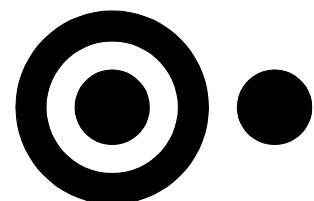


Att mäta ett mått för motiverande samtal.

Bristande interbedömarreliabilitet i MITI 4.2.1.

Av: Ludvig Rylander

Handledare: Per Henrik Hedberg och Lars Forsberg
Södertörns högskola | Institutionen för samhällsvetenskaper
Kandidatuppsats 15 hp
Psykologi | vårterminen 2016



SÖDERTÖRNS HÖGSKOLA | STOCKHOLM
sh.se

ATT MÄTA ETT MÅTT FÖR MOTIVERANDE SAMTAL: BRISTANDE INTERBEDÖMARRELIABILITET I MITI 4.2.1.

Ludvig Rylander

Sammanfattning

Motiverande samtal (MI) är en populär samtalsmetod med fokus på motivations- och förändringsarbete. MI utgår från ett teoretiskt underbyggt förhållningssätt där rådgivaren och klienten tillsammans utforskar möjliga vägar till förändring. För att säkerställa att MI används på bästa sätt kan samtalsledarens skicklighet och förmåga observeras koda enligt olika variabler. Dessa mätningar bör i sin tur besitta en hög interbedömarreliabilitet (IBR) för att ge tillförlitliga resultat. Absolut överensstämmelse i samtliga variabler ger maximal IBR, medan en låg IBR kan tyda på att mätinstrumentet är otydligt, att det finns brister hos bedömare, eller på en kombination av detta. I den här uppsatsen undersöks IBR i The Motivational Interviewing Treatment Integrity Code (MITI), ett kodningsverktyg som tagits fram för att vara reliabelt och lätt att använda. Samtalssekvenser om tjugo minuter koda enligt fjorton variabler. Fyra oberoende kodare har kodat tolv samtal enligt den senaste versionen av mätinstrumentet (MITI 4.2.1). Samtliga kodningar genomfördes inom ramen för en kodarutbildning som hålls av ett professionellt kodningslab. Intraklasskorrelation (ICC), Krippendorffs α (Kalpha) samt Cronbachs α redovisas parallellt som mått på IBR. Flera variabler visar en otillfredsställande eller, beroende på vilken koefficient som används, oacceptabel IBR. De fyra kodarna visar en lägre samstämmighet i bedömningen av kvalitativa variabler samt variabler avseende beteenden ej förenliga med MI. ICC för fyra kodare ligger generellt på en måttlig men godkänd nivå, medan Kalpha genomgående ligger under gränsen för det godtagbara. Cronbachs α ligger överlag något över ICC vilket tyder på en viss systematik i kodarnas bristande överensstämmelse. En tänkbar orsak till en bitvis sviktande IBR är att de undersökta kodarna vid mättillfället var relativt oerfarna och skulle behövt mer praktisk träning i MITI 4.2.1. Vidare kan en bristande förståelse för instrumentets teoretiska grunder ha inverkat negativt på IBR. För att MITI ska kunna utvärderas korrekt i framtida forskning krävs ett enhetligt grepp för hur IBR ska mätas. En högre förståelse för metoderna bakom mätning av IBR är önskvärd.

Att förändra vardagsbeteenden är ingen lätt uppgift. Ändå är just beteendeförändringar något som upptar en stor del av den moderna människans tid och energi. Kostvanor, motion, studieteknik, passivitet – överallt i vardagen finns det vanor som gnager på samvetet och pockar på förändring. För en normalfungerande vuxen människa inom samhällets normer finns det goda förutsättningar att faktiskt lyckas med detta, givet att målen är realistiska och

att arbetsinsatsen stämmer någorlunda överens med fördelarna som motiverar till förändringen. Alla människor har dock inte samma förutsättningar att bryta negativa och destruktiva beteendemönster. För den som söker professionell hjälp med detta finns en uppsjö av terapier och samtalsmetoder att välja bland. Motiverande samtal har under de senaste decennierna kommit att positioneras som en allsidig behandlingsmetod som tycks passa in i de allra flesta sammanhang.

Motiverande samtal (i fortsättningen förkortat MI, av *motivational interviewing*) utvecklades av William R. Miller under början av 1980-talet. Syftet var från början att underlätta motivations- och förändringsarbete i behandling av alkoholmissbruk men metoden har sedan dess kommit att användas som aktiv komponent i ett brett spektrum av förändringsarbete. Utöver missbruksbehandling används MI även för att stödja hälsofrämjande beteenden och dämpa riskbeteenden. Rådgivning om kost- och motionsvanor, spelmissbruk, ätstörningar, social träning, föräldrarådgivning och relationsproblem är andra exempel på mer eller mindre välbeprövade områden (Lundahl & Burke, 2009). I de flesta sammanhang där det är aktuellt att locka fram klienters motivation till ändrat beteende går det att finna en tillämpning för MI. Därmed inte sagt att tillräcklig effekt kan förväntas i alla sammanhang. Tillämpningsområden, effekter och metodiska överväganden behandlas grundligt i fyra omfattande metaanalyser (Burke, Arkowitz & Menchola, 2003; Hetteima, Steele & Miller, 2005; Vasilaki, Hosier & Cox, 2006; Lundahl, Kunz, Brownell, Tollefson & Burke, 2010). De tre generella modellerna för tillämpning av MI inom klinisk behandling är a) som fristående intervention där samtalsessionerna utgör själva behandlingen; b) som additiv komponent, integrerad i annan specifik behandling; eller c) som inledande eller uppföljande intervention i samband med annan specifik behandling (Lundahl et al., 2010).

När MI implementeras i någon form av klinisk verksamhet förväntas det ha en viss specifik effekt på utfallet av behandlingen. Så länge det går att säkerställa att det verkligen är MI som praktiserats är det rimligt att anta att eventuella förändringar i utfallet har med detta att göra (och, omvänt, att utebliven förändring tyder på att metoden inte har någon betydelse för utfallet). I verkligheten är det dock sällan så enkelt: olika individer kan bära på olika fördomar om MI som påverkar utförandet, olika yrkesroller kan ha skilda förhållningssätt till MI, utbildningen kan vara bristfällig eller ojämnt fördelad etc. Det finns helt enkelt många faktorer som vid en närmare granskning kan visa sig ställa det från början avsedda på ända. Motsvarande gäller förstås i vetenskapliga studier om effekter av MI. För att kunna dra slutsatser om vad en uppmätt effektskillnad beror på måste det på något vis gå att fastställa närvaron eller frånvaron av MI. En vetenskaplig undersökning som inte säkerställer detta riskerar att antingen överskatta eller underskatta effekterna av MI.

Hur kan då en forskare eller verksamhetsansvarig gå tillväga för att få klarhet i om uppmätta eller observerade effekter beror på MI eller något annat? För det första behöver behandlingstroheten säkerställas, dvs. att det verkligen är MI som har praktiserats. För det andra måste denna mätning vara tillförlitlig och ge otvetydiga resultat. För att fastställa behandlingstrohet finns det en rad olika mätinstrument att tillgå. Vissa av dem går på djupet och är mycket omfattande medan andra är mer ytliga. Ett mätinstrument som ofta används inom både vetenskapliga studier och i klinisk verksamhet är Motivational Interviewing Treatment Integrity Code (MITI), ett kodningsverktyg som mäter i vilken utsträckning professionella samtalsledare (terapeuter, läkare, rådgivare etc.) använder sig av MI-specifika komponenter och färdigheter. Ett vanligt sätt att bedöma tillförlitligheten i dessa mätningar är att beräkna interbedömarreliabiliteten, dvs. i vilken mån två eller flera oberoende bedömare kommer fram till samstämmiga resultat.

Den här uppsatsen behandlar interbedömarreliabiliteten (IBR) hos MITI 4.2.1 som är den senaste versionen av mätinstrumentet. MITI 4.2.1 har under hösten 2015 översatts till svenska och det här är den första studien av interbedömarreliabiliteten i denna version. För läsare som inte är bekanta med MI och dess teorier följer här en genomgång som syftar till att öka förståelsen för MITI och dess variabler.

MI i praktiken

Den så kallade MI-andan, som kan ses som grundläggande för allt MI-utövande, innebär att terapeuten har ett stödjande och empatiskt förhållningssätt och att det är klienten som vet mest om sina egna behov och möjligheter till förändring. Genom att arbeta med öppna frågor, bekräftelser och reflektioner kan samtalsledaren stärka klientens självkänsla och locka fram klientens egna ord om förändring, så kallat förändringstal. Det är alltså inte samtalsledarens uppgift att tala om vad klienten bör göra, denna kunskap finns redan hos klienten. Miller och Rollnick (2013) inspirerades av Daryl Bems (1967) teori om självperception, som i korthet innebär att människan påverkas mer av det hon hör sig själv säga än det någon annan säger åt henne. Enligt detta sätt att se kommer allt som klienten själv uttrycker att ha starkare påverkan än uppmaningar eller välvilliga råd från någon annan. Tillsammans med Carl Rogers (1957) idéer om acceptans, autonomi och absolut värde utgör detta grunden för MI-andan.

I praktiken inleds MI med att klient och samtalsledare tillsammans identifierar ett mål-beteende, dvs. ett realistiskt mål för den förändring som samtalet sedan kommer att kretsa kring. Miller och Rollnick (2013) beskriver det praktiska MI-förfarandet i fyra grundläggande processer: engagerande, fokuserande, framkallande och planerande. I den *engagerande processen* byggs förtroende och tillit upp. Samtalsledarens roll är då att genom aktivt lyssnande och empatiska bekräftelser skapa en varm och accepterande grund för det fortsatta samtalet. Denna process pågår tills klienten är redo att gå vidare till den *fokuserande processen* där problemen formuleras tydligare och möjliga vägar till förändring tas upp. Här kan det bli aktuellt för samtalsledaren att ge information om olika alternativ. Det är dock viktigt att detta sker i dialog med klienten och inte förmedlas med expertens auktoritet. I detta skede handlar samtalet även om för- och nackdelar med status quo, dvs. att avstå från förändring. Klienten bär ofta på en ambivalens kring status quo kontra förändring, och denna ambivalens utforskas mer specifikt i den *framkallande processen*. Samtalsledaren försöker här förstärka dissonansen mellan förändring och status quo för att på så sätt framkalla förändringstal hos klienten. Denna beslutsbalans är dock inte nödvändig för praktiserandet av MI. I samtal med väldigt ambivalenta klienter kan den till och med motverka sitt syfte medan klienter som känner sig mer redo för en förändring kan stärkas i sitt beslut (Miller & Rose, 2013).

Den framkallande processen är ofta ganska krävande och det kan ibland vara nödvändigt att backa till tidigare skeden för att komma vidare till *planeringsprocessen* som handlar om att komma överens om en realistisk plan för de beslut och åtaganden som uppkommit under samtalsessionerna. Detta förlopp kan sträcka sig över flera samtalsessioner eller utspela sig inom ett enda samtal. Hur mycket tid som krävs beror på problemets art och samtalsledarens empatiska förmåga.

Förloppet i de ovan nämnda processerna har i tidigare utgåvor av Miller och Rollnicks (2013) standardverk om MI beskrivits som två faser, en motivationsbyggande fas och en konsoliderande fas. I den tredje utgåvan diskuteras detta ordval mot bakgrund av svårigheter som

uppstått i praktiserandet av MI. Motivationsarbete följer sällan ett linjärt förlopp där nödvändiga faser passerar i tur och ordning enligt ett från början givet mönster. Att i stället se MI som olika processer implicerar att dessa kan löpa parallellt och att samtalet kan röra sig mellan dem, utan att det skulle innebära någon förlust eller stagnation. Processerna antas visserligen uppträda i en given ordning, men de ingår samtidigt i ett flöde där de överlappar och går in i varandra. Miller och Rollnick (2013) beskriver de fyra processerna som en trappa där det är möjligt att röra sig både uppåt och nedåt.

MI:s teoretiska grunder

Principerna bakom MI inorporerar flera olika teorier om samarbete och förändringsprocesser. Liksom Rogers klientcentrerade terapi har den transteoretiska modellens *stages of change* ofta kopplats till MI (Miller & Rollnick, 2009). Till skillnad från traditionella teorier om terapi förutsätter den transteoretiska modellen inte att klienten kommer väl förberedd till behandling. Många klienter kan i stället vara starkt ambivalenta och inte alls redo att påbörja en krävande förändringsprocess (Prochaska & DiClemente, 1982). Därför är terapeutens första uppgift att tillsammans med klienten utforska vilka förutsättningar som kan ligga till grund för den fortsatta processen, en idé som stämmer väl överens med MI:s grundprinciper. Miller och Rollnick (2009) påpekar dock att medan den transteoretiska modellen syftar till att beskriva förändringsprocesser generellt, är MI en tillämpad samtalsmetod vars specifika teori fortfarande håller på att kartläggas. Modeller och teorier som Carl Rogers klientcentrering, den transteoretiska modellen och Bems teori om självperception är alltså inte synonyma med MI, utan utgör snarare ett teoretiskt ramverk inom vilket olika delar av MI kan förstås och diskuteras (Miller & Rollnick, 2009). Det är dock inte nödvändigt att känna till dessa teorier för att kunna tillämpa MI. I sökandet efter en avgränsad och specifik teori för hur och varför MI fungerar har Miller och Rose (2009) formulerat två hypoteser; den relationella hypotesen och den tekniska hypotesen. Dessa antaganden är viktiga för förståelsen av de grundläggande komponenter som mäts och utvärderas i MITI.

Den relationella hypotesen: empati och partnerskap. MI vilar på ett underliggande förhållningssätt som inbegriper acceptans, empati, samarbete och framkallande av motivation. Dessa relationella faktorer utgör det som Miller och Rollnick kallar MI-andan och har inte så mycket med specifika tekniker att göra. Enligt den relationella hypotesen prediceras förändring av terapeutens förmåga att skapa varma och empatiska relationer. Flera studier har visat resultat som ger stöd för den relationella hypotesen (Miller & Rose, 2009).

MI-andan kan ses som en förlängning av Carl Rogers klientcentrerade rådgivning (Miller & Rollnick, 2013). Rogers (1957) formulerade sex nödvändiga villkor för personlig förändring genom en klientcentrerad terapeutisk relation: 1) två personer står i kontakt med varandra; 2) den ena personen (klienten) upplever någon form av oro, sårbarhet eller ambivalens; 3) den andra personen (terapeuten) deltar i relationen som en ärlig och genuint involverad medmänniska och visar inte upp någon yttre fasad; 4) terapeuten erfar en varm och ovillkorlig acceptans av klienten; 5) terapeuten erfar en empatisk förståelse för klientens inre referensram; 6) klienten uppfattar att villkor 4) och 5) är uppfyllda. Rogers menade att dessa villkor inte bara är nödvändiga utan även tillräckliga. Ur detta formulerade han hypotesen att om a) de sex villkoren är uppfyllda kommer b) en konstruktiv förändring av klientens personlighet att komma till stånd ur en den terapeutiska relationen (Rogers, 1957).

Även om Miller och Rollnick (2009) ser MI-andan som sprungen ur Rogers arbete finns det uppenbara skillnader mellan metoderna. I traditionell klientcentrerad rådgivning är det

klienten som styr samtalet och riktningen avgörs främst av klientens känslor. MI är en metod där terapeuten frammanar en strategisk rörelse och på logisk grund styr samtalet mot specifika mål. Utsagor som uppfattas som stärkande i förändringsprocessen fångas upp och bekräftas medan utsagor som strävar mot status quo dämpas. Det här sättet att styra samtalet ter sig helt främmande inom klientcentrerad rådgivning (Miller & Rollnick, 2009).

Den tekniska hypotesen: förändringstal och bibehållandetal. MI har starka teoretiska kopplingar till såväl Daryl Bems teori om självperception som till Festingers teori om kognitiv dissonans (Miller & Rose, 2009). Enligt Bem blir individen medveten om sina attityder genom att observera sig själv utifrån. Genom att beskriva sig själv och sina beteenden tillåts individen upptäcka möjligheter och förmågor som tidigare legat dolda. En annan aspekt av självbetraktandet är den kognitiva dissonansen, dvs. skillnaden mellan det klienten *vill* göra och det klienten *verkligen* gör. Att sätta ord på denna skillnad kan vara ett sätt att upptäcka nya motiv till, eller avstå från, förändring. Som tidigare nämnts är detta inte en nödvändig komponent för MI, men den är likväl starkt kopplad till den medvetandegörande processen. I ett terapeutiskt sammanhang kan teorierna om självperception och kognitiv dissonans användas för att beskriva funktionen hos olika samtalsbeteenden, dvs. vad som sägs (eller inte sägs), på vilket sätt det uttrycks och hur samspelet mellan klient och terapeut ser ut.

Förståelsen av förändringstalets mekanismer är centralt för MI, såväl i praktiken som i teorin. Enligt den tekniska hypotesen prediceras positiva beteendeförändringar av terapeutens förmåga att framkalla förändringstal hos klienten. Denna hypotes förutsätter två viktiga samband: a) att användandet MI-specifika tekniker verkligen leder till förändringstal och b) att förändringstal verkligen leder till positiv förändring av ett visst beteende. Glynn och Moyers (2010) fann stöd för det första sambandet genom att jämföra hur förändringstal frammanades i MI-samtal respektive KBT-samtal. Sambandet har tidigare undersökts i ett flertal andra studier, men Glynn och Moyers var först med att göra det i en strikt experimentell design (Miller & Rose, 2009).

Det andra sambandet, huruvida förändringstal ger mätbara och beständiga effekter på mål-beteenden, har det varit svårare att finna evidens för (Miller & Rose, 2009). Baserat på sin tidigare forskning om språkliga kategorier presenterade psykologvisten Paul Amrhein, tillsammans med Miller, Yahne, Palmer och Fulcher (2003), ett nytt sätt att analysera och tolka förändringstal. I MI-samtal med 84 klienter i missbruksbehandling differentierades förändringyttranden i kategorierna *åtagande*, *önskan*, *förmåga*, *behov*, *beredskap* och *skäl*. Genom att mäta frekvens och styrka över tid hos dessa kategorier visade det sig att *åtagande* predicerades av de övriga kategorierna samt att *åtagande* var det enda som predicerade minskad droganvändning upp till ett år efter påbörjad behandling. Forskarna rangordnade kategorierna efter styrka och kunde dra slutsatsen att positiva beteendeförändringar kan prediceras av att de svagare kategorierna följs av den starkare (Amrhein et al., 2003). Mängden förändringstal är alltså mindre relevant för utfallet än styrkan i detsamma; styrkan i förändringstalet har större effekt mot slutet av samtalet (Miller & Rose, 2009).

Rollnick och Miller (1995) poängterar att det är MI-andan, snarare än specifika samtals-tekniker, som är grundläggande för MI. Eftersom ingen samtalsituation är den andra lik är variationen stor i hur enskilda terapeuter väljer att använda sig av olika tekniker. MI-andan är mer beständig och genomgående. För att de tekniska komponenterna ska vara förenliga med MI behöver de förstås och tillämpas i enlighet med MI-andan (Rollnick & Miller, 1995).

När är det MI?

I all forskning där slutsatser dras kring tillämpningar och effekter av MI är det nödvändigt att kunna säkerställa i vilken utsträckning och med vilken kvalitet metoden verkligen har använts, det som i kliniska sammanhang brukar kallas behandlingstrohet (Miller & Rollnick, 2014). Miller och Rollnick har formulerat tre enligt dem grundläggande villkor för behandlingstrohet i vetenskapliga studier. För det första krävs utvecklade MI-färdigheter som används i syfte att locka fram förändringstal relaterat till ett eller flera väl definierade förändringsmål. För det andra ska dessa färdigheter dokumenteras och specificeras enligt på förhand angivna kriterier. För det tredje är det nödvändigt att reliabelt mäta och kvalitetssäkra interventionen som levereras i studien. När en studie uppfyller dessa villkor kan dess resultat, positivt eller negativt, med större säkerhet härledas till att det verkligen var en effekt av MI som studerades (Miller & Rollnick, 2014).

Viktiga frågor om reliabilitet och validitet kan besvaras genom att noggrant koda och mäta samtalsledarens och/eller klientens samtalsbeteenden (Jelsma, Mertens, Forsberg & Forsberg, 2015). För att kunna göra reliabla mätningar har flera olika mätinstrument tagits fram. Dessa fyller delvis olika syften men överlappar ofta varandra (Wallace & Turner, 2009). Valet av mätinstrument bör förstås styras av forskningsfrågan. Är det klientens responser eller samtalsledarens färdigheter som ska undersökas? Är det MI som helhet eller en specifik komponent som är av intresse? Handlar det om att predicera utfall eller att analysera samtalsinteraktioner? Vissa frågeställningar berör enbart MI i sig medan andra kan kräva analys av andra faktorer som påverkar (eller påverkas av) MI (Jelsma et al., 2015; Madson & Campbell, 2006).

I klinisk praktik kan det vara önskvärt att snabbt och enkelt ta reda på vilken MI-kompetens som finns i verksamheten för kunna planera eventuella utbildningsinsatser. Det enklaste sättet är förstås att gå direkt till källan, dvs. att fråga de berörda terapeuterna. Detta görs vanligen med ett någon typ av formulär där terapeuten själv skattar sin förståelse för MI:s grundläggande idéer och principer, sin förmåga att tillämpa MI-specifika tekniker och i vilken utsträckning detta görs i det dagliga arbetet (Wain, Kutner, Smith, Carpenter, Hu, Amrhein & Nunes, 2015). Flera studier har dock visat att data inhämtad genom självrapportering korrelerar dåligt med data från objektiva mätmetoder baserade på observation (Wain et al., 2015; Miller & Mount, 2001; Miller et al., 2004).

Trots en mängd väl beprövade instrument saknas konsensus kring vilket som är det ultimata för bedömning av behandlingstrohet till MI (Wallace & Turner, 2009; Madson & Campbell, 2006). McMaster och Resnicow (2015) hänvisar visserligen till MITI som ”gold standard” för mätning av MI-specifika färdigheter, men det påståendet tycks sakna belägg. Möjligen kommer sig denna missuppfattning av att MISC (som ligger till grund för MITI) är det mest grundliga mätinstrumentet, eller så är det bara en förväxling med de professionella kodningslabbens gold standard för IBR. Däremot stämmer det att MITI fått stor spridning och använts i såväl kliniska studier som i utbildning och handledning av nya samtalsledare (McMaster & Resnicow, 2015; Miller & Rollnick, 2014; Moyers, Rowell, Manuel, Ernst & Houck, 2016).

MITI

MITI utvecklades som en enklare och mer ekonomisk variant av det betydligt mer omfattande MISC (Moyers, Martin, Manuel, Hendrickson & Miller, 2005). Syftet var att ta fram ett instrument som snabbt och reliabelt fastställer i vilken utsträckning och med vilken kvalitet terapeuten tillämpar förhållningssätt och tekniker förenliga med MI. En explorativ

faktoranalys visade att de 33 variablerna i MISC fördelade sig på 11 faktorer med egenvärden större än 1. Dessa faktorer reduceras sedan till 9 variabler i den första versionen av MITI. Trots denna drastiska minskning av antalet variabler visade en kanonisk korrelationsanalys att den första versionen av MITI fångade 59% av variansen i MISC. En viktig skillnad mellan instrumenten är att medan MISC mäter variabler relaterade till både klients och terapeutens samtalsbeteenden, mäter MITI endast terapeutens kliniska MI-färdigheter. I undersökningar av kausala samband mellan MI och utfall av behandling krävs att både terapeutens och klientens yttranden analyseras (Moyers et al., 2005). MITI 4.2.1 (bilaga A) innehåller totalt 20 variabler: 4 övergripande skattningar (kvalitativa variabler), 10 beteendefrekvensräkningar (kvantitativa variabler), och 6 summerande variabler (Moyers, Manuel & Ernst, 2014). Indexberäkningarna anges som medelvärden för de övergripande skattningarna och proportioner eller absoluta tal för beteendefrekvenserna. Indexberäkningarna är avsedda att jämföras med gränsvärden för godkänd eller väl godkänd MI-kompetens. Dessa gränsvärden är dock inte validerade, utan baseras på expertutlåtanden (Moyers et al., 2014).

De övergripande skattningarna är indelade i *tekniska komponenter* (*främja förändringstal* och *dämpa bibehållandetal*) och *relationskomponenter* (*partnerskap* och *empati*) vilket tydligt speglar Miller och Rose (2009) teoretiska beskrivning av MI. Skattningarna görs enligt femgradiga likertskalor där utförliga kriterier anges för varje skalsteg (bilaga B). Dessa variabler är alltså uppenbart kvalitativa och syftet med dem är att fånga helhetsintrycket av den kodade samtalssekvensen. Vid kodningssessionens början utgår kodaren från mitten av skalan (dvs. 3) för att sedan justera skattningen enligt angivna kriterier.

Beteendefrekvensräkningarna görs oberoende av de övergripande skattningarna och mäter kvantitativt förekomsten av specifika samtalsbeteenden. Dessa variabler är inte öppna för kodarens tolkning utan sker enligt de beslutsregler som anges i manualen. Terapeutens tal delas upp i ordströmmar och yttranden. En ordström kan ses som en längre sammanhållen mening som kan bestå av flera yttranden. Ordströmmen pågår tills klienten tar till orda, undantaget småord som inte formar ett yttrande, så kallat utfyllnadstal. Ett yttrande definieras som en fullständig tanke eller idé som kan kodas enligt någon av de tio beteendekategorierna. Varje yttrande kan ges endast en beteendekod och varje specifik beteendekategori kan förekomma endast en gång i varje ordström. Om tre frågor ingår i samma ordström eller yttrande ges alltså beteendekoden *fråga* endast en gång. Manualen anger även detaljerade beslutsregler för hur kodaren ska prioritera när alternativa kodningar är tänkbara (Moyers et al., 2014).

Sedan den första officiella versionen har MITI genomgått två större uppdateringar (Moyers, Martin, Manuel & Miller, 2003; Moyers, Martin, Manuel, Miller & Ernst, 2010; Moyers et al., 2014). I och med MITI 3.0 omformulerades de övergripande skattningarna från *empati* och *MI-anda* till *empati*, *styrning*, *locka fram*, *samarbetsfrämjande* och *autonomistöd*, där medelvärdet hos de tre senare variablerna gav värdet för *MI-anda*. Den sjugradiga likertskalan byttes även till en femgradig och skattningarna försågs med tydliga kriterier för varje skalsteg. Vid en jämförelse mellan de svenska översättningarna av MITI 2.0 resp. 3.0 uppmättes en signifikant högre IBR i den senare versionen (Johansson, 2008). I den senaste större uppdateringen av MITI har de övergripande skattningarna förändrats ytterligare för att bättre fånga de tekniska komponenterna i MI-utövandet (Moyers et al., 2016).

Beteendefrekvensräkningarna lämnades i stort sett oförändrade i den första uppdateringen men i och med den senaste versionen har de fått en delvis ny utformning. I tidigare versioner utgjorde *yttranden förenliga/oförenliga med MI* generella kategorier som kodaren förväntades kunna identifiera. Dessa variabler finns kvar i MITI 4.2.1, men nu som summeringar av

mer specifika beteendekategorier: *bekräfta+söka samarbete+betona autonomi=förenliga med MI*; *övertala+konfrontera=oförenliga med MI*. Beteendekategorierna *öppna* resp. *slutna frågor* kodas i MITI 4.2.1 som enbart *frågor*. En annan nytillkommen beteendekategori i MITI 4.2.1 är *övertala med tillstånd* som till skillnad från *övertala* inte påverkar summeringen negativt.

I sammanhang där det inte finns möjlighet att anlita ett professionellt kodningslabb kan det vara ett alternativ att utbilda kodare på plats. Flera studier har visat att personer utan tidigare erfarenhet av MI kan lära sig att använda MITI och nå en tillfredställande IBR (Jelsma et al., 2015). Eftersom varje kodningssession består av en eller två genomlysningar av ett 20-minuterssegment bör själva kodningsproceduren ta mindre än en timme. Tidsåtgången för utbildning och kodning har av vissa beskrivits som effektiv och ekonomisk (Moyers et al., 2005; Wallace & Turner, 2009) medan andra har sett den som tidskrävande och omständlig (Small, Lee, Frey, Seeley & Walker, 2014; Jelsma et al., 2015). I jämförelse med MISC, som kräver omkring tre månaders utbildning och där varje kodningssession tar upp till två timmar i anspråk (Moyers et al., 2005), innebär dock MITI en avsevärd tidsbesparing.

MITI kan fungera både som objektiva mätinstrument i vetenskapliga studier och som en del i den pedagogiska processen (Moyers et al., 2016). I vissa sammanhang kan det vara praktiskt att använda valda delar av instrumentet, t ex. när en terapeut behöver träna på någon enskild teknik i samråd med en erfaren handledare. Det är dock viktigt att särskilja denna typ av riktad handledning från en objektiv användning av instrumentet; mätningar som görs i handledningssyfte bör alltså inte samtidigt användas för att fastställa behandlingstroheten i en vetenskaplig studie (Moyers et al., 2016). Användarmanualen öppnar för modifieringar och anpassade tillämpningar av MITI, men samtidigt påpekas att instrumentet i sådana fall inte kan betraktas som empiriskt validerat. För att få fram reliabel information och kunna jämföra indexpoäng enligt rekommenderade gränsvärden krävs att samtliga variabler kodas enligt manualen (Moyers et al., 2014).

Interbedömarreliabilitet

I beteendevetenskapliga studier mäts behandlingstroheten ofta genom observationer som kodas till kvantifierbar data som sedan kan utvärderas med olika statistiska metoder. Ett problem med data baserad på observationer är att olika individer kan bedöma studieobjektet på olika sätt. En viss individuell varians är säkert ofrånkomlig men om mätningarna blir alltför spretiga begränsas möjligheterna att dra slutsatser från resultaten. Därför är det av yttersta vikt att mätinstrumentet har en tydlig manual och ger likvärdiga resultat vid upprepade mätningar utförda av olika bedömare. Ett mätinstrument som uppfyller dessa krav kan sägas besitta en hög interbedömarreliabilitet.

Två frågor är viktiga att ta ställning till när det gäller val av koefficient för IBR: a) vad avses med IBR, absolut eller relativ överensstämmelse, och b) på vilken skalnivå befinner sig insamlade data. Båda frågorna har ofta lämnats obesvarade i MI-litteraturen. I flera studier har både ICC och Cronbachs α rapporterats med motiveringen att båda mäter IBR, fast ICC anses vara den mer konservativa koefficienten (Moyers et al., 2005; Forsberg, Källmén, Hermansson, Berman & Helgason, 2007; Pierson et al., 2007). Denna uppfattning återfinns även på andra håll, (Moyers et al., 2016; Seng & Lovejoy, 2013). I en översikt över olika mätinstrument påpekar dock Wallace och Turner (2009) att Cronbachs α inte går att använda för att uppskatta IBR och ifrågasätter varför det över huvud taget rapporteras. I mer generella metodöversikter tycks det dock råda samstämmighet om vilken koefficient som är tillämplig

givet omständigheterna (Lombard, Snyder-Duch & Bracken, 2002; Feng, 2015; Hayes & Krippendorff, 2007).

I rent statistiska sammanhang görs det en distinkt skillnad mellan intern konsistens och IBR (Cicchetti, 1994; Shrout & Fleiss, 1979; Cortina, 1993). Den interna konsistensen, dvs. hur två eller flera oberoende bedömningar samvarierar, ska inte förväxlas med IBR, som berättar något om den absoluta överensstämmelsen mellan två eller flera bedömare. En IBR på oacceptabelt låg nivå, orsakad av att bedömarna konstant ligger två skalsteg från varandra, kan motsvaras av en perfekt intern konsistens, så länge avvikelsen är konstant (Cicchetti, 1994). Omvänt kan en låg intern konsistens, orsakad av slumpmässig varians mellan bedömarna, motsvaras av en betydligt högre IBR, förutsatt att avvikelserna är relativt små. Här handlar det alltså inte om vilket mått som är mest konservativt; det är en fråga om vad som mäts, kovarians eller absolut överensstämmelse. Eftersom denna studie handlar om IBR så är det främst dessa resultat som diskuteras. Därmed inte sagt att den interna konsistensen skulle vara irrelevant, men det är viktigt att klargöra vad det är som mäts och att detta görs med ett ändamålsenligt mått. För en översikt över tillämpbara mått för IBR hänvisas till Feng (2015).

Eftersom variablerna i MITI 4.2.1 mäts både på ordinalskala (övergripande skattningar) och på kvotskala (beteendefrekvenser och summerande variabler) krävs olika beräkningar för IBR (Feng, 2015). I tidigare studier av MITI har ICC använts genomgående, men enligt senare rekommendationer är detta inte tillämpligt på det övergripande skattningarna där Krippendorffs α (Kalpha) i stället bör användas (Jelsma et al., 2015). Kalpha är mycket flexibelt i flera avseenden och dess fördelar har lyfts fram på flera håll (Hayes & Krippendorff, 2007; De Swert, 2012; Lombard et al., 2002). Användningen är inte begränsad till en viss datanivå och beräkningen tillåter uteblivna värden, vilket inte är fallet med ICC. Kalpha är även känsligt för sällsynt förekommande kategorier vilket innebär att om en enskild kodare anger ett värde som sällan eller aldrig används i andra kodningar kommer detta att påverka resultatet negativt (De Swert, 2012). Den främsta anledningen till att måttet inte används i större utsträckning tycks vara att det ännu inte inkluderats som standard i populära statistikprogram som t ex SPSS. Jelsma och kolleger (2015) rekommenderar Kalpha för utvärdering av de kvalitativa variablerna och menar att de främsta anledningarna till detta är dess förmåga att hantera uteblivna värden och begränsad spridning.

Moyers och kolleger (2016) använder inte Kalpha enligt rekommendation, utan menar att ICC är att föredra för dessa variabler, dels för att det är mer konservativt än Cronbachs α , dels för att data befinner sig på ordinalnivå. Mot bakgrund av ovanstående kan det te sig lite märkligt att varken Moyers eller Jelsma närmare berör att ICC över huvud taget inte rekommenderas för kvalitativa variabler (Feng, 2015). En förklaring skulle kunna vara att det inom olika vetenskapliga discipliner utvecklas olika syn på vilka metoder som kan och bör användas. Observation och kodning används inom både beteendevetenskap och media- och kommunikationsvetenskap men metoderna verkar skilja sig mellan områdena. I beteendevetenskapligt orienterade metodöversikter är ICC väl beskrivet medan Kalpha bara nämns i förbigående (Hallgren, 2012; Shrout & Fleiss, 1979); omvänt beskrivs Kalpha utförligt inom innehållsanalys medie- och kommunikationsvetenskap där ICC hamnar lite mer i skymundan (Lombard et al., 2002; Krippendorff, 2004; Feng, 2015).

Det råder sedan länge delade meningar om huruvida skattningar enligt likertskalan genererar data på ordinal- eller intervallnivå (se t ex. Knapp, 1990 för en ingående diskussion om detta). Enligt mer konservativa teoretiker kan likertskalor aldrig generera intervalldata medan det är allmän praxis inom beteendevetenskaperna (Stevens, 1946). När det gäller MITI har

uppenbarligen både instrumentets skapare (Moyers et al., 2016) och dess utforskare (Jelsma et al., 2015) valt att betrakta de övergripande skattningarna som ordinaldata. Detta verkar vara en rimlig utgångspunkt med tanke på att likertskalan inte har vare sig någon absolut eller arbiträr nollpunkt samt att de övergripande skattningarna har utformats som än mer kvalitativa i de senare versionerna av MITI (Moyers et al., 2014). Att tillföra fler kategorier till en ordinalskala tenderar att ”kontinuisera” den (Knapp, 1990). I den första uppdateringen av MITI reducerades i stället antalet kategorier i de övergripande skattningarna. Kategorierna kompletterades dessutom med detaljerade kriterier vilket förstärker intrycket av att skalstegen förhåller sig kvalitativt, snarare än kvantitativt, till varandra.

Det primära syftet med denna uppsats är att undersöka IBR för variabelerna i MITI 4.2.1, baserad på data från fyra nyutbildade kodare. Eftersom de nya rekommendationerna för mätning av IBR hos MITI ännu inte tycks ha slagit igenom på bred front är det även av intresse att öppna för en diskussion kring hur de olika koefficienterna ska tolkas och i vilken utsträckning de är rimliga att tillämpa. Förhoppningen är att kunna besvara frågan om IBR i det insamlade materialet. Gällande frågor kring metod och praxis är ambitionen inte att leverera några färdiga svar, utan snarare att öka medvetenheten kring det till synes enkla, men i grunden komplexa, begreppet IBR.

Metod

Kodare, material och kodning

Ett fåtal kodningslabb runt om i världen erbjuder professionell MITI-kodning med högt ställda krav på IBR (Jelsma et al., 2015). I Sverige finns MIC Lab som utför professionell MITI-kodning på uppdrag, bedriver forskning samt ger utbildningar för samtalsledare och kodare. MIC Lab drivs som ett fristående aktiebolag men samverkar med Karolinska Institutet och Stockholms läns landsting. MITI-kodningarna som analyseras i den här uppsatsen har utförts enligt den senaste versionen på svenska, MITI 4.2.1, inom ramen för MIC Labs kodarutbildning våren 2016.

Fyra personer, 26–30 år gamla, deltog i kodarutbildningen på MIC Lab. Parallellt med utbildningen studerade tre av deltagarna vid juristprogrammet och en vid logopedprogrammet. Samtliga har, utöver svenska, mycket goda kunskaper i engelska samt i tre fall ytterligare något språk. Kodarutbildningen sträckte sig över fyra veckor under april–maj 2016, och omfattade tre seminarietillfällen, examination samt avslutande personlig feedback. Kursdeltagarna fick även hemuppgifter som sedan diskuterades under seminarierna. Ingen exakt tidsdokumentation finns tillgänglig, men uppskattningsvis kan utbildningen motsvara de 40 timmar som rekommenderas i litteraturen (Jelsma et al., 2015; Moyers et al., 2010; Moyers et al., 2016).

Kursdeltagarna examinerades genom att tolv inspelade samtal om tjugo minuter kodades individuellt av samtliga deltagare. Kodningarna jämfördes sedan mot facit för att de kodare som uppnått tillfredställande resultat skulle kunna erbjudas tjänst som professionell kodare på MIC Lab. Facit har tagits fram genom att två professionella kodare med flera års erfarenhet, har kodat samtliga samtal oberoende av varandra. Skillnader i kodningarna har sedan diskuterats varpå konsensusbeslut fattats för att nå absolut överensstämmelse, en så kallad Gold Standard.

Det inspelade materialet är hämtat från en studie om MI-utbildning för landstingspersonal (Forsberg, Beckman, Ghaderi, Lindqvist, Öhman & Börjesson-Thot, 2014). För att standar-

disera svårighetsgrad och variation i samtalen har professionella skådespelare spelat olika klienter enligt på förhand givna scenarion.

Statistiska metoder och databearbetning

I denna studie redovisas både ICC och Kalpha som mått på IBR i samtliga variabler. Det finns flera anledningar till att använda dubbla mått för IBR. Det gör det möjligt att jämföra resultaten både med tidigare studier, där enbart ICC redovisas, och studier som redovisar ICC och Kalpha enligt senare rekommendationer (Jelsma et al., 2015). Det möjliggör dessutom jämförelser med studier som redovisar Kalpha för samtliga MITI-variabler, för det fall att detta skulle komma att rekommenderas framgent. Utöver denna fråga om kompatibilitet kan det vara intressant att se i vilken grad ICC och Kalpha skiljer sig åt för olika variabler.

Samtliga beräkningar av medelvärden, ICC, Kalpha och Cronbachs α har genomförts i SPSS Statistics 23. Den variant av ICC som rekommenderas i detta sammanhang är en tvåvägs mixad modell med absolut överensstämmelse (Jelsma et al., 2015). I datasetet har kodarna sorterats som variabler i kolumner, och samtalen som case i rader. Kalpha har beräknats enligt samma princip. Kalpha ingår inte som standard i SPSS, men kan laddas ned som ett separat makro (se t ex De Swert, 2012). Utöver ICC och Kalpha har Cronbachs α angivits som ett mått på intern konsistens (dvs. i vilken grad kodningarna korrelerar med varandra).

Eftersom en av kodarna endast kodat elva av tolv samtal innehåller datasetet en liten andel saknade värden. Dessa har ersatts i SPSS med hjälp av en EM-algoritm (expectation-maximization). Kalpha hanterar normalt sett dataset med saknade värden men eftersom ICC inte gör det har de ersatta värdena använts för samtliga beräkningar. Vid hypotesprövning av ICC, Kalpha och Cronbachs α från olika stora datamängder, har värdena först transformerats till Fishers z och sedan prövats manuellt som skillnader mellan korrelationer (α -nivå = 5 %, tvåsidig prövning).

Resultat

Tabell 1 visar en överblick över samtliga variabler i MITI 4.2.1 samt hur medelvärdena i de genomförda kodningarna förhåller sig till *Gold Standard*. De tolv samtal som kodats inför denna studie har handplockats ur en större studie för att representera en variation av svårighetsgrader och problem som kan tänkas uppstå i verkliga samtalssituationer. Det finns alltså ingen tydlig eller central tendens i samtalen varför dessa medelvärden inte ger någon direkt information om kodningarnas riktighet. Det som kan utläsas i tabell 1 är att den sammantagna kodningen av vissa variabler avviker kraftigt från facit (*ge information, övertala* samt *enkla* respektive *komplexa reflektioner*). Det går även att utläsa att frekvensen för samtalsbeteenden *förenliga med MI* generellt sett ligger något högre än facit medan det omvända gäller för samtalsbeteenden *oförenliga med MI*. Samma typ av korsvisa förhållande, om än i mindre grad, gäller för de tekniska komponenterna *främja förändringstal* respektive *dämpa bibehållandetal*.

Tabell 1. Medelvärden (och standardavvikelser) för MITI-variabler i 12 samtal kodade av kodare W, B, C och A. Höger kolumn visar motsvarande värden för Gold Standard.

MITI 4.2.1	Oberoende kodare <i>M (S)</i>	Gold Standard [†] <i>M (S)</i>
<i>Övergripande skattningar</i>		
Främja förändringstal	2.65 (0.86)	2.17 (1.03)
Dämpa bibehållandetal	2.19 (0.87)	2.42 (0.90)
Partnerskap	2.85 (1.07)	2.50 (1.17)
Empati	2.92 (1.01)	2.75 (1.14)
<i>Beteendefrekvenser</i>		
Ge information	4.35 (4.92)	9.42 (8.46)
Övertala med tillstånd	1.27 (1.62)	0.58 (0.67)
Bekräfta	0.44 (0.80)	0.92 (1.17)
Söka samarbete	1.71 (1.74)	1.67 (1.78)
Betona autonomi	0.92 (1.75)	0.17 (0.58)
Övertala	2.29 (2.81)	4.00 (4.33)
Konfrontera	1.08 (1.72)	1.50 (2.02)
Enkla reflektioner	5.81 (3.20)	7.42 (3.75)
Komplexa reflektioner	5.56 (2.58)	4.50 (3.15)
Frågor	14.92 (8.49)	14.92 (9.25)
<i>Summerande variabler</i>		
Tekniska	2.41 (0.76)	2.29 (0.78)
Relationella	2.89 (1.00)	2.63 (1.07)
Andel komplexa refl.	0.50 (0.20)	0.38 (0.23)
Reflektioner / Frågor	0.94 (0.51)	1.02 (0.61)
Förenliga med MI	3.07 (3.32)	2.75 (2.86)
Oförenliga med MI	3.37 (4.13)	5.50 (5.78)

[†] Gold Standard utgör facit för kodarutbildningens examinationsuppgift. Det har tagits fram genom att två professionella kodare har kodat alla samtal oberoende av varandra och sedan diskuterat fram ett konsensusbeslut vid bristande överensstämmelse.

Tabell 2 visar två olika beräkningarna av ICC, Kalpha och Cronbachs α för respektive variabel. Enligt Cicchettis (1994) riktlinjer för tolkning av ICC kan IBR klassas som utmärkt (>0.75) för tio av fjorton variabler medan två variabler når upp till gränsen för bra (0.60–0.74). IBR för de återstående två variablerna når endast upp till gränsen för det godtagbara (0.40–0.59) vilket tyder på brister i kodningen av dessa. Den ena av dessa variabler, *dämpa bibehållandetal*, är en kvalitativ variabel som inte funnits med i tidigare versioner av MITI, den andra är den kvantitativa variabeln *övertala med tillstånd*.

Som framgår av tabell 2 ligger Kalpha genomgående betydligt lägre än ICC. Krippendorff (2004) själv anser att ett Kalpha på minst 0.8 krävs för att en variabel ska kunna betraktas som reliabel. Värden ned till 0.6 kan vara acceptabla, men i sådana fall bör analysen kompletteras med ett resonemang kring den aktuella variabelns egenskaper – kan den missförstås, är manualen tydlig, är skalstegen distinkta etc. (De Swert, 2012). Samtliga variabler utom en summerande variabel (*reflektioner / frågor*, dvs. förhållandet mellan det totala antalet reflektioner och antalet frågor) ligger långt under gränsen för det acceptabla. När de två minst

samstämmiga kodarna (B och C) utesluts ur analysen ökar både ICC och Kalpha i de flesta variablerna. Efter transformering till Fishers z visar nio av fjorton variabler en signifikant ökning av ICC medan samtliga variabler utom två visar en signifikant ökning av Kalpha ($p < 0.05$). När enbart kodarna W och A ingår i analysen ökar även Cronbachs α signifikant i flera fall, och tenderar även att anta värden närmare ICC.

Tabell 2. ICC, Kalpha och Cronbachs α för 12 samtal kodade av kodare W, B, C och A (enbart kodare W och A).

MITI 4.2.1	ICC	Kalpha	Cronbachs α
<i>Övergripande skattningar</i>			
Främja förändringstal	0.71 ^b (1.00 ^a)*	0.28 (0.95 ^d)*	0.77 (1.00)*
Dämpa bibehållandet	0.47 ^c (1.00 ^a)*	0.14 (0.99 ^d)*	0.55 (1.00)*
Partnerskap	0.78 ^a (0.97 ^a)*	0.45 (0.92 ^d)*	0.79 (0.97)*
Empati	0.81 ^a (0.97 ^a)*	0.38 (0.87 ^d)*	0.82 (0.97)*
<i>Beteendefrekvenser</i>			
Ge information	0.85 ^a (0.66 ^b)	0.35 (0.86 ^d)*	0.86 (0.66)
Övertala med tillstånd	0.49 ^c (0.81 ^a)*	0.20 (0.72 ^e)*	0.54 (0.84)*
Bekräfta	0.79 ^a (0.92 ^a)	0.52 (0.81 ^d)*	0.82 (0.93)
Söka samarbete	0.85 ^a (0.95 ^a)*	0.41 (0.77 ^e)*	0.88 (0.95)
Betona autonomi	0.80 ^a (0.99 ^a)*	0.21 (0.84 ^d)*	0.83 (0.99)*
Övertala	0.70 ^b (0.99 ^a)*	0.49 (0.95 ^d)*	0.72 (0.99)*
Konfrontera	0.76 ^a (0.97 ^a)*	0.46 (0.85 ^d)*	0.77 (0.98)*
Enkla reflektioner	0.86 ^a (0.83 ^a)	0.52 (0.83 ^d)*	0.89 (0.83)
Komplexa reflektioner	0.86 ^a (0.91 ^a)	0.54 (0.72 ^e)	0.89 (0.92)
Frågor	0.97 ^a (1.00 ^a)*	0.59 (0.99 ^d)*	0.98 (1.00)*
<i>Summerande variabler</i>			
Förenliga med MI	0.86 ^a (0.96 ^a)*	0.27 (0.73 ^e)*	0.90 (0.96)
Oförenliga med MI	0.74 ^b (0.98 ^a)*	0.41 (0.92 ^d)*	0.76 (0.98)*
Tekniska komponenter	0.65 ^b (1.00 ^a)*	0.29 (1.00 ^d)*	0.74 (1.00)*
Relationskomponenter	0.80 ^a (0.97 ^a)*	0.52 (0.94 ^d)*	0.81 (0.97)*
Andel komplexa refl.	0.87 ^a (0.88 ^a)	0.55 (0.65)	0.88 (0.89)
Tot. Refl. / Frågor	0.90 ^a (0.99 ^a)*	0.78 ^e (0.98 ^d)*	0.91 (0.99)*

^a Utmärkt, ^b bra, ^c godtagbart enligt Cicchetti (1994).

^e Godtagbart enligt Krippendorff (2004).

* Signifikant ($p < 0.05$) ökning vid analys av enbart kodare W och A.

Som ytterligare en dimension av IBR i de genomförda kodningarna har även en sammantagen ICC, Kalpha och Cronbachs α beräknats för varje kodat samtal (se tabell 3). Detta innebär med andra ord att samstämmigheten mellan kodarna för samtliga variabler i respektive samtal ligger till grund för beräkningen. Såväl ICC som Kalpha och Cronbachs α hamnar på en generellt sett hög nivå och skillnaden mellan de olika koefficienterna är betydligt mindre påtaglig än när mätningen görs för separat för varje variabel (jfr. tabell 2).

Tab. 3. ICC, Kalpha och Cronbachs α för varje samtal.

MI-samtal [‡]	ICC	Kalpha	Cronbachs α
1	0.71 ^b	0.40	0.71
2	0.99 ^a	0.96 ^d	0.99
3	0.99 ^a	0.95 ^d	0.99
4	0.96 ^a	0.85 ^d	0.96
6	0.91 ^a	0.71 ^e	0.91
17	1.00 ^a	0.98 ^d	1.00
18	0.86 ^a	0.59	0.85
23	0.99 ^a	0.94 ^d	0.99
24	0.98 ^a	0.92 ^d	0.98
28	0.74 ^b	0.41	0.77
29	0.96 ^a	0.84 ^d	0.96
30	0.96 ^a	0.85 ^d	0.96

[‡] Efter sortering i den ursprungliga studien (Forsberg et al., 2014)

^a Utmärkt, ^b bra, ^c godtagbart enligt Cicchetti (1994).

^d Utmärkt, ^e godtagbart enligt Krippendorff (2004).

Diskussion

Syftet med denna studie har varit att undersöka interbedömarreliabiliteten i den svenska versionen av MITI 4.2.1. För att kunna föra en diskussion kring olika metoder för mätning av IBR har tre olika mått redovisats parallellt. Resultatet visar mindre eller större brister hos vissa variabler, men även att de huvudsakliga koefficienterna för IBR, intraklasskorrelation och Krippendorffs α , skiljer sig åt på ett markant sätt.

Intraklasskorrelationer och Krippendorffs α har beräknats för samtliga variabler och i flera fall hamnar resultatet under gränsen för det acceptabla. ICC ligger genomgående betydligt högre än Kalpha. Detta verkar rimligt med tanke på att Kalpha påverkas negativt av värden som bara förekommer vid enstaka tillfällen (De Swert, 2012). Vid mätning av de två mest samstämmiga kodarna visade Kalpha i flera fall en signifikant ökning. Detta visar ytterligare på känsligheten hos Kalpha. Även Cronbachs α har beräknats som ett mått på intern konsistens och ligger genomgående på en hög nivå. Det högre värdet på Cronbachs α visar att det ligger någon grad av systematisk variation bakom bristen på absolut överensstämmelse. Även om kodningarna inte överensstämmer i absolut mening så varierar de alltså i samma riktning snarare än helt slumpmässigt.

Kalpa för IBR visar ett betydligt sämre resultat än motsvarande ICC. Ingen av de aktuella variablerna uppnår fullgod eller ens acceptabel nivå enligt Krippendorffs (2004) respektive De Swerts (2012) rekommendationer där allra minst 0.60 men helst 0.80 bör uppnås. Detta blir särskilt tydligt hos de kvantitativa variablerna som i flera fall är relativt lågfrekventa och samtidigt varierar kraftigt mellan de fyra kodarna. Kalpha är visserligen inte rekommenderat för dessa variabler (Jelsma et al., 2015) men i föreliggande studie har det varit av metodiskt intresse att ställa de båda måtten sida vid sida. Å ena sidan kan resultatet tyda på att Kalpha är sämre lämpat för reliabilitetsmätning av dessa variabler; ett mått som drastiskt drar ned resultatet på grund av enstaka avvikelser är knappast användbart i praktiken. Å andra sidan

vore det inte helt orimligt att ställa höga krav på absolut överensstämmelse i kodningen av just dessa variabler; kodningsmanualen och kodningsträningen bör ge otvetydiga definitioner på hur olika samtalsbeteenden ska tolkas och kodas. Manualen är genomgående mycket detaljerad och innehåller beslutsregler och lösningar på motsägelsefulla kodningsexempel (Moyers et al., 2014). Utbildningen av nya kodare bedöms ta minst 40 timmar i anspråk (Jelsma et al., 2015) och bedrivs både i seminarieform och som självstudier. Eftersom kodarna i föreliggande studie varit under utbildning parallellt med andra krävande heltidsstudier är det tänkbart att de inte hunnit nått upp till den rekommenderade studiemängden. Det är med andra ord möjligt att ett bättre resultat skulle kunna uppnås efter ytterligare träning.

Genom att ögna igenom rådatan och göra enkla korrelationer mellan kodningssessioner går det att bilda sig en uppfattning om hur väl enskilda kodare överensstämmer med varandra. Observerad överensstämmelse är dock inte detsamma som reliabilitet. Det IBR berättar är något annat, nämligen sannolikheten för att samma population kodare kommer att överensstämma i sina bedömningar vid framtida tillfällen. Naturligtvis kan resultatet dras ned av brister hos individuella kodare, men IBR säger ytterst något om själva mätinstrumentet och de enskilda variablerna. Det är alltså tänkbart att vissa MITI-variabler varit svåra att förstå, att de överlappar varandra på ett förvirrande sätt, eller att de inte fungerar tillsammans för att mäta det avsedda, alltså i vilken utsträckning samtalsledaren använder MI-specifika färdigheter.

En del av förändringarna i den senaste versionen syftar till att göra MITI mer användarvänligt. Där tidigare versioner krävt kunskap om, och erfarenhet av MI, är variablerna i MITI 4.2.1 utformade för att bättre kunna användas av kodare utan tidigare erfarenhet. Det är t ex enklare att koda alla frågor som en enda beteendekategori än att särskilja öppna och slutna frågor. Även kodningen av förenliga och oförenliga samtalsbeteenden är enklare utformad i MITI 4.2.1.

Samtidigt har flera variabler fått en tydligare koppling till varandra och till teoribildningen bakom MI. Beteendefrekvensräkningarna är i högre grad kopplade till de övergripande skattningarna genom beslutsregler som anger hur kodning av t ex. *frågor*, *övertalning med tillstånd* och *ge information* återspeglas i de kvalitativa variablerna (Moyers et al., 2014). Vidare har de övergripande skattningarna delats upp i tekniska och relationella komponenter och på så vis kopplingen till Miller och Rose (2009) relationella och tekniska hypoteser blivit tydligare. Även om ett av syftena med MITI 4.2.1 har varit att utveckla ett lättanvänt mätinstrument som inte kräver tidigare erfarenhet kan en bristande förståelse för teorin bakom MI göra det svårt för en oerfaren kodare att göra dessa kopplingar. Det återstår att se hur dessa syften (ökad användbarhet och starkare teoretisk förankring) kommer att gå hand i hand. Förhoppningsvis kan effektivare utbildning och träning av nya kodare resultera i att teori och praktik sammanbinds.

Eftersom variabler tillkommit och tagits bort i och med MITI 4.2.1 har det inte varit meningsfullt att göra jämförelser med tidigare versioner. Johansson (2008) har gjort en jämförelse av de två föregående versionerna, men det var möjligt tack vare att materialet dubbelkodades av samma kodare i båda versionerna. Detta skulle förstås vara intressant för att göra motsvarande jämförelse mellan de senaste versionerna, men inför denna studie har det saknats tid och resurser att anlita erfarna kodare. Att använda material från kodare under utbildning är ett enkelt och billigt sätt att samla in data men som underlag för en större jämförelse skulle denna typ av material inte vara tillräckligt tillförlitligt. I takt med att fler professionella kodare samlar erfarenhet av flera versioner av MITI skulle det vara intressant

att se fler jämförande studier av de olika versionernas reliabilitet och validitet. Både utforskandet och utövandet av MI har en relativt kort historia och nya fynd skulle kunna bidra till såväl teoretiskt som praktiskt förståelse.

Miller och Mount (2001) påpekar att intervjuer där klientrollen spelas av skådespelare inte motsvarar verkligheten. Detta kanske främst har betydelse för träningen av samtalsledare, men skulle även kunna påverka samtalskodarna, och därmed IBR. Ett iscensatt klientbeteende som inte verkar trovärdigt skulle kunna leda till att samtalsledaren inte använder sina färdigheter med samma skärpa som i en autentisk situation. Å andra sidan är det sannolikt att ett manus som definierar klienten skrivits med olika specifika variabler i åtanke, och därmed riskerar att bli övertydligt i dessa avseenden. I båda fallen skulle kodningen kunna färgas av observatörernas subjektiva tolkningar och ansträngningar att identifiera olika samtalskomponenter som passar in i MITI. Det är förstås praktiskt att använda iscensatta samtal eftersom det går att standardisera innehållet genom att skapa olika berättelser och samtalsituationer. Att samla in autentiska samtal är mer resurskrävande samtidigt som det inte på förhand går att veta vad de kommer att handla om. För att kunna mäta verkliga effekter, och samtidigt säkerställa dessa mätningars reliabilitet, är dock verkliga samtal med autentiska intentioner att föredra.

En uppenbar begränsning i den här uppsatsen är min egen bristande kunskap om statistiska metoder och beräkningar. Arbetet med den har till stor del gått ut på att försöka förstå och förhålla sig till komplexa begrepp och metoder som ofta förutsätter kunskap och erfarenhet på avancerad nivå. MI som forskningsområde kräver inte bara djupa kunskaper om MI, det kräver även en djupare förståelse av metoderna för att undersöka och utvärdera MI. Tydlighet kring vilka mått som används, och varför de används, skulle kunna bidra till att förtydliga hur olika typer av mått på reliabilitet förhåller sig till varandra. Att klassificera ett visst mått som mer eller mindre konservativt än ett annat säger helt enkelt inte så mycket så länge det inte framgår vad som ska mätas (Krippendorff, 2004). Förhoppningsvis kommer mer specifika och initierade metodstudier att öka förståelsen för likheter och skillnader mellan statistiska metoder och mått som används för att utvärdera MI. Detta skulle säkerligen uppskattas av såväl studenter som erfarna forskare.

Själva datainsamlingen inför denna studie utgör både dess brister och styrkor. Samtidigt som urvalet hade kunnat vara större, både till antalet kodare och mängden kodade samtal, är tillgången till professionellt insamlad data en uppenbar styrka. Att på egen hand administrera en såpass omfattande kodningsprocedur hade varit näst intill omöjligt inom ramen för en uppsats på denna nivå. Det faktum att kodarna inte hunnit skaffa sig någon större erfarenhet kan ses som en svaghet då resultatet kan tänkas spegla individuella brister snarare än själva mätinstrumentet i sig. För framtida forskning skulle det kunna vara av intresse att närmare undersöka skillnader mellan olika grupper av kodare, olika samtals typer och olika samtalsmiljöer. På så sätt skulle bättre kunskap om analys av IBR kunna bidra både till utvecklingen av enskilda MITI-variabler och det pedagogiska arbetet med att utbilda kompetenta kodare.

Slutsatsen av denna studie är att den svenska versionen av MITI 4.2.1 än så länge visar uppenbara brister i IBR hos flera variabler. Vidare blottläggs stora skillnader i själva mätningen av IBR när olika koefficienter ställs bredvid varandra. Bättre träning av samtalskodare och ökad förståelse för metoderna bakom IBR är en liten, men samtidigt väldigt viktig del av den fortsatta utvecklingen av MI som samtalsmetod.

Referenser

- 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), 862–878.
- Bem, D. J. (1967). Self-perception: an alternative interpretation of cognitive dissonance phenomena. *Psychological review, 74*, 183–200.
- Burke, B. L., Arkowitz, H., & Menchola, M. (2003). The efficacy of motivational interviewing: A meta-analysis of controlled clinical trials. *Journal of consulting and clinical psychology, 71*(5), 843–861.
- Cicchetti, V. (1994). Guidelines, criteria and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological assessments, 6*, 284–290.
- Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of applied psychology, 78*(1), 98–104.
- De Swert, K. (2012). Calculating inter-coder reliability in media content analysis using Krippendorff's Alpha. *Center for politics and communication*, 1–15.
- Feng, (2015). Mistakes and how to avoid mistakes in using intercoder reliability indices. *Methodology, 11*(1), 13–22.
- Forsberg, L., Beckman, M., Ghaderi, A., Lindqvist, H., Öhman, L., & Börjesson-Thot, K., (2014). *Utvärdering och utveckling av utbildningar i motiverande samtal inom hälso- och sjukvården*. (Forskningsrapport, november 2014). Karolinska Institutet: Institutionen för klinisk neurovetenskap. Hämtad från Socialstyrelsen 2016-05-02.
- 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.
- Glynn, L. H., & Moyers, T. B. (2010). Chasing change talk: The clinician's role in evoking client language about change. *Journal of substance abuse treatment, 39*(1), 65–70.
- 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.
- Hayes, A. F., & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication methods and measures, 1*(1), 77–89.
- Hettema, J., Steele, J., & Miller, W. R. (2005). Motivational interviewing. *Annual review of Clinical psychology, 1*, 91–111.
- Jelsma, J. G., Mertens, V. C., Forsberg, L., & Forsberg, L. (2015). How to measure motivational interviewing fidelity in randomized controlled trials: practical recommendations. *Contemporary clinical trials, 43*, 93–99.
- Johansson, N. (2008). *Interbedömarreliabilitet i kodningsverktyget "motivational interviewing treatment integrity code" 3.0 (MITI 3.0)*. (opublicerad c-uppsats). Stockholm universitet, Psykologiska institutionen.
- Knapp, T. R. (1990). Treating ordinal scales as interval scales: An attempt to resolve the controversy. *Nursing research, 3*(2), 121–123.
- Krippendorff, K. (2004). Reliability in content analysis. *Human communication research, 30*(3), 411–433.
- Lombard, M., Snyder-Duch, J., & Bracken, C. C. (2002). Content analysis in mass communication: Assessment and reporting of intercoder reliability. *Human communication research, 28*(4), 587–604.

- 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., Kunz, C., Brownell, C., Tollefson, D., & Burke, B. (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., & Campbell, T. C. (2006). Measures of fidelity in motivational enhancement: A systematic review. *Journal of substance abuse treatment, 31*(1), 67–73.
- McMaster, F., & Resnicow, K. (2015). Validation of the one pass measure for motivational interviewing competence. *Patient education and counseling, 98*(4), 499–505.
- Miller, W. R., & Mount, K. A. (2001). A small study of training in motivational interviewing: Does one workshop change clinician and client behavior? *Behavioural and cognitive psychotherapy, 29*(04), 457–471.
- Miller, W. R., & Rollnick, S. (2009). Ten things that motivational interviewing is not. *Behavioural and cognitive psychotherapy, 37*, 129–140.
- Miller, W. R., & Rollnick, S. (2013). *Motiverande samtal: att hjälpa människor till förändring*. (3:e rev. utg.). Stockholm: Natur & kultur.
- Miller, W. R., & Rollnick, S. (2014). The effectiveness and ineffectiveness of complex behavioral interventions: Impact of treatment fidelity. *Contemporary clinical trials, 37*(2), 234–241.
- Miller, W. R., & Rose, G. S. (2009). Toward a theory of motivational interviewing. *American psychologist, 64*(6), 527–537.
- Miller, W. R., & Rose, G. S. (2013). Motivational interviewing and decisional balance: Contrasting responses to client ambivalence. *Behavioural and cognitive psychotherapy, 43*(2), 129–141.
- Miller, W. R., Yahne, C. E., Moyers, T. B., Martinez, J., & Pirritano, M. (2004). A randomized trial of methods to help clinicians learn motivational interviewing. *Journal of consulting and clinical psychology, 72*(6), 1050–1062.
- Moyers, T. B., Martin, T., Manuel, J. K., & Miller, W. R. (2003). *The motivational interviewing treatment integrity (MITI) code: Version 2.0*. Unpublished manual.
- Moyers, T. B., Martin, T., Manuel, J. K., Miller, W. R., & Ernst, D. (2010). *Revised global scales: Motivational interviewing treatment integrity coding manual 3.1.1*. Unpublished manual.
- Moyers, T. B., Manuel, J. K., & Ernst, D. (2014). *Motivational interviewing treatment integrity coding manual 4.2.1*. Unpublished manual.
- 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., 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.
- 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.
- Prochaska, J. O., & DiClemente, C. C. (1982). Transtheoretical therapy: Toward a more integrative model of change. *Psychotherapy: theory, research & practice, 19*(3), 276–288.

- Rogers, C. R. (1957). The necessary and sufficient conditions of therapeutic personality change. *Psychology: Journal of consulting psychology, 21*(2), 95–103.
- Rollnick, S., & Miller, W. R. (1995). What is motivational interviewing? *Behavioural and cognitive psychotherapy, 23*(04), 325–334.
- Seng, E. K., & Lovejoy, T. I. (2013). Reliability and validity of a treatment fidelity assessment for motivational interviewing targeting sexual risk behaviors in people living with HIV/AIDS. *Journal of clinical psychology in medical settings, 20*(4), 440–448.
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological bulletin, 86*(2), 420–428.
- Small, J. W., Lee, J., Frey, A. J., Seeley, J. R., & Walker, H. M. (2014). The development of instruments to measure motivational interviewing skill acquisition for school-based personnel. *Advances in school mental health promotion, 7*(4), 240–254.
- Stevens, S. S. (1946). On the theory of scales of measurement. *Science, 103*, 677–680
- Vasilaki, E. I., Hosier, S. G., & Cox, W. M. (2006). The efficacy of motivational interviewing as a brief intervention for excessive drinking: A meta-analytic review. *Alcohol and alcoholism, 41*(3), 328–335.
- Wain, R. M., Kutner, B. A., Smith, J. L., Carpenter, K. M., Hu, M. C., Amrhein, P. C., & Nunes, E. V. (2015). Self-report after randomly assigned supervision does not predict ability to practice motivational interviewing. *Journal of substance abuse treatment, 57*, 96–101.
- Wallace, L., & Turner, F. (2009). A systematic review of psychometric evaluation of motivational interviewing integrity measures. *Journal of teaching in the addictions, 8*(1–2), 84–123.

Bilaga A

Kodningsprotokoll för variabler i MITI 4.2.1



Kodning av behandlares yttranden i Motiverandesamtal (MITI 4.2.1)

Samtal: Kodat segment: 00:00-20:00 (min) Kodare: MIC Lab

Målbeteende:

Datum:

Övergripande skattningar

Tekniska komponenter		Skala 1-5
	Främja förändringstal	
	Dämpa bibehållandetal	
Relationskomponenter		Skala 1-5
	Partnerskap	
	Empati	

Beteendefrekvens

		S:a
Ge information		
Övertala med tillstånd		
Bekräfta		
Söka samarbete		
Betona autonomi		
	Summa Förenliga med MI	
Övertala		
Konfrontera		
	Summa Oförenliga med MI	
Reflektioner	Enkla reflektioner	
	Komplexa reflektioner	
	Summa reflektioner	
Frågor		

Relationell = (≥3.5)

Teknisk = (≥3.0)

Komplexa reflektioner = (≥0.4)
Enkla + Komplexa reflektioner

Enkla + Komplexa reflektioner = (≥1.1)
Frågor

Totalt antal Förenliga med MI =

Totalt antal Oförenliga med MI =

Kodarens/kodarnas kommentarer:

Övrigt:

Bilaga B

Beskrivning av skalstegen i de övergripande variablerna i MITI 4.2.1

Främja förändringstal				
Låg			Hög	
1	2	3	4	5
Rådgivaren fäster ingen tydlig uppmärksamhet mot eller visar någon preferens för klienttal som förespråkar förändring.	Rådgivaren uppmärksammar bara sporadiskt förändringsinriktat tal och missar ofta tillfällen att uppmuntra förändringstal.	Rådgivaren uppmärksammar ofta förändringsinriktat klienttal, men missar några tillfällen att uppmuntra förändringstal.	Rådgivaren uppmärksammar konsekvent förändringsinriktat klienttal och anstränger sig för att uppmuntra det.	Rådgivaren uppvisar tydliga och konsekventa ansträngningar för att öka djup och styrka, eller drivkraften framåt, i klientens förändringstal

Dämpa bibehållandetal				
Låg			Hög	
1	2	3	4	5
Rådgivaren svarar konsekvent på klientens yttranden på ett sätt som främjar frekvensen eller djupet i argument för att bibehålla status quo.	Rådgivaren väljer generellt sett att utforska, fokusera på och svara på klienttal som förespråkar status quo.	Rådgivaren fäster större uppmärksamhet vid klienttal som förespråkar status quo men flyttar vid några tillfällen fokus bort från status quo.	Rådgivaren undviker vanligen att ge tonvikt åt klienttal som förespråkar status quo.	Rådgivaren uppvisar tydliga och konsekventa ansträngningar för att minska djupet, styrkan eller drivkraften hos klienttal som förespråkar status quo

Partnerskap				
Låg			Hög	
1	2	3	4	5
Rådgivaren intar aktivt en expertroll under merparten av interaktionen med klienten. Samarbete eller partnerskap saknas.	Rådgivaren tillvaratar samarbetsmöjligheter endast på ett ytligt plan.	Rådgivaren inkorporerar klientens bidrag men gör det på ett halvdant eller sporadiskt sätt.	Rådgivaren främjar samarbete och maktindelning på ett sätt som gör att klientens bidrag inverkar på samtalet på ett sätt som de annars inte skulle ha gjort.	Rådgivaren främjar och uppmuntrar maktindelning i samtalet på ett sådant sätt att klientens bidrag på ett påtagligt sätt påverkar samtalets karaktär.

Empati				
Låg			Hög	
1	2	3	4	5
Rådgivaren ger klientens perspektiv liten eller ingen uppmärksamhet.	Rådgivaren gör sporadiska försök att utforska klientens perspektiv. Rådgivarens förståelse kan vara felaktig eller skilja sig från vad klienten verkligen menade.	Rådgivaren försöker aktivt förstå klientens perspektiv med begränsad framgång.	Rådgivaren gör aktiva och upprepade försök att förstå klientens utgångspunkt. Demonstrerar korrekt förståelse för klientens världsbild, vilken dock mestadels är begränsad till det som uttryckts explicit.	Rådgivaren demonstrerar en djup förståelse för klientens utgångspunkt, vilken inte begränsas till det som uttryckligen sagts utan också omfattar det klienten menade men ännu inte har uttalat.