi kokee jonkin osatestin ikäväksi tai hän ei ymmärrä osatestin

ohjeistusta, minkä vuoksi hän ei anna osatestissä parastaan, hän voi kuitenkin kokea toisen osatestin mielekkääksi tai ymmärrettävämmäksi, jolloin hän saattaa panostaa sen tekemiseen enemmän.

Kahden kognitiivisen kykytestistön suorittaminen on kognitiivisesti kuormittavaa, mikä voi johtaa niin sanottuun testiväsymykseen tai kognitiiviseen väsymykseen. Väsymiseffektin on todettu olevan yhteydessä muun muassa kognitiivisen suoritustason laskuun ja tylsistymisen tunteeseen tehtäviä tehdessä (ks. esim. American Psychological Association, n.d.; Van Cutsem ym., 2022). Tutkittavia ohjeistettiin tekemään kognitiiviset testistöt erillisinä päivinä ja IST-testistö ennen MU CT -testistöä. Täten IST-testistön suorittamisesta aiheutunut kuormitus on voinut vaikuttaa tutkittavan suoriutumiseen MU CT -testistössä – varsinkin, jos tutkittava on ohjeistuksesta huolimatta suorittanut testistöt samana päivänä. Väsymiseffekti saattaa osittain selittää, miksi testattavat ovat saaneet keskiarvollisesti korkeampia pisteitä IST-testistön visuaalisspatiaalisesta alaskaalasta ja yleisestä kognitiivisesta kyvykkyydestä kuin MU CT -testistön vastaavista osa-alueista.

Erot testistöjen välillä voivat johtua myös siitä, että testattavat ovat tehneet testistöt erilaisissa olosuhteissa. Testattavat suorittivat testistöt itsenäisesti testilinkkien kautta omilla tietokoneillaan haluamanaan ajankohtana ja haluamassaan tilassa. Testitilanteiden kontrolloimattomuuden vuoksi testitilanteiden välillä voi siis olla suuriakin eroja. Testattava on esimerkiksi saattanut tehdä IST-testistön aamupäivällä ollessaan virkeä ja MU CT -testistön myöhään illalla väsyneenä. Liian erilaisten testitilanteiden kuitenkin luulisi aiheuttavan systemaattista eroa testattavan kokonaissuoriutumisessa eli jokaisessa alaskaalassa, mutta tällaista systemaattista eroavaisuutta ei esiintynyt.

Yhteenvedona voidaan siis todeta, että Mercuri Urvalin kognitiivisen kykytestistön validiteetti sai tukea yleisen kognitiivisen kyvykkyyden kartoittamisen osalta. Myös testistöjen numeeristen alaskaalojen lähes voimakas yhteys viittaa siihen, että alaskaalat kartoittavat samantyyppistä numeerista prosessointia. Abstraktin ja visuaalisspatiaalisen päättelykyvyn välinen kohtalainen korrelaatio sekä kielellisen päättelykyvyn alaskaalojen välinen kohtalainen korrelaatio voidaan tulkita kahdella tavalla. Kohtalainen korrelaatio voidaan nähdä todisteena siitä, että alaskaalat vangitsevat samantyyppistä prosessointia, mutta esimerkiksi mittausvirheen tai alaskaalojen rakenteellisten eroavaisuuksien takia yhteys jää kohtalaiseksi. Toisaalta kohtalaisten korrelaatioiden voidaan ajatella kertovan alaskaalojen

taustalla vallitsevasta g-faktorista. Tarkempien johtopäätöksiä tekeminen vaatii lisätutkimusta.

4.1.2 MU CT -testistön tulokset suhteessa aiempaan tutkimukseen

Tässä tutkimuksessa havaittuja MU CT -testistön tuloksia vertailtiin aiemmissa tutkimuksissa havaittuihin tuloksiin riippumattomien otosten t-testillä sekä tarkastelemalla tässä tutkimuksessa havaittujen korrelaatiokertoimien luottamusvälejä. T-testien perusteella osatesteistä saadut raakapistemäärät ovat korkeampia tässä tutkimuksessa kuin mitä testistölle tehdyissä aiemmissa tutkimuksissa on havaittu. Toisin sanoen tämän tutkimuksen otos on suoriutunut testistöstä keskimääräistä korkeammilla pistemäärillä aiempien tutkimuksien otoksiin verrattuna.

Tämän tutkimuksen osallistujat rekrytoitiin lähettämällä tutkimuskutsu yli sadan ainejärjestön sähköpostilistalle eli tutkimuksen otos koostuu siis yliopisto-opiskelijoista. Korkeakoulussa opiskeleminen vaatii monia kognitiivisia taitoja, joten yliopistossa opiskelevien kognitiivinen kyvykkyys saattaa jo lähtökohtaisesti olla korkeampaa. On myös mahdollista, että tällaiseen vapaaehtoiseen kognitiota kartoittavaan tutkimukseen hakeutuvat todennäköisemmin henkilöt, joiden kognitiivinen kyvykkyys on korkeaa. Jos henkilö mieltää kognitiiviset testit vaikeiksi tai epämiellyttäväiksi, kynnys lähteä tällaiseen vapaaehtoiseen tutkimukseen mukaan voi olla suurempi.

Sen sijaan aiempien tutkimuksien otokset koostuvat Mercuri Urvalin soveltuvuusarviointiprosessiin osallistuneista henkilöistä. Suurin osa Mercuri Urvalin asiakkaista ovat korkeakoulutettuja, mutta asiakkaina on myös muun koulutustaustan omaavia henkilöitä. Laajempi kirjo koulutustaustasta saattaa tarkoittaa myös laajempaa kirjoa kognitiivisessa kyvykkyudessa. Soveltuvuusarviointiprosessiin osallistuneet henkilöt ovat myös lähtökohtaisesti osallistuneet rekrytointiprosessiin, eivätkä kognitiiviseen tutkimukseen. Soveltuvuusarviointiin tuleva ihmismassa todennäköisesti eroaa pelkästään kognitiiviseen tutkimukseen hakeutuvasta ihmismassasta. Nämä erot otosten välillä voivat selittää eroja otosten keskimääräisessä suoriutumisessa.

Yksi ero Mercuri Urvalin asiakaskunnan ja tutkimukseen hakeutuneen otoksen välillä on myös ikä. Tässä tutkimuksessa testattavien ikä vaihteli välillä 19–53, mutta noin 82 % testattavista olivat 19–30-vuotiaita. Mercuri Urvalin asiakaskunta kuitenkin painottuu tätä vanhempiin henkilöihin. Prosessointinopeus yleensä hidastuu iän myötä, minkä vuoksi

aikarajalliset kykyrakennetestit saattavat jossain määrin suosia nuorempia henkilöitä (Psykologiliitto, 2019). Erot tämän otoksen ja Mercuri Urvalin asiakaskunnan ikärakenteessa saattavat siis osittain selittää eroa keskimääräisessä suoriutumisessa.

On myös huomioitavaa, että opiskelijat saattavat olla tottuneempia toimimaan koetilanteissa kuin työelämässä olevat henkilöt. Opiskelijat ovat tottuneet tekemään tenttejä ja tenttitilanteet muistuttavat testitilanteita muun muassa siinä, että tenttitilanteet ovat aikarajattuja suoritustilanteita, joissa henkilön tulisi keskittyä vaan omaan tenttisuoritukseensa.

Työelämässä vastaavanlaisia yhtä rajattuja koetilanteita saattaa tulla eteen harvemmin, jolloin niissä toimiminen ei ole yhtä tuttua. Täten yliopisto-opiskelijoiden kohdalla testitilanteiden tietynlainen tuttuus ja testitilanteisiin liittyvä harjoiteltu paineensietokyky saattavat parantaa heidän suoriutumistaan myös kognitiivisessa testitilanteessa, kun taas työelämässä olevien henkilöiden kohdalla testitilanteeseen liittyvät vieraat elementit saattavat vaikuttaa heidän suoriutumiseensa alentavasti.

Vaikka tämän tutkimuksen otoksen suoriutuminen olikin keskiarvallisesti korkeampaa, MU CT -testistön osa-alueiden väliset korrelaatiot olivat pitkälti samankaltaisia tässä ja aiemmissa tutkimuksissa. Ainoastaan numeerisen ja kielellisen alaskaalan sekä abstraktin alaskaalan ja yleisen päättelykyvyn väliset korrelaatiot aiemmissa tutkimuksissa eivät istuneet tässä tutkimuksissa havaittujen samojen korrelaatioiden luottamusvälien sisään. Nämä aiemmissa tutkimuksissa havaitut korrelaatiot olivat kuitenkin vain kahden ja yhden mittayksikön verran korkeampia kuin vastaavien korrelaatioiden luottamusvälien yläraja, joten ero ei ole dramaattisen suuri.

Uusi testialusta ja uudet normit siis vaikuttavat toimivan vanhojen lailla, vaikka tämän tutkimuksen otos olikin keskiarvallisesti kognitiivisesti kyvykkäämpi. Suurin osa MU CT -testistön osa-alueiden välisistä korrelaatioista olivat samansuuruisia ja erot kahden korrelaation kohdalla olivat pieniä. Otoksen suurempi kognitiivinen kyvykkyys selittyy todennäköisemmin muilla tekijöillä kuin testialustaan tai normistoon liittyvillä tekijöillä.

4.2 Tutkimuksen vahvuudet ja rajoitukset

Tutkimuksen otos vastaa haluttua kohderyhmää eli Mercuri Urvalin asiakaskuntaa. Asiakaskunta koostuu pitkälti korkeakoulutetuista työssäkäyvistä henkilöistä ja yliopisto-opiskelijat ovat tulevaisuuden korkeakoulutettuja työntekijöitä. Tutkimuskutsut myös lähetettiin monien eri alojen opiskelijoille, joten tutkimuksen otos todennäköisesti koostuu

useampaa eri pääainetta opiskelevista henkilöistä. Tutkimuksen ilmoittautumisen yhteydessä ei kuitenkaan kysytty opiskelualaa, joten testattavien pääaineista ei ole varmuutta.

Tutkimuksen tuloksia voidaan siis jokseenkin yleistää eri alojen ammattilaisiin, mutta yleistämisessä tulee käyttää harkintaa. Otos on myös suhteellisen pieni, mikä rajoittaa tulosten yleistettävyyttä.

Intelligence Structure Test -testistö on tähän tarkoitukseen toimiva vertailutesti monesta syystä. IST-testistön suomenkielisen version psykometriset ominaisuudet eli reliabiliteetti ja valideetti ovat tutkimuksien perusteella hyvät (Hogrefe, 2017), mikä tekee IST-testistön tuloksista päteviä vertailutuloksia. IST-testistön teoriatausta (Hogrefe, n.d. -a) on myös hyvin samankaltainen kuin MU CT -testistön (Carlsted ym., 2020), jolloin käsitevaliditeetin arvioiminen rinnakkaisvaliditeetin kautta on testistöjen välillä mielekästä. Teoriataustansa mukaisesti kummatkin testistöt mittaavat yleistä päättelykykyä sekä samoja päättelykyvyn alaskaaloja. Testistöjen tulokset myös standardoidaan samalle numeeriselle välille, mikä mahdollistaa standardoitujen tulosten välisen vertailun.

On kuitenkin huomioitavaa, että vaikka testistöjen abstrakti ja visuaalisspatiaalinen alaskaala molemmat kartoittavat päättelykyvyn visuaalisspatiaalista ulottuvuutta, IST-testistön alaskaala painottaa visuaalisspatiaalista kyvykkyyttä enemmän. Alaskaalojen painotusero rajoittaa niiden tuloksien vertailua. Testistöjen tulosjakaumista kävi myös ilmi, että testistöjen erottelukyky erosi toisistaan tällä otoksella, sillä testattavat saivat korkeampia pisteitä IST-testistön visuaalisspatiaalisesta alaskaalasta ja yleisestä päättelykyvystä kuin MU CT -testistön abstraktista alaskaalasta ja yleisestä päättelykyvystä.

MU CT -testistön normisto on kerätty Mercuri Urvalin soveltuvuusarviointiprosesseissa (Carlsted ym., 2020). IST-testistö tarjoaa pisteytykseen kahdet henkilöarvioinneissa kerätyt normit (Hogrefe, n.d. -a), joista tähän tutkimukseen valittiin ylemmälle korkea-asteelle ja tutkijakoulutuksen käyneille testattaville soveltuvat normit. Täten molempien testistöjen normistot ovat sekä vertailukelpoiset keskenään että soveltuvat halutulle populaatiolle eli Mercuri Urvalin asiakaskunnalle. Käytetyt normistot tukevat siis testistöistä saatujen tulosten yleistettävyyttä halutussa populaatiossa.

Sekä MU CT -testistö että tutkimuksessa käytetty IST-testistön internetversio on suunniteltu tietokoneella tehtäväksi. Testistöjen samanlaiset suoritustavat tukevat testistöjen tulosten vertailemista. Tietokoneella suoritettavat testistöt ovat myös jokaiselle tutkittavalle täysin samanlaiset, sillä testien ohjeistuksessa tai aikarajoissa ei ole eroja testattavien välillä. Tämä

tukee testitilanteiden yhteneväisyyttä otoksessa. Lisäksi voidaan olettaa, että nykypäivänä opiskeleminen vaatii sen verran tietoteknisiä taitoja, että opiskelijaotoksella tietoteknisten taitojen puute ei ole vääristänyt tietokoneella suoritetuista testistöistä saatuja tuloksia.

Testattavia oli ohjeistettu suorittamaan testistöt erillisinä päivinä rauhallisessa tilassa ja varaamaan kummankin testistön tekemiseen puolitoista tuntia aikaa. Testitilanteita ei siis kontrolloitu muuten kuin ohjeistamalla testattavia toimimaan tietyllä tavalla. Tämä ei poikkea Mercuri Urvalin soveltuvuusarviointiprosessin käytännöistä, sillä prosessin aikana asiakkaat suorittavat Mercuri Urvalin testipatteriston tyypillisesti itsenäisesti ennen haastatteluun tulemistä. Testitilanteita ei ollut mahdollista tässä tutkimuksessa kontrolloida enempää, mutta kontrolloimattomuus on voinut aiheuttaa tutkimustuloksiin epähaluttua varianssia.

Testattavat ovat saattaneet tehdä testistöt toisiin testattaviin verrattuina hyvin erilaisissa tilanteissa, ja sama testattava on myös saattanut tehdä IST-testistön ja MU CT -testistön toisiinsa verrattuna erilaisissa tilanteissa. Testattavien vireystilaa tai keskittymistä testien suorittamiseen ei ole voitu arvioida etätestauksessa. Testitilanteiden kontrolloimattomuus mahdollistaa myös vilpin niin, että vaikka jokainen testattava saikin personoidun testilinkin antamaansa sähköpostiosoitteeseen, testattava on voinut antaa toisen henkilön suorittaa testin itsensä sijasta. Tämä ei tässä tutkimuksessa haittaa, jos toinen henkilö on tehnyt molemmat testit, mutta se vääristäisi tuloksia, jos testattava on itse tehnyt toisen testin ja toinen henkilö toisen testin.

Kognitiivisten testistöjen validiteettia on tutkittu muissakin tutkimuksissa rinnakkaisvaliditeetin avulla (ks. esim. Krach ym., 2009; Lam ym., 2013), minkä vuoksi saman menetelmän hyödyntäminen tässäkin tutkimuksessa on perusteltua. MU CT -testistöä verrattiin vain yhteen toiseen kognitiiviseen testistöön, mutta validiteetin todentamista voidaan perustella painokkaammin, jos mittaria verrataan useampaan samaa ilmiötä kartoittavaan pätevään mittariin. Useampien kognitiivisten testistöjen ottaminen mukaan tutkimukseen ei kuitenkaan ollut mahdollista.

4.3 Jatkotutkimusehdotukset

Mikäli Mercuri Urvalin kognitiivisen testistön käsitevaliditeettia halutaan kartoittaa lisää rinnakkaisvaliditeetin avulla, testistöä tai sen osia voidaan verrata muihin kognitiivisiin testistöihin. Esimerkiksi MU CT -testistön abstraktin päättelykyvyn osa-aluetta eli BIG-matriisitehtävää voitaisiin verrata Raven Progressive Matrices -testiin. Ravenin matriisit

kartoittavat yleistä kognitiivista kyvykkyyttä abstraktin ja ei-kielellisen päättelykyvyn kautta (Mills ym., 2010), mikä voisi vastata MU CT -testistön abstraktia alaskaalaa paremmin kuin IST-testistön visuaalisspatiaalisen päättelykyvyn alaskaala. Lisävertailujen avulla voidaan myös pyrkiä tarkentamaan sitä, millaista prosessointia MU CT -testistön abstrakti ja kielellinen alaskaala mittaavat rinnakkaisvaliditeetin näkökulmasta, kun tässä tutkimuksessa havaitut kohtalaiset yhteydet IST-testistön vastaavien alaskaalojen kanssa ovat monitulkintaisia.

Jos käsitevaliditeetin ajatellaan sisältävät kaikki validiteetin alalajit (Hiltunen, 2009), MU CT -testistön käsitevaliditeettia voitaisiin arvioida keskittymällä myös muihin validiteetin alalajeihin. Esimerkiksi testistön sisältövaliditeettia voitaisiin tutkia tarkastelemalla testistön faktorirakennetta ja sitä, kartoittavatko testistön faktorit tarpeeksi laajasti haluttuja kognition osa-alueita. Tarvittaessa testistöön voisi lisätä osatestejä, jotta tehtäväkokonaisuudet kattavat halutun ilmiön tarpeeksi laajasti.

Kriteerivaliditeetin suhteen voitaisiin rinnakkaisvaliditeetin lisäksi tarkastella myös testistön ennustevaliditeettia. Ennustevaliditeetin tutkimisessa on kuitenkin muistettava se, että MU CT -testistön ei ole tarkoitus yksinään ennustaa hakijan työssä suoriutumista. Testistöä käytetään Mercuri Urvalin soveltuvuusarviointiprosessin osana, joten ennustevaliditeettitarkastelussa on mielekkäämpää ennemminkin tarkastella koko arviointiprosessin ennustevaliditeettia ja sitä, onko kognitiivinen testistö merkittävä ja mielekäs osa tätä prosessia.

On kuitenkin tärkeää pohtia, onko kriteerivaliditeetin kartoittaminen aina mielekästä soveltuvuusarviointikontekstissa (Wuorivirta, 2020). Soveltuvuusarviointiprosessissa pyritään kartoittamaan työnhakijan ominaisuuksia ja kykyjä, joiden pohjalta voidaan pohtia hänen sopivuuttaan erilaisiin työtehtäviin tai organisaatioihin (Psykologiliitto, 2019). Henkilön soveltuvuuden arvioiminen on kuitenkin aina tapauskohtaista eli henkilön ominaisuuksia ei pidä tarkastella lähtökohtaisesti sopivina tai sopimattomina (Wuorivirta, 2020). Myös työnkuvaa ja työtehtäviä voidaan muokata ja valikoida työntekijän mukaan, mikä tekee soveltuvuuden pohtimisesta vastavuoroista. Soveltuvuusarviointikontekstissa tarkkojen kriteerien asettaminen työntekijän soveltuvuudelle ei siis ole välttämättä aina mielekästä, jolloin kriteerivaliditeetin eli jonkin ennustajan ja kriteerin välisen yhteyden tarkastelukaan ei ole aina mielekästä.

Lähteet

- American Psychological Association. (n.d.). *Fatigue effect*. APA Dictionary of Psychology. Haettu 28.12.2022 osoitteesta <https://dictionary.apa.org/fatigue-effect>
- Bryan, V. M., & Mayer, J. D. (2020). A meta-analysis of the correlations among broad intelligences: Understanding their relations. *Intelligence*, 81(1):101469. <https://doi.org/10.1016/j.intell.2020.101469>
- Carlstedt, L., Hagafors, R., & Jonsson, E. (2020). *The Mercuri Urval Assessment Method* [tekninen raportti]. Mercuri Urval.
- Cattell, R. B. (1967). The theory of fluid and crystallized general intelligence checked at the 5–6-year-old level. *British Journal of Educational Psychology*, 37(2), 209–224. <https://doi.org/10.1111/j.2044-8279.1967.tb01930.x>
- Cronbach, L., & Meehl, P. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4), 281–302. <https://doi.org/10.1037/h0040957>
- GraphPad (n.d.). *T test calculator*. Haettu 19.10.2022 osoitteesta <https://www.graphpad.com/quickcalcs/ttest1/>
- Gustafsson, J-E. (1984). A unifying model of the structure of intellectual abilities. *Intelligence*, 8(3), 179–203. [https://doi.org/10.1016/0160-2896\(84\)90008-4](https://doi.org/10.1016/0160-2896(84)90008-4)
- Heineman, K. R., Middelburg, K. J., Bos, A. F., Eidhof, L., La Bastide-Van Gemert, S., Van Den Heuvel, E. R., & Hadders-Algra, M. (2013). Reliability and concurrent validity of the Infant Motor Profile. *Developmental Medicine & Child Neurology*, 55(6), 539–545. <https://doi.org/10.1111/dmcn.12100>
- Hiltunen, L. (18.02.2009). *Validiteetti ja reliabiliteetti*. [luento]. Graduryhmä, Jyväskylän yliopisto. http://www.mit.jyu.fi/ope/kurssit/Graduryhma/PDFt/validius_ja_reliabiliteetti.pdf

- Hogrefe Psychologien Kustannus Oy. (n.d. -a). *IST (Intelligence Structure Test)*. Haettu 12.03.2022 osoitteesta <https://www.hogrefe.fi/testit/IST>
- Hogrefe Psychologien Kustannus Oy. (n.d. -b). *IST – Intelligence Structure Test*. Haettu 28.12.2022 osoitteesta https://www.hogrefe.fi/tuote?product_id=110
- Hogrefe Psychologien Kustannus Oy. (2017). *Soveltuvuusarvioinnin erityisnormit IST-testissä* [tutkimusraportti].
- Kuuskorpi, T., & Heikkinen, J. (2014). *Psykologiset testit ja testaukset: tosiasiat, uskomukset ja pelot*. Luovin.
- Krach, S. K., Loe, S. A., Jones, W. P., & Farrally, A. (2009). Convergent validity of the Reynolds Intellectual Assessment Scales (RIAS) using the Woodcock–Johnson Tests of Cognitive Ability, Third Edition (WJ-III) with university students. *Journal of Psychoeducational Assessment*, 27(5), 355–365.
<https://doi.org/10.1177/0734282909331749>
- Laki yksityisyyden suojasta työelämässä 759/2004. Haettu 28.12.2022 osoitteesta <https://www.finlex.fi/fi/laki/ajantasa/2004/20040759>
- Latvala, A., Silventoinen, K., & Vuoksima, E. (2020). Mitä tiedetään perimän ja ympäristötekijöiden vaikutuksesta älykkyyteen? *Duodecim* 136(3), 285–90.
<https://www.duodecimlehti.fi/duo15364>
- Lam, B., Middleton, L. E., Masellis, M., Stuss, D. T., Harry, R. D., Kiss, A., & Black, S. E. (2013). Criterion and convergent validity of the Montreal Cognitive Assessment with screening and standardized neuropsychological testing. *Journal of the American Geriatrics Society*, 61(12), 2181–2185. <https://doi.org/10.1111/jgs.12541>
- Lenhard, W., & Lenhard, A. (2016). *Computation of effect sizes*. Psychometrica.
https://www.psychometrica.de/effect_size.html

- Mercuri Urval. (25.3.2022). *A guide for the measurement of G through Mercuri Urval Cognitive Tests* [testiraportti].
- Mercuri Urval. (n.d.). *Organisaatiomme*. Haettu 19.12.2022 osoitteesta <https://www.mercuriurval.com/fi-fi/keita-me-olemme/organisaatiomme/>
- Mercuri Urval. (2020). *The Mercuri Urval Cognitive Tests (MU CT)* [testiraportti].
- Mills, C. J., Ablard K. E., & Brody, L. E. (2010). The raven's progressive matrices: Its usefulness for identifying gifted/talented students. *Longitudinal Studies in Gifted Education*, 15(3), 183–186. <https://doi.org/10.1080/02783199309553500>
- Psykologiliitto. (2019). *Henkilöarviointi työelämässä – ohjeistus hyviksi käytännöiksi* [Psykologiliiton ohjeistus].
- Rubio-Codina, M., Araujo, M. C., Attanasio, O., Munoz, P., & Grantham-McGregor, S. (2016). Concurrent validity and feasibility of short tests currently used to measure early childhood development in large scale studies. *Plos One*, 11(8). <https://doi.org/10.1371/journal.pone.0160962>
- Salgado, J. F. (2017). Personnel Selection. Teoksessa O. Braddick (toim.), *Oxford Encyclopedia of Research in Psychology* (s. 1–32). Oxford University Press <https://doi.org/10.1093/acrefore/9780190236557.013.8>
- Schneider, W. J., & Newman, D. A. (2015). Intelligence is multidimensional: Theoretical review and implications of specific cognitive abilities. *Human Resource Management Review*, 25(1), 12–27. <https://doi.org/10.1016/j.hrmr.2014.09.004>
- Schmidt, F. L., & Hunter, J. E. (2004). General mental ability in the world of work: Occupational attainment and job performance. *Journal of Personality and Social Psychology*, 86(1), 162–173. <https://doi.org/10.1037/0022-3514.86.1.162>
- Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research

- findings. *Psychological Bulletin*, 124(2), 262–74. <https://doi.org/10.1037/0033-2909.124.2.262>
- Scroggins, W. A., Thomas, S. L., & Morris, J. A. (2008). Psychological testing in personnel selection, part II: The refinement of methods and standards in employee selection. *Public Personnel Management*. <https://doi.org/10.1177/009102600803700204>
- Spearman, C. (1904). General intelligence, objectively determined and measured. *American Journal of Psychology*, 15(2), 201–292. <https://doi.org/10.2307/1412107>
- Valkeinen, H., Anttila, H., & Paltamaa, J. (2014). *Opas toimintakyvyn mittarin arviointiin TOIMIA-verkostossa (1.0)*. [https://thl.fi/documents/974257/1449823/Mittariopas_VALMIS_090614+\(2\).pdf/b53595b9-15b8-4fa3-8765-23cd9221de8f](https://thl.fi/documents/974257/1449823/Mittariopas_VALMIS_090614+(2).pdf/b53595b9-15b8-4fa3-8765-23cd9221de8f)
- van Aken, L., van der Heijden, P. T., Oomens, W., Kessels, R., & Egger, J. (2019). Predictive value of traditional measures of executive function on broad abilities of the Cattell-Horn-Carroll theory of cognitive abilities. *Assessment*, 26(7), 1375–1385. <https://doi.org/10.1177/1073191117731814>
- Van Cutsem, J., Van Schuerbeek, P., Pattyn, N., Raeymaekers, H., De Mey, J., Meeusen, R., & Roelands, B. (2022). A drop in cognitive performance, whodunit? Subjective mental fatigue, brain deactivation or increased parasympathetic activity? It's complicated! *Cortex*, 155, 30–45. <https://doi.org/10.1016/j.cortex.2022.06.006>
- Vuoksima, E. (2019). Kognitiivisten toimintojen muutokset – mikä on ikääntymistä, mikä sairautta? *Duodecim*, 135(11), 1075–84. <https://www.duodecimlehti.fi/duo14952>
- Weber, M., Van Ancum, J., Bergquist, R., Taraldsen, K., Gordt, K., Mikolaizak, A. S., Nerz, C., Pijnappels, M., Jonkman, N. H., Maier, A. B., Helbostad, J. L., Vereijken, B., Becker, C., & Schwenk, M. (2018). Concurrent validity and reliability of the Community Balance and Mobility scale in young-older adults. *BMC Geriatrics*, 18(1), 1–10. <https://doi.org/10.1186/s12877-018-0845-9>

Wuorivirta, P. (2020). *MPS-toimintatyylitestin validiteetti* [pro gradu -työ, Turun yliopisto].

UTUPub Turun yliopiston julkaisuarkisto.

<https://www.utupub.fi/handle/10024/150235>