

Tartu Ülikool
Sotsiaalteaduste valdkond
Psühholoogia instituut

Terje Saunoris

**MITI 4.2.1 hindamissüsteemi rakendatavus
ja hindajatevaheline kooskõla simuleeritud tingimustes**
Magistritöö

Juhendaja: Inga Karton, PhD

Tartu 2025

MITI 4.2.1 hindamissüsteemi rakendatavus ja hindajatevaheline kooskõla simuleeritud tingimustes

Lühikokkuvõte

Motiveeriv intervjuerimine (MI) on tõenduspõhine nõustamisviis, mille tõhusus sõltub suurel määral selle rakendamise kvaliteedist. Usaldusväärne ja süsteemne hindamine on vajalik nii MI õpetamisel kui ka teadusuuringutes. Käesolev magistritöö keskendub MI hindamissüsteemi MITI 4.2.1 (*Motivational Interviewing Treatment Integrity*) rakendatavusele Eesti kontekstis ning hindajatevahelise kooskõla analüüsile olukorras, kus MI-hindamine toimub standardiseeritud patsientidega simuleeritud intervjuude põhjal.

Uuring viidi läbi Tartu Ülikooli õpetamise ja arendamise grandi raames meditsiinitudengitega, kes viisid läbi MI-seansid koolitatud standardiseeritud patsientidega. Analüüsi kaasati 14 salvestatud intervjuud, mida hindasid neli väljaõppega kodeerijat ning üks eksperthindaja. Hindamised viidi läbi MITI 4.2.1 alusel, kasutades globaalskaalasid ja käitumiskoode. Hindajatevahelist kooskõla hinnati klassisisese korrelatsioonikordaja (ICC) abil ning eksperthindajaga võrreldavust täiendati Pearsoni korrelatsioonide ja Cronbachi alfa analüüsiga.

Tulemused näitasid, et enamik skaalasid saavutas hea või väga hea hindajatevahelise usaldusväarsuse (ICC 0,73-0,97). Kõrgeim kooskõla ilmnis struktureeritud käitumiste, näiteks küsimuste ja lihtsate peegelduste puhul. Globaalskaalad nagu empaatia näitasid madalamat kooskõla. Analüüs toetab MITI kasutatavust Eesti õppekeskkonnas simuleeritud MI intervjuude hindamisel, kuid osutab ka vajadusele täpsemaks koolituseks keerukamate skaalade hindamisel.

Märksõnad: Motiveeriv intervjuerimine, MITI hindamissüsteem, hindajatevaheline kooskõla, kodeerija, standardiseeritud patsient

Applicability of the MITI 4.2.1 Assessment System and Inter-Rater Reliability in Simulated Conditions

Abstract

Motivational Interviewing (MI) is an evidence-based counseling method, and its effectiveness largely depends on the quality of its implementation. Reliable and systematic assessment is essential in both MI training and research. This master's thesis focuses on the applicability of the Motivational Interviewing Treatment Integrity (MITI) 4.2.1 coding system in the Estonian context and analyzes inter-rater reliability in the specific setting of MI assessments conducted on simulated interviews with standardized patients.

The study was conducted as part of a University of Tartu teaching development grant and involved medical students conducting MI sessions with trained standardized patients. A total of 14 recorded interviews were analyzed, each independently assessed by four trained coders and one expert rater. The evaluations followed the MITI 4.2.1 coding system, using both global scores and behavior counts. Inter-rater reliability was measured using the Intraclass Correlation Coefficient (ICC), with additional comparisons to the expert ratings made using Pearson correlations and Cronbach's alpha.

The results showed that most MITI scales achieved good to excellent inter-rater reliability (ICC 0.73-0.97). The highest agreement was observed for structured behaviors such as open questions and simple reflections. Global ratings, such as empathy, showed lower consistency. The findings support the use of MITI 4.2.1 in Estonian training contexts for simulated MI interviews and highlight the need for more targeted training when assessing more complex global dimensions.

Keywords: Motivational Interviewing, MITI assessment system, inter-rater reliability, coder, standardized patient

SISSEJUHATUS

Tänapäeva ühiskonnas on vaimne tervis üks keskseid heaolu määravaid tegureid. Kasvav teadlikkus, kõrgeenenud ootused ja suurenev koormus vaimse tervise spetsialistidele on toonud fookusesse vajaduse mitmekesiste, inimkesksete ja samas tõenduspõhiste sekkumisviiside järele (Sackett et al., 1996). Spetsialistid vajavad tööriistu, mis aitavad tõhusalt toetada inimesi elumuutuste tegemisel, enese paremal mõistmisel ja probleemidega toimetulekul.

Kaasaegne vaimse tervise tugi ei saa põhineda vaid ühel meetodil ega lähenemisel - vajalik on erinevate meetodite kombinatsioon, mis arvestab inimese individuaalsust, konteksti ja valmisolekut muutusteks. Samas peab iga kasutatav meetod vastama tõenduspõhise praktika standarditele: olema empiiriliselt toetatud, kohaldatav ning kvaliteetselt teostatav (Sackett et al., 1996).

Üks olulisemaid muutust toetavaid protsesse abisuhetes on inimese sisemine motivatsioon. Just motivatsiooni olemasolu või selle puudumine määrab tihti, kas sekkumised on tulemuslikud. See on suunanud tähelepanu nõustamisviisidele, mis mitte ainult ei vahenda teavet ega paku lahendusi, vaid aitavad kliendil leida oma sisemised põhjused ja valmisoleku muutuseks (Deci & Ryan, 2000; Miller & Rollnick, 2002).

Motiveeriv intervjuerimine (*Motivational Interviewing*, MI) on lähenemine, mis põhineb koostööl, empaatial ja autonoomiat toetaval suhtlemisel. Selle keskmes on põhimõte, et inimesed on võimelised muutuma, kui neile luuakse sobiv psühholoogiline keskkond, kus nad saavad oma kahtlusi uurida, väärtusi sõnastada ja sihte selgemalt teadvustada. MI ühendab endas kliendikeskse ja suunava lähenemise, olles seetõttu sobiv tööriist väga erinevates olukordades ja süsteemides - alates sõltuvusnõustamisest kuni krooniliste haiguste käsitlemiseni (Miller & Rollnick, 2002). MI on eesmärgistatud suhtlemisviis, mille sihiks on tugevdada inimese sisemist motivatsiooni ja pühendumust muutusele, tuues esile tema enda põhjused muutumiseks empaatilises ja toetavas õhkkonnas. See erineb direktiivsest või õpetavast lähenemisest, rõhutades partnerlust ja inimese õigust teha oma elu üle tähenduslikke otsuseid. MI on eriti tõhus olukordades, kus esineb ambivalentsus, see tähendab kui soov muutuda on olemas, kuid selle elluviimist takistavad sisemised või välised kahtlused ja hirmud. Sellisel juhul aitab MI klientidel neid vastuolusid uurida ning toetab refleksiooni oma väärtuste ja eesmärkide üle (Miller & Rose, 2009).

MI rakendamine Eestis on viimastel aastatel laienenud, seda nii koolituste kui ka praktilise kasutamise kaudu sotsiaaltöös, tervishoius ja hariduses. Üha enam rõhutatakse, et lisaks MI õpetamisele on vaja tagada ka oskuste kvaliteetne hindamine. See võimaldab nõustajal saada objektiivset tagasisidet ja toetab ühtlasi kutsealast arengut. MI hindamiseks on välja töötatud mitmeid süsteeme, millest üks levinumaid ja enim kasutatud on Motiveeriva intervjuerimise hindamissüsteem (*Motivational Interviewing Treatment Integrity*, MITI) (Moyers et al., 2015).

Kuigi MITI on rahvusvaheliselt levinud hindamisvahend ja selle kasutamine on Eestis viimastel aastatel laienenud, on selle rakendamine simuleeritud õppeolukordades, kus kasutatakse standardiseeritud patsiente, siiski vähe uuritud. Selline formaat erineb traditsioonilistest tööalastest näidisseanssidest ning eeldab kohandatud lähenemist hindamisele. Käesoleva uurimuse eesmärk on hinnata, kas MITI 4.2.1 hindamissüsteem toimib usaldusväärselt ja rakendatavalt olukorras, kus MI-hindamine toimub simuleeritud seansside põhjal, kasutades standardiseeritud patsiente. Keskne uurimisküsimus on, kas eesti keelde kohandatud MITI 4.2.1 võimaldab saavutada hindajatevahelist kooskõla sellistes tingimustes.

MI ajalooline kujunemine

Motiveeriv intervjuerimine kujunes 1980-ndate alguses Ameerika Ühendriikides, kui William R. Miller töötas sõltuvushäiretega klientidega ja otsis tõhusamat viisi, kuidas neid kaasata ja motiveerida. Esimesed ideed said kirjaliku kuju 1983. aastal avaldatud artiklis ning koostöös Stephen Rollnickiga arenes meetod välja 1990-ndatel, kui avaldati esimene MI käsiraamat (Björk, 2014). Alguses keskendus MI eelkõige alkoholi ja narkootikumide tarvitamise probleemidele, kuid juba 1990ndate lõpuks laienes selle kasutusala tervisekäitumise muutmisele, vaimse tervise sekkumistele ning sotsiaaltöö praktikasse (Burke et al., 2003; Madson et al., 2009). Aastate jooksul on MI kontseptuaalne raamistik süvenenud ning saanud üha rohkem teaduslikku tuge (Miller & Rose, 2009).

MI arengulugu peegeldab ka sekkumise enda muutuvat olemust. Kui esialgu oli rõhk tehnikatel ja strateegiatel, siis hiljem sai keskseks MI *spirit* ehk vaimsus (Björk, 2014). MI vaimsuse kontseptsioon, mis rõhutab koostööd, esiletoomist ja autonoomiat toetavat hoiakut, võeti kasutusele vastusena MI laialdasele levikule ja murele selle väärkasutuse üle (Rollnick & Miller, 1995, viidatud Björk, 2014). Björk (2014) toob välja, et motiveeriv intervjuerimine

kujunes algselt kui nn voolav tehnoloogia (*fluid technology*) - paindlik ja kontekstitundlik lähenemine, millel puudusid jäigad piirid ning mida sai kasutada väga erinevates olukordades ja vormides. See paindlikkus soodustas MI kiiret levikut eri valdkondades, kuid tekitas samal ajal ka väljakutseid sekkumise identiteedi ja tõenduspõhisuse stabiliseerimisel. Selleks loid kesksed arendajad mitmeid täpsustavaid mehhanisme: välja töötati MI hindamissüsteemid (nt MISC ja MITI), loodi koolitajate võrgustik MINT ning avaldati täpsustavaid käsitlusi, et MI tähendus ei taanduks vaid tehnikate kogumikuks. Aja jooksul kujunes olukord, kus metaanalüüsid hakati viitama MI tõhususele tervikuna, kuigi uuritud olid sageli erinevad MI kohandused või versioonid. See tähendas omamoodi retoorilist sulgumist: teaduskirjanduses juurdus arusaam, et „MI töötab“, hoolimata selle sisulise mitmekesisuse püsimisest (Björk, 2014).

MI rakendusvõimalused

MI sobib hästi erinevate käitumuslike muutustega seotud olukordadesse, näiteks sõltuvusprobleemid, kroonilised haigused, elustiilivalikud, toitumisharjumused ja psühhosotsiaalne riskikäitumine (Fu et al., 2015; Lundahl et al., 2010). Lisaks kasutatakse MI-d üha enam ka hariduses, karjäärinõustamises, kriminaalhoolduses ja perega seotud sotsiaaltöös (Hohman, 2021; Self et al., 2023).

MI kasutusvõimalused ei piirdu üksnes individuaalse nõustamisega. Seda rakendatakse ka rühmasekkumistes, telefonikonsultatsioonides ja digitaalselt vahendatud sekkumistes. Erinevad uuringud on näidanud, et MI saab edukalt integreerida ka teiste sekkumismudelitega, näiteks kognitiiv-käitumisteraapia või kogukonnapõhiste programmidega (Britt et al., 2023; Madson et al., 2009).

Tervishoius on MI näidanud efektiivsust elustiili muutmise ja krooniliste haigustega toimetuleku toetamisel (Fu et al., 2015; Lundahl et al., 2010). Vaimse tervise valdkonnas on MI-d kasutatud ravi alustamise ja jätkamise toetamiseks, sh depressiooni, ärevuse ja PTSD kontekstis (Burke et al., 2003). Sotsiaaltöös aitab MI tugevdada klientide autonoomiat ja osalust, näiteks noorukite või perevägivaldaga ohvritega töötades (Hohman, 2021).

Karistusõiguses ja kriminaalhoolduses on MI-d rakendatud retsidiivsusvastaste programmide osana, et toetada elumuutusi ja vältida korduvat õigusrikkumist (Britt et al., 2023).

Samuti on MI leidnud rakendust veterinaarmeditsiinis ja hariduses, toetades klientide koostööd raviskeemide järgimisel ning suurendades õppimismotivatsiooni (Bard et al., 2023).

MI tõhususe hindamine

MI tõhusus sõltub suuresti sellest, kuidas ja millise kvaliteediga meetodit rakendatakse. Miller ja Rose (2009) toovad esile kaks omavahel seotud komponenti, mis selgitavad MI toimemehhanismi: suhtetasandi komponent, mis hõlmab empaatilist ja koostöist suhet, ning tehniline komponent, mis keskendub muutusjutu (*change talk*) esilekutsumisele ja samaksjäämisejutu (*sustain talk*) vältimisele. Muutusjutt viitab kliendi enda avaldustele, mis väljendavad motivatsiooni või valmisolekut muutuseks, samas kui samaksjäämisejutt sisaldab põhjusi, miks inimene soovib oma käitumist samaks jätta.

Lisaks mõjutavad MI tulemusi ka seansi kestus, nõustaja oskused, kliendi muutusvalmidus ja sekkumise kultuuriline sobivus (Lundahl et al., 2010). Kuna uuringute tulemused on tugevalt seotud sellega, kui hästi MI põhimõtteid järgiti, on praktikas ja teadusuuringutes hakatud järjest enam rõhutama vajadust hindamissüsteemide järele, mis võimaldavad hinnata MI rakendamise kvaliteeti. Üks selline süsteem on MITI, mille abil saab hinnata nii suhtlemisviisi kui ka konkreetsete tehnikate kasutamist. MITI võimaldab eristada juhuslikku motiveeriva keelekasutuse sarnasust teadlikust ja järjepidevast MI rakendamisest ning on kujunenud oluliseks tööriistaks nii koolituses, supervisioonis kui teadusuuringutes (Moyers et al., 2005; Moyers et al., 2016).

MOTIVEERIVA INTERVJUEERIMISE HINDAMINE

Motiveeriv intervjuerimine (MI) saab olla tõhus muutuse toetamisel vaid juhul, kui seda rakendatakse oskuslikult ja meetoodiliselt. Selleks, et hinnata, kas nõustamisel järgitakse MI põhimõtteid, on loodud motiveeriva intervjuerimise hindamisvahend MITI (*Motivational Interviewing Treatment Integrity Scale*). MITI võimaldab anda struktureeritud ja mõõdetavat tagasisidet nõustaja suhtlemisviisi ja tehniliste võtete kohta ning seeläbi toetada nõustamisoskuste arendamist. Tegemist on töövahendiga, mida kasutatakse nii MI tehnikate hindamiseks teadusuuringutes kui ka formaalse tagasiside andmiseks treeningutes ja supervisioonis. MITI abil saab käitumuslikult kirjeldada ja kvantitatiivselt hinnata nõustaja

tegevust vestluse käigus, sealhulgas selliseid komponente nagu refleksioonide kasutamine, muutusjutu esilekutsumine ja empaatia väljendamine (Moyers et al., 2015; Moyers et al., 2016).

MITI kujunemine ja areng

MITI esimene versioon töötati välja 2003. aasta paiku eesmärgiga luua ökonoomne ja usaldusväärne hindamisinstrument MI treeningu ja supervisiooni tarbeks (Moyers et al., 2005). See tugines varasemalt loodud MISC süsteemile, kuid erines sellest, keskendudes üksnes nõustaja verbaalsele käitumisele ning võimaldades oluliselt lihtsamat ja ajasäästlikumat kodeerimist (Pierson et al., 2007). Varasemates versioonides oli rõhk peamiselt käitumuslikel näitajatel, nagu avatud küsimuste ja refleksioonide kasutamine. MITI arendus jätkus läbi mitme versiooni. Alates versioonist 4.0 töötati välja uus struktuur, mis hõlmab nelja globaalskaalat (partnerlus, empaatia, muutusjutu esilekutsumine, samaksjäämisejutu pehmenemine) ja täiendatud käitumuslikke koode, mis peegeldavad MI põhikomponentide rakendamist (Moyers et al., 2016). Viimane versioon, MITI 4.2.1, ei lisa uusi skaalasisid, kuid täpsustab olemasolevate hindamiskriteeriumide rakendamist. Eesmärk on piirata hinnangute andmisel liiga liberaalset skoorimist ning aidata hindajatel paremini eristada struktureeritud MI kasutamist üldisest toetavast suhtlemisest. See saavutatakse kitsamate definitsioonide ja kõrgema hindamiskünnise abil, näiteks empaatia või partnerluse skaalal, et vältida MI vaimu pealiskaudset hindamist (Moyers et al., 2016).

Üks MITI peamisi eeliseid teiste seansikodeerimise tööriistade, näiteks MISC-i (*Motivational Interviewing Skill Code*) või BECCI (*Behaviour Change Counselling Index*) ees, seisneb selles et see võimaldab anda nõustaja tegevusele objektiivseid ja mõõdetavaid hinnanguid. MISC on ajamahukam ning sobib pigem põhjalikeks uurimuslikeks analüüsiks, kuna kodeerib ka kliendi kõnet. BECCI seevastu on lihtsam kontrollnimekirja tüüpi tööriist, mis sobib pigem lühinõustamiste hindamiseks, kuid ei võimalda põhjalikku MI kvaliteedi analüüsi ning on pigem subjektiivne (Lane et al., 2005).

Oluline on rõhutada, et kuigi MITI ei paku täielikku ülevaadet kliendiprotsessidest ega MI keerukamate strateegiatest (nt väärtuste ja käitumise vahelise vastuolu loomine), on see osutunud väga tõhusaks tööriistaks MI oskuste hindamisel, treeningutes ja supervisioonis. Kuna MITI mõõdab vaid nõustaja käitumist, ei asenda see protsessimeetmeid, kuid annab tõenduslikult usaldusväärse pildi seansi kvaliteedist ja MI-põhimõtetest kinnipidamisest (Moyers et al., 2016).

MITI kasutusvaldkonnad

MITI on laialdaselt kasutusel erinevates valdkondades, sealhulgas sõltuvusravis, krooniliste haiguste nõustamisel, vaimse tervise teenustes, HIV-ennetuses, karistusõiguses, hariduses ja isegi veterinaarmeditsiinis (Bard et al., 2023; Martino et al., 2007; Owens et al., 2017). Hindamissüsteemi saab kohandada ka lühikonsultatsioonide ja simuleeritud patsientidega läbiviidud intervjuude analüüsimiseks, mistõttu on see sobiv vahend ka MI õpetamisel ja supervisioonis (Beckman et al., 2022; Lane et al., 2008).

Rahvusvaheliselt on MITI kujunenud üheks enim kasutatavaks tööriistaks MI kvaliteedi tagamiseks nii teadustöös kui kliinilises praktikas. Selle eeliseks on mõõdetavus, usaldusväärsus ja rakendatavus mitmesugustes kontekstides (Moyers et al., 2016).

Eestis on MITI kasutamine arenenud koos MI koolituste ja kogukonna kasvuga. Tähtsat rolli on mänginud Eesti Motiveeriva Intervjueerimise ja Treeningu Assotsiatsioon (EMITA), mis on alates 2011. aastast korraldanud MI ja MITI koolitusi ning toetanud rahvusvahelist koostööd, sh MINT treenerite koolitusi ja foorumeid. MITI on Eestis rakendust leidnud sotsiaaltöös, hariduses, tervishoius ja sõltuvusnõustamises ning seda on kasutatud mitmes teadusprojekti ja lõputöös.

MITI struktuur ja tööprotsess

MITI hindamissüsteem koosneb kahest põhilisest osast: globaalskaaladest ja käitumiskoodidest. Need täiendavad teineteist, pakkudes nii kvalitatiivset kui kvantitatiivset teavet nõustaja tegevuse kohta MI vestluses (Moyers et al., 2015).

Globaalskaalad peegeldavad hindaja üldmuljet nõustaja suhtlemisviisist neljas valdkonnas:

- *Muutusejutu kultiveerimine* - nõustaja võime esile kutsuda kliendi sisemisi muutusmotive,
- *Samaksjäämisejutu pehmendamine* - kuidas nõustaja reageerib ambivalentsusele või vastupanule,
- *Partnerlus* - koostöine ja võrdne suhe kliendiga,
- *Empaatia* - nõustaja tundlikkus ja arusaamine kliendi kogemusest.

Kõiki skaalasid hinnatakse 5-pallisel Likerti skaalal, kus kõrgemad väärtused viitavad tugevamale vastavusele MI põhimõtetele ning need hinnangud põhinevad hindaja tervikmuljel.

Käitumiskoodid tähistavad konkreetseid nõustaja verbaalseid tegevusi, mida hindaja loendab kogu hinnatava lõigu jooksul. MITI 4.2.1 süsteemis eristatakse järgmisi käitumiskategooriaid:

- Peegeldused - jagunevad lihtsateks (nõustaja kordab või parafraseerib kliendi öeldut) ja keerulisteks (nõustaja lisab tähendust, tõlgendab või esitab tunnetuslikke järeldusi).
- Küsimused - kõik nõustaja küsimused registreeritakse
- MI-le kohased käitumised - sisaldavad näiteks kliendi tugevuste kinnitamist, koostööle kutsumist ning autonoomia rõhutamist.
- MI-le mittekohased käitumised - nende alla kuuluvad veenmine ilma loata ning konfronteeruv suhtlemine, mille eesmärk on mõjutada klienti ilma koostöö või valikuvabaduseta.
- Info jagamine - neutraalne faktide või soovitude esitamine, mille eesmärk ei ole klienti mõjutada ega veenda.

Nende alusel arvutatakse suhtarvud (nt peegelduste ja küsimuste suhe, MI-kohaste vs mittekohaste käitumiste osakaal), mis võimaldavad anda struktureeritud tagasisidet (Moyers et al., 2016).

MITI rakendamine hõlmab MI-sessiooni salvestamist, selle transkribeerimist või otse kuulamist ning hindamist koolitatud kodeerija poolt. Tavaliselt hinnatakse ligikaudu 20-minutilist lõiku, mis valitakse juhuslikult, et säilitada esinduslikkus (Moyers et al., 2005). Hindamisprotsessis hinnatakse nelja üldhinnangut ning loendatakse käitumiskoode. Neid kasutatakse mitmesuguste suhete ja koondskooride arutamiseks, mis võimaldavad anda struktureeritud tagasisidet nii koolituse, supervisiooni kui teadustöö kontekstis (Moyers et al., 2016). Hindajad läbivad struktureeritud koolitusprotsessi, mille käigus arendatakse nende kodeerimisoskust ja järjepidevust. Töös on oluline hindamisoskusi regulaarselt ühtlustada ja kasutada supervisiooni (Forsberg et al., 2007). Beckman et al. (2022) rõhutavad, et enesehinnang ei asenda objektiivset hindamist, sest praktikud kipuvad üle hindama oma MI-oskusi.

MITI tõhususe hindamise vajadus

Kuigi MITI ei võimalda hinnata kõiki motiveeriva intervjuerimise (MI) nüansse, on see üks enim kasutatud hindamisinstrumente MI kvaliteedi tagamiseks. 2020. aastal läbiviidud metaanalüüs MI hindamissüsteemide kasutamise kohta (Hurlocker et al., 2020) näitas, et MITI oli märkimisväärselt kõige sagedamini kasutatav tööriist - seda rakendati 111 artiklis võrreldes 31 artikliga, kus kasutati selle põhilist konkurenti MISC-i (*Motivational Interviewing Skill Code*). MITI leidis enim kasutust koolitusuuringutes (47 artiklit 72-st; 65,3%) ja tõhususuuringutes MI oskuste täpsuse hindamiseks (52 artiklit 99-st; 52,5%).

Ka Eesti kontekstis on MITI otstarbekaim valik olemasolevate hindamisinstrumentide seast, kuna see on suhteliselt vähem ressursimahukas. Siiski juhivad Schmidt et al. (2019) tähelepanu sellele, et MITI rakendamine nõuab põhjalikku hindajate koolitust, piisavat kooskõlalisust ja ajamahukat kodeerimisprotsessi, mis võib piirata selle laiemat kasutust praktikas.

Käesoleva magistritöö aluseks olevas uuringus hinnati MITI 4.2.1 kehtivust olukorras, kus hindamine viidi läbi standardiseeritud patsientidega (SP) salvestatud õppeolukordade põhjal. Uuringu eesmärk oli muuta MITI kasutamine koolituskontekstis lihtsamini teostatavaks ja kättesaadavamaks - võimaldades tudengitel praktiseerida MI-oskusi kontrollitud, kuid realistlikes tingimustes ning saada samal ajal ka struktureeritud tagasisidet. Selline lähenemine loob võimaluse hinnata MITI sobivust mitte ainult teadusuuringutes, vaid ka praktilistes õpikeskkondades. Et selgitada, kui usaldusväärselt ja kehtivalt MITI sellistes tingimustes toimib, viidi läbi käesolev uurimus.

UURIMUSE EESMÄRK JA UURIMISKÜSIMUSED

Käesoleva töö eesmärk on hinnata MITI 4.2.1 hindamissüsteemi usaldusväärsust ja rakendatavust Eesti kontekstis, kasutades standardiseeritud patsientidega salvestatud motiveeriva intervjuerimise simulatsioone. Uuring keskendub sellele, kas eesti keelde kohandatud MITI võimaldab saavutada hindajatevahelist kooskõla olukorras, kus MI hindamine toimub simuleeritud, kuid struktureeritud õppeolukorras.

- Uurimust suunab järgmine peamine uurimisküsimus:
Mil määral võimaldab MITI 4.2.1 hindamissüsteem saavutada hindajatevahelist kooskõla standardiseeritud patsientidega salvestatud MI intervjuude hindamisel Eesti kontekstis?

Täpsustavad uurimisküsimused:

- Millisel tasemel on hindajatevaheline kooskõla MITI globaalskaalade ja käitumiskoodide lõikes?
- Kuidas erinevad üksikhindajate hinnangud eksperthindajast?
- Millised MITI skaalad on simuleeritud õppeolukorras usaldusväärsemalt hinnatavad?

MEETOD

Uuringu disain

Uuring viidi läbi Tartu Ülikooli õpetamise ja arendamise grandis (projekt nr 03-21-TT-21, 2021-2023) raames (Karton, 2021), mille eesmärk oli rakendada standardiseeritud patsiendi (SP) meetodit motiveeriva intervjuueerimise (MI) õpetamisel arstiteaduse tudengitele. Pilootuuring disainiti klassikalise mõju-uuringuna AB-disainiga, kus kõik osalejad viisid SP-ga läbi kaks intervjuud: ühe õppetöö alguses ja teise lõpus. Uus disain tugines varasema pilootuuringu tulemustele, mis viitasid, et SP-de kaasamine toetab MI-oskuste arengut tõhusamalt kui tudengite omavahelised rollimängud (Karton, 2022, viidatud Luik, 2023). Iga tudeng töötas sama SP-ga ning intervjuud tuginesid eelnevalt ühtlustatud stsenaariumitele ja protokollidele. Intervjuud salvestati ja hinnati MITI 4.2.1 hindamissüsteemi alusel.

Käesolev magistritöö moodustab ühe osa läbiviidud uuringust ning keskendub hindajatevahelise kooskõla analüüsile MITI 4.2.1 hindamissüsteemi põhjal, kasutades arstiteaduse tudengite standardiseeritud patsientidega läbiviidud intervjuusid. Töö eesmärk on hinnata hindamissüsteemi rakendatavust MI õpetamise ja oskuste hindamise kontekstis Eesti tingimustes

Standardiseeritud patsient

Standardiseeritud patsient (SP) on koolitatud isik, kes kehastab kontrollitud stsenaariumi põhjal tüüpilist patsienti. SP-meetod võimaldab õppijatel harjutada keerulisi suhtlusolukordi

turvalises ja struktureeritud keskkonnas, vähendades riski õppijale ja välistades kahju patsiendile (Abrahamson et al., 1969, viidatud Hamman, 2004). Simulatsioon võimaldab eksida ja õppida ilma tegelike tagajärgedeta ning seda on edukalt rakendatud ka teistes valdkondades nagu lennundus, sõjavägi ja meditsiin (Aabersold, 2016; Rutherford-Hemming & Jennrich, 2013).

Meditsiiniõppes kasutatakse SP-d üha enam ka suhtlemisoskuste, sealhulgas MI õpetamiseks (Kelly et al., 2014). SP võimaldab harjutada ambivalentsuse äratundmist, muutusjutu esilekutsumist ja sobimatu, nt konfronteeruva suhtlusstiili vältimist (Galal et al., 2018; Martino et al., 2007). Lisaks loob see võimaluse standardiseeritud hindamiseks ning sobib hästi sekkumiste mõju või oskuste arengu hindamiseks (Bard et al., 2023; Svensson et al., 2020). Samal ajal rõhutavad mitmed allikad, et SP-ga läbiviidud intervjuud ei pruugi peegeldada täielikult reaalses nõustamises toimuvat - SP-d võivad tugevdada mõne MI-elementi esinemist, nt empaatia või partnerlus (Atkins et al., 2016; Stokoe, 2013). Samas on leitud, et SP-olukorras hinnatud oskused korreleeruvad hästi tegelikus praktikas ilmnevatega (Bard et al., 2023). Mõned autorid soovivad kaaluda hindamiskünniste kohandamist vastavalt olukorra realistlikkusele, kuna mõnes kontekstis võib olla keerukam saavutada kõrgeid hindepunkte (Decker et al., 2013).

Eestis on SP-meetodi kasutamine MI õpetuses alles algusjärgus. Alates 2021/2022. õppeaastast on Tartu Ülikoolis tehtud süsteemset arendustööd. Sellele eelnesid EMITA koolitusuuring ja arstiteaduse tudengite pilootuuring (Karton & Jürjen, 2015; Karton, 2022, viidatud Luik, 2023). Mõlemas kasutati MITI hindamissüsteemi ning järeldati, et SP-d toetavad MI-oskuste arengut.

Valim

Uuringus osalesid Tartu Ülikooli arstiteaduskonna kolmanda kursuse eestikeelse õppekava tudengid, kes olid eelnevalt läbinud kursuse „Suhtlemise ja tervisekäitumise alused“, sh MI teoreetilised alused ja baastehnikad. Osalemine põhines vabatahtlikkusel ja informeeritud nõusolekul. Valimisse kaasati 41 tudengit 133 huvilise seast juhuvaliku teel. Iga tudeng viis läbi kolm SP-intervjuud, millest kaks - esimene ja viimane - salvestati ning kaasati analüüsi. Kokku analüüsiti 82 pseudonüümitud intervjuud. Valimi suurus määrati tuginedes pilootuuringule ja varasemale kirjandusele, mis näitas, et see maht võimaldab hinnata hindajatevahelist usaldusväarsust piisava täpsusega (Oh et al., 2015; Schwalbe et al., 2014).

Käesoleva magistritöö raames valiti salvestatud intervjuudest juhusliku valiku alusel välja 14 intervjuud, mida hindasid neli kodeerijat. Intervjuude ajaline jaotus oli tasakaalustatud ning kodeerimine viidi läbi ühtlaselt kolme kuu jooksul. Selline disain vastab MITI hindamispraktikatele, mille kohaselt usaldusvärsuse hindamise eelduseks on, et ühte salvestist hindab vähemalt kaks sõltumatut kodeerijat. Mida rohkem hindajaid osaleb sama intervjuu kodeerimises, seda täpsem ja usaldusväärsem on kooskõla hinnang (Moyers et al., 2016).

Kodeerijate koolitamine

Uuringu jaoks valiti neli sõltumatut kodeerijat, kes said väljaõppe MITI 4.2.1 hindamissüsteemi kasutamiseks. Koolitus hõlmas teoreetilist osa, praktilisi harjutusi ja ühiseid arutelusid. Hindamine viidi läbi pimemeetodina - kodeerijatel ei olnud infot tudengi isiku ega intervjuu ajastuse kohta.

Magistritöö autor osales uuringus nii SP rollis kui ka ühe kodeerijana. Autor läbis MI baaskoolituse 2022. aasta suvel ning sügissemestril MITI kodeerijate koolituse, osaledes aktiivselt hindamispraktikates ja -koosolekutel.

Protseduur

Intervjuud viidi läbi BigBlueButton (BBB) keskkonnas ning salvestati TÜ Moodle'i platvormile. Salvestistelt eemaldati isikutuvastavad andmed ja need pseudonüümiti enne hindajatele edastamist. Intervjuude sisuks olid eelnevalt koostatud SP-stsenaariumid, mis käsitlesid elustiili teemasid, näiteks kehaline aktiivsus ja toitumine. Esimene intervjuu toimus 2022. aasta sügissemestri 4.-5. nädalal, pärast lühitutvustust MI metoodikasse. Seejärel järgnes moodulipõhine õpe koos praktikumi- ja harjutusvoorudega. Teine intervjuu viidi läbi 11.-13. nädalal. Hindamine toimus 2023. aasta kevadsemestril ning iga kodeerija hindas 22 intervjuud.

Eetika

Uuringus ei kaasatud reaalseid patsiente ega töödeldud tundlikke isikuandmeid. Kõik osalejad andsid kirjaliku teadliku nõusoleku (Eetikaveeb, 2022). SP-ga sõlmiti konfidentsiaalsuslepingud. Üliõpilased ei saanud osalemise eest tasu; SP-d osalesid kas tasustatud rollis või õppeaine „Standardiseeritud patsient“ raames, mille eest omandati 3 EAP. Salvestisi kasutati üksnes uurimiseesmärgil ja neid säilitatakse kuni viis aastat. Isikustatud

andmetele pääsesid ligi ainult volitatud uurimisrühma liikmed. Andmetöötlus toimus kooskõlas Tartu Ülikooli andmekaitse reeglitega. Oluliseks peeti, et uue meetodika rakendamine ei kahjustaks õppijate heaolu ega põhjustaks neile stressi või ebamugavust.

STATISTILINE ANALÜÜS

Statistilise analüüsi eesmärk ja ülesehitus

Käesolevas uuringus hinnati MITI 4.2.1 hindamissüsteemi rakendatavust Eesti kontekstis, keskendudes hindajatevahelise kooskõla analüüsile standardiseeritud patsientidega läbiviidud motiveeriva intervjuerimise (MI) intervjuude põhjal. Analüüs tugines 14 salvestatud intervjuule, mida hindasid neli sõltumatut kodeerijat ning üks eksperthindaja.

Andmeanalüüs algas kirjeldava statistikaga, et saada ülevaade skooride jaotusest, skaalade sisemisest varieeruvusest ja hindajate esmastest erinevustest. Seejärel hinnati hindajatevahelist usaldusväarsust, kasutades klassisisest korrelatsioonikordajat (*Intraclass Correlation Coefficient*, ICC), mille abil saab hinnata, kui palju hinnangute varieeruvusest tuleneb hinnatavate tegelikest erinevustest ning kui palju jääb hindajate mõõtmisvigade või ebaühtlase hindamisstiili arvele.

Täiendavalt võrreldi iga hindaja hinnanguid eksperthindaja tulemustega, kasutades Pearsoni korrelatsioonikordajat (r), et hinnata hinnangute omavahelist lineaarset seost. Lisaks analüüsiti iga kodeerija kooskõla eksperthindajaga eraldi ka ICC abil, kasutades vastavat mudelit (ICC[3,1]), mis võimaldab hinnata kahe fikseeritud hindaja absoluutset vastavust skaala tasandil.

ICC rakendamine MITI kontekstis

ICC valimisel tuleb teha mitu meetodilist otsust, mis mõjutavad tulemuste tõlgendatavust: valida tuleb sobiv mudel nt ühe- või kahefaktoriline (*one-way or two-way*-), juhusliku (*random*) või fikseeritud efektiga (*fixed effect*), otsustada, kas hinnata üksiku hindaja või keskmise hindajate grupi kooskõla (*single vs average measures*), ning kas huvi pakub hindajate omavaheline järjestuslik kooskõla (*consistency*) või absoluutne kooskõla (*absolute agreement*) (Koo & Li, 2016; ten Hove et al., 2022).

Näiteks Moyers et al. (2016) kasutasid MITI 4.2.1 valideerimisuurings *two-way mixed-effects* mudelit, lähtudes eeldusest, et huvi pakkus just nende konkreetsete hindajate kooskõla, mitte üldistatavus kõigile võimalikele hindajatele. Valiti *average-measures* ICC koos *absolute agreement* määratlusega, kuna uurijate eesmärk oli hinnata, kui usaldusväärne on mitme hindaja keskmine hinnang, s.t kui ühtselt suudab hindajate rühm hinnata konkreetseid tunnuseid, mitte kui täpselt teeb seda üksikhindaja. Tulemused näitasid väga head usaldusväarsust: globaalskaalade ICC ulatus 0.86-0.91 ning käitumisloendite puhul oli vahemik 0.71-0.97, kusjuures harvemini esinevate käitumiste puhul (nt *Persuade with Permission*) esines madalamaid väärtusi.

Sarnast lähenemist järgis ka Schmidt et al. (2019) uuring, kus kasutati *two-way mixed effects* mudelit koos *average-measures* ICC ja *consistency* määratlusega. Autorid selgitasid, et valisid *consistency of agreement*, kuna huvi keskendus sellele, kas hindajad hindavad järjestusi samas suunas, mitte täpselt samu skooriväärtusi. *Average-measures* ICC valiti, kuna analüüsis kasutati mitme hindaja keskmist hinnangut. *Two-way mixed* model oli asjakohane, kuna hindajad olid kindlalt määratletud grupp, mitte juhuslik valim.

ICC metoodiline valik

Ten Hove et al. (2022) toovad välja, et ICC valik võib oluliselt mõjutada tulemuste tõlgendust. Näiteks *average-measures* ICC annab sageli kõrgema usaldusväarsuse hinnangu kui *single-measures* ICC, kuna keskmistamine vähendab juhuslikku viga. Kui uurimistöö eesmärk on hinnata, kui täpne oleks ühe hindaja hinnang, siis tuleks kasutada *single-measures* ICC-d. Kui aga huvi on treeningprogrammi või hindamissüsteemi üldise usaldusväarsuse vastu - st kui usaldusväärne oleks hindajate rühma keskmine tulemus - on sobivam *average-measures* ICC (Ten Hove et al., 2022). Vale mudelivalik võib viia üle- või alahinnatud usaldusväarsuseni. Näiteks kui kasutatakse *average-measures* ICC-d olukorras, kus tegelik huvi on üksiku hindaja usaldusväarsuse vastu, võib see anda eksitavalt kõrged väärtused (Koo & Li, 2016). Samuti on oluline määratleda, kas hindajate mõju tuleks käsitleda juhusliku või fikseeritud efektina - viimast soovitatakse ainult juhul, kui kõik võimalikud hindajad osalevad analüüsis, mis on harva realistlik (Ten Hove et al., 2022).

Käesolevas uuringus eeldati kõigi hindajate puhul eelnevat väljaõpet ja ühtsete hindamispõhimõtete järgimist, nagu nõuavad MITI rakendamise standardid (Moyers et al., 2016; Schmidt et al., 2019). ICC tüübi valik tehti Ten Hove et al., (2021) ja Koo ja Li (2016) soovitude alusel, lähtudes uuringudisainist ning hindamisskeemi ülesehitusest. Selline lähenemine vastab rahvusvaheliselt aktsepteeritud praktikatele MITI-süsteemi valideerimisel ja rakendamisel (Moyers et al., 2016; Schmidt et al., 2019). ICC arvutamiseks kasutati *two-way random effects* mudelit, mis on sobiv juhul, kui nii hindajaid kui hinnatavaid käsitletakse juhuslike teguritena, st et mõlemad esindavad suuremat võimalikku populatsiooni ning tulemusi soovitakse üldistada sellest valimist kaugemale (Hallgren, 2012; Koo & Li, 2016; Ten Hove et al., 2022). Tõlgendamiseks kasutati *absolute agreement* tüüpi ICC-d, mille eesmärk on hinnata mitte üksnes hindajate skooride korrelatsiooni, vaid ka nende täpset kattuvust. See on asjakohane olukordades, kus hindamiskriteeriume tuleb rakendada võimalikult ühtselt ning individuaalsed erinevused ei tohiks mõjutada tulemuste tähendust (Hallgren, 2012; Koo & Li, 2016; Moyers et al., 2016). ICC tüüp ja parameetrite seadistus järgivad MITI valideerimisartiklite soovitusi, milles rõhutatakse vajadust hinnata hinnangute täpset kooskõla, mitte pelgalt korrelatsioonilist sarnasust (Moyers et al., 2016)

Kasutatud statistiliste meetodite ülevaade

Ten Hove et al., (2022) käsitlevad klassisiseset korrelatsioonikordajat (ICC) kui universaalset usaldusväarsuse näitajat, mis on tõlgendatav nii klassikalise testiteooria (CTT) kui ka üldistusteooria (GT) raames. ICC väljendab, kui suur osa hinnangute koguhajuvusest on tingitud hinnatavate objektiivsetest erinevustest (tõelise skoori varieeruvusest) võrreldes veavariatsiooniga, mis hõlmab mõõtmisvigu ja hindajatega seotud allikaid (Ten Hove et al., 2022). Usaldusväarsuse üldvalem selles käsitluses on kujul:

$$ICC = \frac{\sigma_s^2}{\sigma_s^2 + \sigma_e^2}$$

kus: σ_s^2 tähistab hinnatavate varieeruvust (tõelist skoori) ja σ_e^2 mõõtmisviga (sh hindajate varieeruvus ja interaktsioon).

Käesolevas analüüsis kasutati *two-way random effects* mudelit, eeldades, et nii hindajad kui hinnatavad on juhuslikud tegurid, mis võimaldab üldistada tulemusi laiemale populatsioonile.

Valitud ICC tüüp oli *absolute agreement*, mis arvestab hindajate hinnangute täpset kattuvust, mitte ainult nende korrelatsiooni. Lisaks kasutati keskmistatud skoori mitme hindaja põhjal (*average-measures*), mis vähendab mõõtmisviga ning suurendab usaldusväarsust. Sellisel juhul vastab kasutatud mõõdik tähistusele ICC(2,k) (McGraw & Wong, 1996; Shrout & Fleiss, 1979), mille valem on järgmine:

$$ICC(2, k) = \frac{\sigma_s^2}{\sigma_s^2 + \frac{\sigma_r^2 + \sigma_{sr}^2}{k}}$$

kus: σ_s^2 tähistab hinnatavate varieeruvust (s.o skoor, mis ei sõltu hindajatest ega mõõtmisveast), σ_r^2 on hindajate varieeruvus, σ_{sr}^2 hinnatava ja hindaja interaktsioon ning mõõtmisviga ja k hindajate arv.

ICC väärtused jäävad teoreetiliselt vahemikku 0 kuni 1, kusjuures väärtus lähemal 1-le viitab suuremale hindajatevahelisele kooskõlale. Siiski võib väikese valimi või madala varieeruvuse korral ICC väärtus langeda ka alla nulli, mis näitab, et mõõtmisviga ületab hinnatavate üksuste tegelikud erinevused ning hindajate hinnangud ei ole sel juhul usaldusväärsed (Hallgren, 2012; McGraw & Wong, 1996).

Tõlgendamisel lähtuti üldtunnustatud piiridest (Koo & Li, 2016):

- ICC < 0.50 - madal usaldusväarsus
- 0.50-0.75 - mõõdukas
- 0.75-0.90 - hea
- ICC > 0.90 - väga hea usaldusväarsus

MITI 4.2.1 hindamissüsteemi rakendamisel tuleb arvestada, et erinevate kategooriate mõõdetavus võib varieeruda. Globaalskaalade puhul (nt empaatia, partnerlus) peetakse rahuldavaks usaldusväarsuseks sageli taset $ICC \geq 0.75$, samas kui sagedasemate ja kvantitatiivselt mõõdetavate käitumiskoodide puhul (nt küsimused, lihtsad peegeldused) võib piisavaks pidada ka taset $ICC \geq 0.60$ (Moyers et al., 2016; Schmidt et al., 2019).

Pearsoni korrelatsioonianalüüsi kasutamine võimaldas hinnata, kui tugev lineaarne seos esines üksikhindajate ja eksperthindaja skooride vahel skaala tasandil. Selline analüüs annab esialgset teavet selle kohta, kui lähedale hindajate hinnangud jäävad eksperthinnangutele ning kas hindamissüsteemi kriteeriumid on rakendatavad kooskõlas referentsstandardiga. See toetab lisaks usaldusvärsusele ka MITI 4.2.1 hindamissüsteemi konvergentset valiidsust, mida on varasemates ülevaadetes käsitletud olulise näitajana selle rakendatavuse hindamisel koolituste ja juhendamise kontekstis (Gill et al., 2020). Sarnast lähenemist on kasutatud ka eelnevates MITI hindajatevahelise reliaabluse uuringutes, kus Pearsoni r on esitatud koos ICC ja Cronbachi alfaga reliaabluse erinevate aspektide kirjeldamiseks (Jürjen, 2013).

Arvestamaks mitte ainult hindajate hinnangute lineaarset seost, vaid ka nende täpset kooskõla absoluutsel tasandil, arvutati igale üksikhindaja / eksperthindaja paarile klassisisene korrelatsioonikordaja (ICC). Kuna eksperthindaja oli fikseeritud isik (st mitte juhuslikult valitud, vaid esindas hindamisstandardit), kasutati analüüsis ICC(3,1) mudelit, mis sobib olukorda, kus üks hindajatest on fikseeritud (*fixed rater*) ning huvi keskmes on hinnangute absoluutne kooskõla (*absolute agreement*). Mudel põhineb *two-way mixed effects* disainil ja võimaldab hinnata täpset vastavust kahe kindlaksmääratud hindaja hinnangute vahel. Fikseeritud hindajate korral võtab valem arvesse ka hindaja ja hinnatava võimaliku interaktsiooni, mida korrigeeritakse hindajate arvuga:

$$C(3, 1) = \frac{\sigma_s^2 - \frac{\sigma_{r:s}^2}{k-1}}{\sigma_s^2 + \sigma_{r:s}^2 + \sigma_e^2}$$

kus: σ_s^2 on hinnatavate (intervjuude) vahelise varieeruvuse komponent, $\sigma_{r:s}^2$ tähistab hindaja ja hinnatava interaktsiooni, σ_e^2 mõõtmisviga ja k on hindajate arv (antud paarivõrdluse puhul $k=2$).

Selline lähenemine on kooskõlas mõõtmisusaldusvärsuse käsitlemisega klassikalises testiteoorias (Lord & Novick, 1968) ning ICC rakendamise juhistega usaldusvärsusuuringutes (Hallgren, 2012; McGraw & Wong, 1996; Shrout & Fleiss, 1979; Ten Hove et al., 2022).

TULEMUSED

Esmase ülevaate saamiseks leiti kirjeldavad statistikud skaalade kaupa. Osade käitumiskoodide väärtuste jaotused erinesid oluliselt normaaljaotusest: asümmeetriakordajad > 3 ning ekstsess > 9 . See viitab väärtuste kuhjumisele jaotuskõvera vasakule poolele ning jaotuskõvera teravale tipule, mis tähendab, et enamik hinnanguid olid nullilähedased ning vaid üksikud väärtused ulatusid kõrgemale. Nii Hallgren (2012) kui ka Koo ja Li (2016) rõhutavad, et väikese varieeruvusega muutujad (nt väga harva esinevad käitumised) võivad ICC analüüsis anda moonutatud tulemusi. Seetõttu osa vähemkasutatavaid üksikkoode koondati kahte sisuliselt eristuvasse kategooriasse: MI-le kohased ja MI-le mittekohased käitumised, et suurendada statistilist stabiilsust ning tagada usaldusväärsuse hindamise sobivus. Selline lähenemine on kooskõlas ka MITI rakendussoovitustega (Moyers et al., 2016).

Hindajatevahelise kooskõla hindamiseks arvutati iga skaala kohta klassisisene korrelatsioonikordaja (ICC). Tulemused on esitatud tabelis 1. Enamik skaaladest näitas head kuni väga head hindajatevahelist kooskõla, kusjuures ICC väärtused jäid vahemikku 0,73–0,97.

Tabel 1

Kodeerijatevaheline kooskõla

Skaala	ICC (average)	95% CI	Cronbachi α
MJ - muutusejutt	0,841	[0,632, 0,943]	0,879
SJ -samaksjäämisejutt	0,802	[0,559, 0,928]	0,837
P - partnerlus	0,893	[0,763, 0,961]	0,898
E - empaatia	0,765	[0,488, 0,914]	0,794
I - info jagamine	0,807	[0,576, 0,930]	0,824
VL - veenmine loaga	0,808	[0,576, 0,930]	0,816
K - küsimused	0,950	[0,888, 0,982]	0,951
PL - lihtsad peegeldused	0,909	[0,795, 0,967]	0,919
PK - keerulised peegeldused	0,733	[0,386, 0,905]	0,721
MIA - MI-le kohane käitumine	0,930	[0,844, 0,975]	0,931
MIN - MI-le mittekohane käitumine	0,969	[0,930, 0,989]	0,969
P/K - peegelduste/küsimuste suhe	0,930	[0,815, 0,976]	0,953

Märkus: Kõik ICC väärtused ($p < .001$)

Kõige kõrgem kooskõla saavutati skaala MI-le mittekohaste käitumiste puhul (ICC = 0,97, 95% CI [0,93, 0,99]), samas kui madalaim väärtus ilmnes keeruliste peegelduste (PK) skaalal (ICC = 0,73, 95% CI [0,39, 0,91]). Kõik ICC väärtused olid statistiliselt olulised ($p < .001$). Lisaks arvutati iga skaala kohta Cronbachi alfa, mis toetas üldiselt ICC tulemusi ja viitas skaalade heale sisemisele järjekindlusele.

Nagu soovitatud Koo ja Li (2016) poolt, tuleks ICC väärtuste tõlgendamisel arvestada ka nende 95% usaldusvahemikke, et hinnata hinnangute täpsust ja usaldusvärsuse ulatust. Usaldusvahemik, mis ulatub mitme tõlgenduskategooria vahele, võib viidata ebakindlusele hinnangu täpsuses - eriti juhul, kui alumine piir langeb alla mõõduka usaldusvärsuse läve. Käesolevas analüüsis oli mitme skaala punktihinnang formaalselt hea või isegi väga hea, kuid usaldusvahemike laius viitab vajadusele tõlgendada neid ettevaatlikult. Näiteks skaala samaksjäämisejutt (SJ) punktihinnang ICC = 0,80 paigutub hea usaldusvärsuse kategooriasse, kuid selle 95% usaldusvahemik [0,56, 0,93] ulatub mõõdukast väga hea usaldusvärsuseni. Sama kehtib ka empaatia (E) skaala kohta, mille punktihinnang oli ICC = 0,77, ent usaldusvahemik [0,49, 0,91] langeb alla mõõduka piiri. Sellised tulemused viitavad sellele, et kuigi kooskõla võib olla hea, pole hinnang kuigi stabiilne. Keeruliste peegelduste (PK) puhul oli usaldusvahemik eriti lai - [0,39, 0,91], ulatudes madalast kuni hea usaldusvärsuseni. Selline laius viitab suurele varieeruvusele hindajate hinnangutes ning võib tähendada, et see skaala on hindamiseks raskemini ühtlustatav või vajab täpsemat definitsiooni ja täiendavat koolitust.

Lisaks ICC-le arvutati iga skaala kohta ka Cronbachi alfa, et hinnata hindajatevahelist järjekindlust samade mõõdikute hindamisel. Kõikide skaalade alfa väärtused jäid vahemikku [0,72, 0,97], viidates vähemalt rahuldavale sisemisele kooskõlale. Kuigi Cronbachi alfa ei arvesta süsteemset erinevust hindajate tasemes, toetab see koondnäitaja üldiselt ICC tulemusi, andes lisainfot hindamisskaalade rakendatavuse ja stabiilsuse kohta. Siiski tuleb arvestada, et Cronbachi alfa ei erista, kas hindajad hindavad samu üksusi samale või erinevale tasemele, vaid mõõdab üksnes nende hinnangute omavahelist korrelatsiooni. Seetõttu võib alfa olla kõrge ka juhul, kui kõik hindajad hindavad järjekindlalt süsteemselt liiga kõrgelt või madalalt (McGraw & Wong, 1996).

Pearsoni korrelatsioonikordajate tõlgendamisel (tabel 2 ja 3) lähtuti Cohen'i (1988) soovitatud piiridest, mille kohaselt r -väärtust vahemikus 0.1- 0.29 loetakse nõrgaks, 0.3 - 0.49 mõõdukaks ning 0.5 või enam tugevaks lineaarseks seoseks.

Tabel 2

Pearsoni korrelatsioonikordajad globaalskaalade lõikes

Skaala	Kõrgeim r /(kodeerija nr)	Madalaim r /(kodeerija nr)	Üldine tendents (Cohen'i järgi)
MJ	0.751** (1)	0.567* (3)	Keskised seosed
SJ	0.498 (1)	0.206 (4)	Nõrgad/keskmised
P	0.585* (2)	0.401 (4)	Keskised seosed
E	0.376 (3)	0.188 (1)	Nõrgad seosed

Märkus: * $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$

Analüüsis keskenduti seega eeskätt sellele, kas kodeerijate ja eksperthindaja hinnangud näitasid vähemalt mõõdukat või tugevat seost. Lisaks võeti arvesse, kas saadud seosed olid statistiliselt olulised ($p < .05$).

Tugevaimad korrelatsioonid ilmnemid käitumisloendite skaaladel (tabel 3): küsimused (K) $r=0,899$, $p < .001$, MI-le mittesoovitav käitumine (MIN) $r=0,913$, $p < .001$, peegelduste ja küsimuste suhe (P/K) $r=0,899$, $p < .001$ ja lihtsad peegeldused (PL) $r=0,882$, $p < .001$, kus kõikide hindajate Pearsoni r väärtused ületasid 0.5 ja olid enamasti statistiliselt olulised (tabel 3). Eriti kõrged kooskõlad eksperdiga ilmnemid hindajal nr 4 (tabel 3), kellel oli kõrgeim korrelatsioon kolmel skaalal: info jagamine (I) $r=0,542$, $p < 0,05$, MI-le kohased käitumised (MIA) $r=0,723$, $p < 0,01$, keerulised peegeldused (P/K) $r=0,899$, $p < .001$. See võib viidata hindaja oskusele tabada MI kvaliteeti eksperdi sarnaselt.

Tabel 3

Pearsoni korrelatsioonikordajate vahemik käitumisloendite lõikes

Skaala	Kõrgeim r /(kodeerija nr)	Madalaim r /(kodeerija nr)	Üldine tendents (Cohen'i järgi)
I	0.542* (4)	0.023 (3)	Nõrgad/keskmised
VL	0.678** (1)	0.503 (3)	Keskmised seosed
K	0.899*** (3)	0.844*** (1)	Tugevad seosed
PL	0.882*** (2)	0.743** (1)	Tugevad seosed
PK	0.630* (3)	0.217 (2)	Nõrgad/keskmised
MIA	0.723** (4)	0.549* (2)	Keskmised seosed
MIN	0.913*** (4)	0.787** (2)	Tugevad seosed
P/K	0.899*** (4)	0.677** (2)	Tugevad seosed

Märkus: * $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$

Mõned skaalad, nagu empaatia (E) $r=0,188$, samaksjäämisejutt (SJ) $r=0,206$, info jagamine (I) $r=0,023$ ja keerulised peegeldused (PK) $r=0,217$, näitasid madalamaid korrelatsioone, viidates kas raskustele nende skaalade rakendamisel või väiksemale tundlikkusele MI kvaliteedierinevuste suhtes. Samuti ilmnes, et hindaja nr 2 (tabel 3) korrelatsioonid eksperdiga olid mitmel juhul madalamad: keerulised peegeldused (PK) $r=0,217$, Mi-le mittekohane käitumine (MIN) $r=0,787$, peegelduste ja küsimuste suhe (P/K) $r=0,677$, mis võib viidata vajadusele täiendava juhendamise või harjutamise järele, kuigi kõik hindajad saavutasid mitmel skaalal tugeva seose.

Pearsoni korrelatsioonide analüüs ei anna infot hindamiste absoluutse kokkulangevuse kohta (selleks kasutati ICC väärtusi), kuid toetab tõlgendust, millised skaalad ja hindajad on paremini kooskõlas eksperdihinnangutega kvalitatiivses mõttes. Selline kodeerijate lõikes tehtud võrdlus aitab hinnata, kui ühtlaselt on hindamiskriteeriumid rakendatavad erinevate hindajate poolt ning milliste skaalade puhul võib esineda suuremat tõlgendamisruumi.

Tabelis 4 esitatud ICC väärtused näitavad hindajate hinnangute kooskõla eksperthindajaga, MITI 13 skaala lõikes.

Tabel 4

Klassisisene korrelatsioon, nelja kodeerija tulemused eraldi vs ekspert

Skaala	ICC / 95% CI Kodeerija nr 1	ICC / 95% CI Kodeerija nr 2	ICC / 95% CI Kodeerija nr 3	ICC 95% CI Kodeerija nr 4
MJ	0.762 *** (0.399 - 0.918)	0.604 * (0.111 - 0.855)	0.432 * (-0.070 - 0.772)	0.536 ** (0.045 - 0.822)
SJ	0.500 * (-0.016 - 0.806)	0.309 (-0.180 - 0.699)	0.413 * (-0.057 - 0.754)	0.215 (-0.378 - 0.664)
P	0.552 * (0.076 - 0.828)	0.487 * (0.004 - 0.796)	0.406 * (-0.068 - 0.751)	0.321 (-0.121 - 0.695)
E	0.188 (-0.353 - 0.638)	0.138 (-0.211 - 0.546)	0.276 (-0.134 - 0.660)	0.221 (-0.134 - 0.607)
I	0.390 (-0.187 - 0.757)	0.403 * (-0.064 - 0.748)	0.022 (-0.493 - 0.529)	0.473 * (0.0 - 0.787)
VL	0.606 ** (0.159 - 0.852)	0.428 * (-0.084 - 0.772)	0.309 * (-0.121 - 0.691)	0.318 * (-0.12 - 0.701)
K	0.829 *** (0.559 - 0.941)	0.823 *** (0.546 - 0.939)	0.905 *** (0.731 - 0.969)	0.892 *** (0.701 - 0.964)
PL	0.713 ** (0.318 - 0.898)	0.827 *** (0.544 - 0.941)	0.751 *** (0.303 - 0.918)	0.852 *** (0.612 - 0.949)
PK	0.509 * (-0.023 - 0.813)	0.220 (-0.336 - 0.661)	0.592 * (0.110 - 0.848)	0.557 * (0.053 - 0.833)
MIA	0.809 *** (0.515 - 0.934)	0.475 * (0.003 - 0.788)	0.531 ** (0.035 - 0.820)	0.676 ** (0.270 - 0.881)
MIN	0.647 ** (0.225 - 0.869)	0.737 *** (0.363 - 0.907)	0.762 *** (0.403 - 0.917)	0.894 *** (0.711 - 0.964)
P/K	0.833 *** (0.568 - 0.943)	0.605 ** (0.121 - 0.855)	0.789 *** (0.322 - 0.934)	0.896 *** (0.716 - 0.965)

Märkus: * $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$

Tulemustes eristuvad mitmed skaalad, mille puhul oli kooskõla järjekindlalt kõrge. Skaalad - küsimused (K), lihtsad peegeldused (PL), MI-le kohane (MIA) ja mittekohane käitumine (MIN) ja peegelduste küsimuste suhe (P/K) saavutasid kõigi või vähemalt kolme hindaja puhul ICC väärtuse üle 0.75, mis oli ühtlasi statistiliselt oluline. Näiteks küsimuste skaala puhul olid kõikide hindajate ICC väärtused vahemikus 0,823 - 0,905, osutudes kõige ühtlasemalt kõrge kooskõlaga skaalaks kogu analüüsis.

Mõõduka kooskõlaga, kuid statistiliselt olulisi tulemusi esines mitme skaala puhul (tabel 4), sh veenmine loaga (VL) ICC=0,318 - 0,606 ning MI-le kohased käitumised (MIA) ICC=0,475 - 0,809, kus mõne kodeeri hinnangud olid tugevalt vastavuses eksperthinnangutega, samas kui teistel jäid need madalaks või mitteolulisteks. See viitab võimalikele erinevustele kodeerijate tõlgendustes või skaalade raskusastmes.

Kõige madalamad ICC väärtused (tabel 4) ilmnisid skaaladel empaatia (E) ICC=0,138, samaksjäämisejutt (SJ) ICC=0,215 ja info jagamine (I) ICC=0,022, kus kooskõla ei olnud statistiliselt oluline ühegi hindaja puhul. Näiteks skaala empaatia (E) ICC väärtused jäid vahemikku 0.138 - 0.276, kusjuures ükski neist ei ületanud statistilise olulisuse piiri. Nende skaalade puhul võib oletada, et hinnatavad nähtused olid kas harva esinevad või keerukamad üheselt määratleda.

Hindajate võrdluses (tabel 4) paistis silma kodeeri nr 2, kellel oli 9 skaala puhul ICC väärtus vähemalt 0.403, need 9 ka statistiliselt olulised. Kodeerijal nr 4 oli samuti 9 olulist tulemust, kuid varieeruvus skaalade lõikes oli suurem. Kodeerijate nr 1 ja nr 3 puhul esines samuti mitmeid kõrgeid kooskõlasid (nt skaala küsimused (K) ICC=0,829 / 0,925), kuid ka rohkem madalaid või mitteolulisi väärtusi, eriti skaaladel empaatia (E) ICC=0,188 / 0,276 ja info jagamine (I) C=0,390 / 0,022.

ARUTELU

Tõendus põhiste psühholoogiliste sekkumiste usaldusväärne hindamine on olulise tähtsusega nii koolituspraktikas kui ka rakendusuringutes. Motiveeriva intervjuerimise (MI) puhul, kus suhtlusstiil ja kliendikesksus on kesksel kohtadel, eeldab hindamine mitte ainult

tehniliste oskuste tabamist, vaid ka keerukate suhtlemiselementide mõtestamist. Käesolev töö keskendus MITI 4.2.1 hindamissüsteemi kasutamisele Eesti kontekstis, eesmärgiga hinnata selle usaldusväärsust ja sobivust motiveeriva intervjuerimise õpetamisel standardiseeritud patsientide simulatsioonide põhjal.

Andmeanalüüsi algul leitud kirjeldavad statistikud viitasid mitmete MITI skaalade puhul tugevale väärtuste koondumisele madalale skoorivahemikule, kusjuures mõned käitumiskoorid esinesid andmestikus harva või olid sageli nullväärtusega. Selline jaotus ei ole erandlik, eriti olukordades, kus hindamisele tulevad algaja taseme õppijad ja hinnatavateks on keerulisemad või vähem kasutatavad MI-käitumised. Sarnaseid tulemusi on kirjeldatud ka varasemates uuringutes. Näiteks Schmidt et al. (2019) juhivad tähelepanu sellele, et mitmete käitumiskoodide, nagu veenmine loaga (VL) või konfronteerumine (K), esinemissagedus kipub olema sedavõrd madal, et nende usaldusväärne hindamine statistiliste meetoditega on piiratud. Britt et al. (2022) ja Hallgren (2012) on omakorda märkinud, et sellistel juhtudel võivad hindajate skoorid koonduda nullpunkti ümber, mis vähendab varieeruvust ja moonutab kooskõla statistilisi hinnanguid. Seetõttu rõhutatakse mitmes allikas, et hindajatevahelise kooskõla hindamisel tuleb selliseid jaotusi arvesse võtta ning vajadusel kohandada andmete struktureerimist.

Käesolevas analüüsis lähtuti MITI juhendmaterjalidest (Moyers et al., 2016), mille kohaselt võib madala sagedusega üksikkoodide korral kasutada ka koondkategoriaid nagu MI-le kohane ja MI-le mittekohane käitumine. See võimaldab suurendada mõõdikute stabiilsust, säilitades samas sisulise tähenduse MI hindamises. Niisugune liigitus toetab mitmes mõttes ka koolituspraktikat, kus eesmärk ei ole iga üksikkäitumise eraldiseisev kvantifitseerimine, vaid üldise MI kompetentsuse kujunemise jälgimine.

Käitumiskoodide vähene kasutus võib viidata mitmele erinevale nähtusele. Ühelt poolt võib see peegeldada õppurite ettevaatlikkust või ebakindlust keerukamate MI elementide kasutamisel - näiteks keerulised peegeldused või muutusejutu esilekutsumine eeldab lisaks teadmisele ka kogemust ja enesekindlust. Teisalt võib madal esinemissagedus olla ka soovitud tulemus - näiteks kui õppijad väldivad MI-le mitesobivaid käitumisi, nagu otsekohene veenmine (V) või konfrontatsioon (K), mille puudumine võib olla märgiks MI põhimõtete omaksvõtmisest.

Hindajatevahelise kooskõla hindamiseks kasutati klassisisesse korrelatsiooni mudelit, mis sobib olukorda, kus hinnatakse mitme määratletud hindaja keskmist hinnangut ning soovitakse tulemusi üldistada laiemale hindajate populatsioonile. Selline lähenemine on soovitatud vaatluspõhiste andmete usaldusväärsuse hindamisel (Hallgren, 2012) ning on levinud ka MITI süsteemi usaldusväärsusuuringutes (Moyers et al., 2016; Schmidt et al., 2019). See võimaldab hinnata, kui ühtselt suudavad koolituse saanud kodeerijad rakendada hindamisskaalasid reaalsel vestlusmaterjalil. Tulemused viitasid sellele, et enamus skaalasid saavutas hea või väga hea hindajatevahelise kooskõla. See vastab töö kesksele uurimisküsimusele ja kinnitab, et MITI 4.2.1 hindamissüsteem võimaldab simuleeritud, kuid struktureeritud õppeolukorras saavutada hindajatevahelist kooskõla Eesti kontekstis. Kõige ühtlasema kooskõlaga olid skaalad, mis põhinesid konkreetsetel ja hästi määratletavatel käitumistel, näiteks küsimused (K) ja lihtsad peegeldused (PL). Samalaadseid tulemusi on leitud ka varasemates uuringutes, kus nenditi, et struktureeritud, vormilised käitumised on kergemini hinnatavad kui abstraktsemad suhtlemiselemendid (Britt et al., 2022; Schmidt et al., 2019). Seega võib järeldada, et just need skaalaelemendid - näiteks küsimused (K) ja lihtsad peegeldused (PL) - on simuleeritud õppeolukorras usaldusväärsemalt hinnatavad, mis vastab töö teisele uurimisküsimusele (Britt et al., 2022, Schmidt et al., 2019).

Osade globaalskaalade ja keerukamate käitumiste puhul ilmnes suuremat varieeruvust. Näiteks empaatia (E) ja keeruliste peegelduste (PK) skaala puhul oli hindajate kooskõla madalam ning usalduspiirid laiemad, viidates ebaühtlusele hindamistes. Sellised tulemused pole haruldased - mitmed autorid, näiteks Pierson et al. (2006) ja Schmidt et al. (2019), on rõhutanud, et just abstraktsemad või subjektiivsemad kategooriad, nagu empaatia või partnerlus, võivad olla hinnatavad mitmeti, sõltuvalt hindaja kogemusest ja MI mõistmise sügavusest. See viitab, et tehniliselt selgemini piiritletavad skaalad ei pruugi olla sisuliselt kõige tähenduslikumad, kuid nende hindamine on ühtlasem ja kooskõllalisem. Samas võivad keerukamad ja tõlgendusele avatumad kategooriad, nagu empaatia, kujutada hindajatele suuremat väljakutset. Sellisel juhul ei pruugi madalam usaldusväärsus viidata skaala ebatäpsusele, vaid hoopis vajadusele täiendava väljaõppe, jagatud tõlgendusraamistiku või selgemate hindamisjuhiste järele. Seega tuleb MITI rakendamisel arvestada, et kõik skaalad ei pruugi olla võrdselt tundlikud või stabiilsed. Usaldusväärsuse tõlgendamisel tuleb lisaks punktihinnangutele arvesse võtta ka skaala sisulist keerukust, hindamise konteksti ning hindajate treeningtaset.

Üks osa käesolevast analüüsist käsitles üksikhindajate ja eksperthindaja kui hindamisstandardi hinnangute omavahelist seost ja kooskõla. Analüüsimiseks kasutati nii Pearsoni korrelatsiooni kui ka vastavat klassisisese korrelatsiooni mudelit, mis mõõdab hinnangute täpset kattuvust skaala tasandil kahe hindaja vahel. Selle analüüsi eesmärk ei olnud ainult kooskõla hindamine, vaid ka ülevaate saamine, millised hindajad ja skaalad vajasis suuremat ühtlust või potentsiaalset täiendavat juhendamist. Tulemused näitasid, et mitmete skaalade puhul - näiteks küsimused, MI-le mittekohased käitumised (MIN), lihtsad peegeldused (PL) ja peegelduste/küsimuste suhe (P/K) - saavutati tugev kooskõla eksperthindajaga. Need olid üldjoontes samad skaalad, mille puhul oli kõrge hindajatevaheline kooskõla ka grupipõhises analüüsis. Seega võib järeldada, et teatud käitumised - eriti need, mis on selge struktuuriga ja vormiliselt lihtsamini äratuntavad - on mitte ainult hindajate vahel stabiilsemad, vaid ka kergemini joondatavad referentshinnangutega. See vastab kolmandale uurimisküsimusele ning viitab, et hindajate hinnangud eksperthindajaga langevad kõige paremini kokku just nende käitumiste puhul, mis on standardiseeritud ja kergemini määratletavad. ja vormiliselt lihtsamini äratuntavad - on mitte ainult hindajate vahel stabiilsemad, vaid ka kergemini joondatavad referentshinnangutega.

Teisalt ilmnes, et näiteks empaatia (E), samaksjäämisejutt (SJ) ja keerulised peegeldused (PK) olid skaalad, kus hindajate ja eksperdi hinnangute vahelised seosed olid nõrgemad ja ebahütlasemad. Mõnel juhul ei olnud kooskõla statistiliselt oluline, mõnel juhul olid tulemused varieeruvad hindajate lõikes. See toetab varem kirjanduses esitatud järeldusi (Pierson et al., 2006), et just need skaalad nõuavad hindajalt rohkem tõlgendamisoskust, kontekstitaju ning arusaamist MI põhimõtetest. Selliste skaalade hindamine võib sõltuda nii hindaja kogemusest kui ka sellest, kuidas ta mõtestab näiteks empaatia mitteverbaalseid ja emotsionaalseid komponente (Schmidt et al., 2019).

Erinevused hindajate ja eksperdi vahel võivad peegeldada ka individuaalseid hinnangustiile või erinevusi selles, kuidas nad MI hindamise juhiseid rakendasid. Näiteks kui üks hindaja kaldub andma konservatiivsemaid hinnanguid empaatia osas või ei tuvasta keerulist peegeldust, võib see viidata mitte pahatahtlikule veaallikale, vaid hoopis erinevale arusaamale hindamiskriteeriumist. Sarnaseid tähelepanekuid on tehtud ka Britt et al. (2022) uurimuses, kus rõhutati vajadust pideva juhendamise ja ühise hindamispraktika arendamise järele.

Kokkuvõtlikult viitavad analüüsi tulemused sellele, et kuigi MITI rakendamine võimaldab enamikku skaalasad hinnata usaldusväärset ka üksikhindaja tasandil, on osa skaalasad sisuliselt keerukamad ning nende hindamisel võib vaja minna täpsemat juhendamist, supervisiooni või koolituskogemust. Seega on MI hindamises lisaks metoodilisele korrektsusele oluline ka pedagoogiline ja sisuline tugi hindajate arengule.

Andmestiku piirangud ja tulevikusuunad

Käesoleva uuringu tulemuste tõlgendamisel tuleb arvestada mõningate metoodiliste piirangutega. Valimi suurus oli piiratud ning mitmete skaalaelementide varieeruvus andmestikus madal. Sellistes olukordades võib ka väikene erinevus hindajate hinnangutes jätta mulje madalast kooskõlast, kuigi tegelikud väärtused on omavahel sarnased. Eriti kehtib see käitumiste kohta, mis esinesid harva või olid valdavalt nullilähedased.

Hindajate usaldusväärse kujunemisel mängivad rolli nii väljaõpe kui ka võimalus saada tagasisidet ja arutada keerukamaid juhtumeid. Kuigi käesolevas uuringus olid olemas hindamisele juhendamine ja koolitus, rõhutavad mitmed autorid (Britt et al., 2022; Owens et al., 2017), et hindamise usaldusväärset toetab ka jätkuv tagasiside ning arutelu hindamise põhimõtete ja keerukamate skaalade üle. Selline reflektiivne hindamispraktika võib aidata kujundada ühiseid tõlgendusraamistikke ning vähendada erinevusi hindajate lähenemises.

Eksperthindaja kaasamine võimaldas luua võrdluspunkti, mille suhtes hindajate hinnanguid võrrelda. On oluline rõhutada, et eksperthinnang ei tähenda absoluutset tõe, vaid esindab koolitatud, standardiseeritud tõlgendust, mis aitab hinnata, kui võrd hindajad suudavad referentslävendiga joonduda. Nagu Beckman et al. (2022) märgivad, võib isegi kogenud hindajatel olla erinevusi teatud suhtlemiselementide tõlgendamisel, sõltuvalt kontekstist ja hinnatava olukorra nüanssidest. See ei vähenda eksperthinnangu väärtust, vaid peegeldab inimfaktorist tulenevat paratamatut varieeruvust.

Tulevikusuundadena võiks MITI rakendatavust Eestis uurida laiemates ja mitmekesisemates valimites, näiteks kombineerides eri kogemustasemega hindajaid, erinevaid koolituslähenevusi või rakenduskontekste.

KOKKUVÕTE

Käesoleva töö fookuses oli MITI 4.2.1 hindamissüsteemi rakendatavuse ja hindajatevahelise kooskõla analüüs standardiseeritud patsientidega simuleeritud intervjuude põhjal. Tulemused näitasid, et MITI süsteem võimaldab sellises kontekstis saavutada usaldusväärseid hinnanguid, eriti struktuursete ja sagedamini esinevate käitumiskoodide puhul. Globaalskaalad ja keerukamad suhtlemiselemendid osutusid hindamise seisukohalt keerulisemaks ja nõudsid suuremat tõlgendamisoskust. Hindajate ja eksperthindaja hinnangute võrdlus kinnitas, et standardiseeritumad käitumised on stabiilsemalt hinnatavad. Töö järeldab, et MITI 4.2.1 sobib kasutamiseks simuleeritud õppekeskkonnas, kuid hindajate koolitusprotsess vajab täiendavat tähelepanu, et tagada hindamiskriteeriumide ühtlane mõistmine ja rakendamine.

Tänuavaldus

Siirad tänusõnad juhendaja Inga Kartonile toetava suhtumise, kannatlikkuse ja asjatundliku suunamise eest - oli suurepärane kogemus osaleda uuringus, mis on läbi viidud nii inspireeriva pühendumusega. Tänan ka kõiki, kellega ühiselt uuringu praktilist osa ellu viisime. See koostöö ja ühine panustamine andsid minu tööle olulise aluse.

Kasutatud allikad

- Aebersold, M. (2018). Simulation-based learning: No longer a novelty in undergraduate education. *Online Journal of Issues in Nursing, 23*(2).
- Amrhein, P. C., Miller, W. R., Yahne, C. E., Palmer, M., & Fulcher, L. (2003). Client commitment language during motivational interviewing predicts drug use outcomes. *Journal of consulting and clinical psychology, 71*(5), 862.
- Atkins, S., Roberts, C., Hawthorne, K., & Greenhalgh, T. (2016). Simulated consultations: a sociolinguistic perspective. *BMC Medical Education, 16*, 1–9.
- Barrows, H. S. (1993). An overview of the uses of standardized patients for teaching and evaluating clinical skills. *Academic Medicine, 68*(6), 443–451.
- Bard, A., Forsberg, L., Wickström, H., Emanuelson, U., Reyher, K., & Svensson, C. (2023). Clinician motivational interviewing skills in ‘simulated’ and ‘real-life’ consultations differ and show predictive validity for ‘real life’ client change talk under differing integrity thresholds. *PeerJ, 11*, e14634.
- Beckman, M., Lindqvist, H., Öhman, L., Forsberg, L., Lundgren, T., & Ghaderi, A. (2022). Correspondence between practitioners’ self-assessment and independent motivational interviewing treatment integrity ratings. *Frontiers in Psychology, 13*, 890579.
- Björk, A. (2014). Stabilizing a fluid intervention: The development of Motivational Interviewing, 1983–2013. *Addiction research & theory, 22*(4), 313-324.
- Britt, E., Soleymani, S., Wallace-Bell, M., & Garland, A. (2023). Motivational interviewing for employment: An exploration of practitioner skill and client change talk. *Journal of Employment Counseling, 60*(1), 42-59.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.
- Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry, 11*(4), 227–268.
- Decker, S., Jeffries, P. R., Elliott, M. M., & Goldman, N. (2013). Standards of best practice: Simulation standard VI: The debriefing process. *Clinical Simulation in Nursing, 9*(6), S26–S29.
- Forsberg, L., Källmén, H., Hermansson, U., Berman, A. H., & Helgason, Á. R. (2007). Coding counsellor behaviour in motivational interviewing sessions: Inter-rater reliability for the Swedish Motivational Interviewing Treatment Integrity code (MITI). *Cognitive Behaviour Therapy, 36*(3), 162–169.
- Fu, S. S., Roth, C., Battaglia, C. T., Nelson, D. B., Farmer, M. M., Do, T., ... & Zillich, A. J. (2015). Training primary care clinicians in motivational interviewing: a comparison of two models. *Patient Education and Counseling, 98*(1), 61-68.

- Galal, S., Vyas, D., Mayberry, J., Rogan, E. L., Patel, S., & Ruda, S. (2018). Use of standardized patient simulations to assess impact of motivational interviewing training on social-emotional development. *Pharmacy*, 6(3), 65.
- Gill, I., Oster, C., & Lawn, S. (2020). Assessing competence in health professionals' use of motivational interviewing: A systematic review of training and supervision tools. *Patient education and counseling*, 103(3), 473-483.
- Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology*, 8(1), 23-34.
<https://doi.org/10.20982/tqmp.08.1.p023>
- Hamman, W. R. (2004). Will simulation fly in medicine as it has in aviation? *Quality & Safety in Health Care*, 13(5), 397-399.
- Hohman, M. (2021). *Motivational interviewing in social work practice*. Guilford Publications.
- Hurlocker, M. C., Madson, M. B., & Schumacher, J. A. (2020). Motivational interviewing quality assurance: A systematic review of assessment tools across research contexts. *Clinical Psychology Review*, 82, 101909.
- Jürjen, T. (2013). *Motiveeriva intervjuerimise hindamissüsteemi 3.1.1 eestindamine* (magistritöö). Tartu Ülikool.
- Kelly, M. A., Forber, J., Conlon, L., Roche, M., & Stasa, H. (2014). Empowering the registered nurses of tomorrow: Students' perspectives of a simulation experience for recognising and managing a deteriorating patient. *Nurse Education Today*, 34(5), 724-729.
- Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2), 155-163.
<https://doi.org/10.1016/j.jcm.2016.02.012>
- Lane, C., Huws-Thomas, M., Hood, K., Rollnick, S., Edwards, K., & Robling, M. (2005). Measuring adaptations of motivational interviewing: the development and validation of the behavior change counseling index (BECCI). *Patient Education and Counseling*, 56(2), 166-173.
- Lane, C., Hood, K., & Rollnick, S. (2008). Teaching motivational interviewing: using role play is as effective as using simulated patients. *Medical Education*, 42(6), 637-644.
- Lewis, K. L., Bohnert, C. A., Gammon, W. L., Hölzer, H., Lyman, L., Smith, C., ... & Gliva-McConvey, G. (2017). The association of standardized patient educators (ASPE) standards of best practice (SOBP). *Advances in Simulation*, 2, 1-8.
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley. Vaadatud:2025
<https://ia601405.us.archive.org/32/items/in.ernet.dli.2015.139135/2015.139135.Statistical-Theories-Of-Mental-Test-Scores.pdf>

- Luik, K. (2023). Standardiseeritud patsiendi rakendamise mõju motiveerivast intervjuerimisest lähtuva nõustamismeetodi alase väljaõppe tulemuslikkusele. *Uurimistöö, TÜ*.
- Lundahl, B., & Burke, B. L. (2009). The effectiveness and applicability of motivational interviewing: A practice-friendly review of four meta-analyses. *Journal of clinical psychology, 65*(11), 1232-1245.
- Lundahl, B. W., Kunz, C., Brownell, C., Tollefson, D., & Burke, B. L. (2010). A meta-analysis of motivational interviewing: Twenty-five years of empirical studies. *Research on social work practice, 20*(2), 137-160.
- Madson, M. B., Loignon, A. C., & Lane, C. (2009). Training in motivational interviewing: A systematic review. *Journal of substance abuse treatment, 36*(1), 101-109.
- Martino, S., Haeseler, F., Belitsky, R., Pantalon, M., & Fortin, A. H. (2007). Teaching brief motivational interviewing to Year three medical students. *Medical Education, 41*(2), 160–167.
- McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods, 1*(1), 30–46. <https://doi.org/10.1037/1082-989X.1.1.30>
- McMaster, F., & Resnicow, K. (2012). Motivational interviewing: Moving from why to how with autonomy support. *International Journal of Behavioral Nutrition and Physical Activity, 9*, 19.
- Miller, W. R., & Rollnick, S. (2002). Motivational interviewing: Preparing people for change. Book Review. Vaadatud 2025: <https://bluepeteraustralia.wordpress.com/wp-content/uploads/2012/12/motivational-interviewing.pdf>
- Miller, W. R., & Rose, G. S. (2009). Toward a theory of motivational interviewing. *American psychologist, 64*(6), 527.
- Moyers, T. B., Martin, T., Manuel, J. K., Hendrickson, S. M., & Miller, W. R. (2005). Assessing competence in the use of motivational interviewing. *Journal of Substance Abuse Treatment, 28*(1), 19–26
- Moyers, T. B., Manuel, J. K., & Ernst, D. (2014). Motivational interviewing treatment integrity coding manual 4.1. *Unpublished manual, 1*, 3.
- Moyers, T. B., Manuel, J. K., & Ernst, D. (2015). Motivational Interviewing Treatment Integrity Coding Manual 4.2.1 (MITI 4.2.1). *Unpublished manual, June*.
- Moyers, T. B., Rowell, L. N., Manuel, J. K., Ernst, D., & Houck, J. M. (2016). The Motivational Interviewing Treatment Integrity Code (MITI 4): Rationale, preliminary reliability and validity. *Journal of Substance Abuse Treatment, 65*, 36–42.

- Oh, P. J., Jeon, K. D., & Koh, M. S. (2015). The effects of simulation-based learning using standardized patients in nursing students: A meta-analysis. *Nurse Education Today*, *35*(5), e6–e15.
- Owens, M. D., Rowell, L. N., & Moyers, T. (2017). Psychometric properties of the motivational interviewing treatment integrity coding system 4.2 with jail inmates. *Addictive Behaviors*, *73*, 48–52.
- Pierson, H. M., Hayes, S. C., Gifford, E. V., Roget, N., Padilla, M., Bissett, R., ... & Fisher, G. (2007). An examination of the motivational interviewing treatment integrity code. *Journal of Substance Abuse Treatment*, *32*(1), 11–17.
- Rutherford-Hemming, T., & Jennrich, J. A. (2013). Using standardized patients to strengthen nurse practitioner competency in the clinical setting. *Nursing Education Perspectives*, *34*(2), 118–121.
- Sackett, D. L., Rosenberg, W. M., Gray, J. M., Haynes, R. B., & Richardson, W. S. (1996). Evidence based medicine: what it is and what it isn't. *BMJ*, *312*(7023), 71–72.
- Schmidt, L. K., Andersen, K., Nielsen, A. S., & Moyers, T. B. (2019). Lessons learned from measuring fidelity with the Motivational Interviewing Treatment Integrity code (MITI 4). *Journal of Substance Abuse Treatment*, *97*, 59–67.
- Schmidt, L. K., Moyers, T. B., Nielsen, A. S., & Andersen, K. (2019). Is fidelity to motivational interviewing associated with alcohol outcomes in treatment-seeking 60+ year-old citizens? *Journal of Substance Abuse Treatment*, *101*, 1–11.
- Schwalbe, C. S., Oh, H. Y., & Zweben, A. (2014). Sustaining motivational interviewing: A meta-analysis of training studies. *Addiction*, *109*(8), 1287–1294.
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological bulletin*, *86*(2), 420.
- Self, K. J., Borsari, B., Ladd, B. O., Nicolas, G., Gibson, C. J., Jackson, K., & Manuel, J. K. (2023). Cultural adaptations of motivational interviewing: A systematic review. *Psychological Services*, *20*(S1), 7.
- Stokoe, E. (2013). The (in)authenticity of simulated talk: Comparing role-played and actual interaction and the implications for communication training. *Research on Language and Social Interaction*, *46*(2), 165–185.
- Svensson, C., Wickström, H., Emanuelson, U., Bard, A. M., Forsberg, L., & Reyher, K. K. (2020). Training in motivational interviewing improves cattle veterinarians' communication skills for herd health management. *Veterinary Record*, *187*(5), 191.
- Ten Hove, R., Brabers, A. E., & Arntz, A. (2022). Guidelines for calculating and reporting interrater reliability for behavioral coding research in psychology. *Journal of Behavior Therapy and Experimental Psychiatry*, *75*, 101719.

Trickey, A. W., Newcomb, A. B., Porrey, M., Wright, J., Bayless, J., Piscitani, F., ... & Dort, J. (2016). Assessment of surgery residents' interpersonal communication skills: validation evidence for the communication assessment tool in a simulation environment. *Journal of Surgical Education*, 73(6), e19–e27.

Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina, Terje Saunoris,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose

„MITI 4.2.1 hindamissüsteemi rakendatavus

ja hindajatevaheline kooskõla simuleeritud tingimustes“

mille juhendaja on Inga Karton,

reprodutseerimiseks eesmärgiga seda säilitada, sealhulgas lisada digitaalarhiivi DSpace kuni autoriõiguse kehtivuse lõppemiseni.

2. annan Tartu Ülikoolile loa teha punktis 1 nimetatud teos üldsusele kättesaadavaks Tartu Ülikooli veebikeskkonna, sealhulgas digitaalarhiivi DSpace kaudu Creative Commons'i litsentsiga CC BY NC ND 4.0, mis lubab autorile viidates teost reprodutseerida, levitada ja üldsusele suunata ning keelab luua tuletatud teost ja kasutada teost ärieesmärgil, kuni autoriõiguse kehtivuse lõppemiseni.

3. olen teadlik, et punktides 1 ja 2 nimetatud õigused jäävad alles ka autorile.

4. kinnitan, et lihtlitsentsi andmisega ei riku ma teiste isikute intellektuaalomandi ega isikuandmete kaitse õigusaktidest tulenevaid õigusi.

Terje Saunoris

15.05.2025