



UMEÅ UNIVERSITET

# **PSYKOMETRISK UTVÄRDERING AV INSTRUMENT FÖR ADL-FÖRMÅGA HOS PATIENTER MED PSYKOSSJUKDOMAR**

Yar Shakely, Disa Williamsson

Examensarbete, 30 hp  
Psykologprogrammet, 300 hp  
Vt 2018

Handledare: Steven Nordin

Tack till Daniel Abrams och Maria Wikeby vid Kompetenscentrum för Schizofreni och Psykiatri Psykos, Sahlgrenska Universitetssjukhuset i Göteborg, för samarbetet och möjligheten att delta i projektet. Tack till Steven Nordin, Ingrid Schéle och Tova Stenlund vid Umeå universitet för er hjälp och era värdefulla synpunkter.

### Abstrakt

Syftet var att utvärdera instrumentet Min Förmåga genom att undersöka 1) interbedömarreliabilitet och 2) konvergent validitet. Min Förmåga utarbetades för att möjliggöra enklare bedömningar av Aktiviteter i Dagligt Liv (ADL) på inneliggande patienter i psykiatri. Delarna i instrumentet som undersöktes innehåller fasta svarsalternativ och motiveringar i fritext. Deltagarna var 36 patienter med psykosjukdomar. Varje patient bedömdes först av två olika skötare med Min Förmåga och därefter med de liknande instrumenten Functional Independence Measure (FIM) och Global Assessment of Functioning (GAF). De fasta svarsalternativen analyserades med Cohens kappas för interbedömarreliabilitet och överensstämmelsen av motiveringarna analyserades kvalitativt. För konvergent validitet undersöktes korrelationen mellan Min Förmåga och FIM och GAF med Spearmans r. Resultatet visade hög interbedömarreliabilitet och hög konvergent validitet för de fasta svarsalternativen, samt hög grad av överensstämmelse mellan motiveringarna i fritext. En möjlig orsak till den höga interbedömarreliabiliteten kan vara att det endast fanns två fasta svarsalternativ, vilket tidigare forskning har visat kan leda till en överskattning av interbedömarreliabiliteten. Att FIM och GAF liknade Min Förmåga och administrerades under samma tidsperiod kan vara en orsak till den höga konvergenta validiteten. Då detta var en preliminär undersökning och antalet deltagare var bristfälligt bör fler utvärderingar av Min Förmåga göras i framtiden.

### Abstract

The aim was to evaluate the instrument Min Förmåga by examining 1) inter-rater reliability and 2) convergent validity. The purpose of Min Förmåga is to make assessments of Activities in Daily Life (ADL) on psychiatric inpatients easier. The participants were 36 patients with psychotic disorders. Each patient was first assessed by two different assessors with Min Förmåga and then with the similar instruments Functional Independence Measure (FIM) and Global Assessment of Functioning (GAF). Min Förmåga had both fixed response options and responses in free text. The fixed response options were analyzed with Cohen's kappa for inter-rater reliability and the responses in free text were analyzed qualitatively. The convergent validity was examined with Spearman's correlation between Min Förmåga and FIM and GAF. The result showed high inter-rater reliability and high convergent validity for the fixed response options, as well as high degree of consistency between the answers in free text. One possible reason for the high inter-rater reliability may be that there were only two fixed response options, as previous research has shown can lead to an overestimation of inter-rater reliability. That FIM and GAF resembled Min Förmåga and were administered during the same period of time can be a reason for the high convergent validity. As this was a preliminary examination and the number of participants inadequate, further evaluations of Min Förmåga should be made in the future.

## Psykometrisk utvärdering av instrument för ADL-förmåga hos psykospatienter

Instrumentet Min förmåga utarbetades som en del i verksamheten vid Psykiatri Psykos och Kompetenscentrum för Schizofreni vid Sahlgrenska universitetssjukhuset i Göteborg. Instrumentet utformades av Maria Wikeby, legitimerad arbetsterapeut, som uppfattade ett behov av ett instrument som skulle möjliggöra för skötare i psykiatri att göra bedömningar av funktionsförmåga på ineliggande patienter med psykosjukdomar. Utgångspunkten var att patientgruppen ofta har en mycket låg funktionsförmåga, men att det inte alltid uppmärksammas av personalen. Instrumentet avser mäta Aktiviteter i Dagligt Liv (ADL), ett mått för att mäta funktionsförmåga. Denna rapport avser utvärdera de psykometriska egenskaperna av instrumentet Min Förmåga.

Att bedöma funktionsförmåga i sjukvården är viktigt för att kunna undersöka hur patienten klarar vardagliga aktiviteter och de krav som ställs i hemmet, arbete, skola och sociala sammanhang (Världshälsoorganisationen, 2015). Patienter med samma sjukdom klarar olika saker och det är av betydelse att patienten själv, anhöriga och personalen har en bra bild av funktionsförmågan. Kunskap om patientens funktionsförmåga används för planering, behandling, utvärdering och prognos. Den är också ett underlag för att undersöka arbetsprestation, förmåner och planering för anpassning till samhället (Världshälsoorganisationen, 2015).

Det är svårt att definiera och mäta funktionsförmåga eftersom många aspekter av vardagligt liv och kontexten som personen ingår i berörs (Världshälsoorganisationen, 2015). Ett sätt att bedöma funktionsförmåga är genom att undersöka ADL (Sonn & Törnquist, 2016). Det finns ingen gemensam begreppsdefinition för ADL (Sonn & Törnquist, 2016; McDowell, 2006) men en definition är att ADL avser aktiviteter som återkommer i en persons boende, i kommunikation med andra och i vården. Kommunikation räknas in i begreppet eftersom det ofta är till grund för andra aktiviteter. Utifrån denna definition bör det först göras en bedömning av vad personen gör, och därefter om personen kan eller vill göra aktiviteten (Sonn & Törnquist, 2016). Historiskt så har ADL som koncept och instrument utvecklats från att beröra aktiviteter i dagliga livet för institutionaliserade patienter, men i takt med att trycket ökat på tidigare utskrivning av patienter har det utvecklats till att idag även inkludera aktiviteter i dagliga livet för människor delaktiga i samhället utanför institutioner (McDowell, 2006). Det förstnämnda brukar kallas personlig ADL, och innefattar personlig hygien och förflyttning, och det sistnämnda brukar kallas instrumentell ADL och innefattar aktiviteter som matlagning, städning, inköp och tvätt (McDowell, 2006). Analyser har visat att personlig ADL och instrumentell ADL faller under en och samma underliggande dimension. Instrumentell ADL reflekterar dock en något högre nivå av funktion än personlig ADL då aktiviteterna ofta också är avhängiga kognitiva förmågor och sociala roller (Spector & Fleishman, 1998). Instrument som avser mäta personlig ADL reflekterar ofta en relativt allvarlig brist i funktionsförmåga, exempelvis frågor som avser mäta blåsfunktion och förflyttning. I utformandet av ett ADL-instrument är det viktigt att utgå från gruppens förväntade funktionsförmåga så att måtten är känsliga för variationer (McDowell, 2006).

Social förmåga beskrivs av McDowell (2006) som en viktig del i en persons rehabilitering framförallt sedan trycket på snabb utskrivning från institution ut till samhället har ökat. Denna förändring har gjort att man börjat studera social förmåga hos patienter som är inlagda och ska skrivas ut, t ex för äldre patienter (McDowell, 2006).

Schizofreni och schizofreniliknande tillstånd är sjukdomar som innebär en förändrad verklighetsuppfattning och nedsatt social och kognitiv förmåga. I de nationella riktlinjerna för schizofreni och schizofreniliknande tillstånd beskrivs att den nedsatta kognitiva förmågan ofta är angränsande till autism och autismliknande tillstånd (Socialstyrelsen, 2017), det vill säga svårigheter med social interaktion. Till schizofreniliknande tillstånd omfattas schizofreniform

psykos, schizoaffektivt syndrom, kroniska vanföreställningssyndrom, korta och övergående psykotiska syndrom samt ospecificerad ickeorganisk psykos. I Sverige insjuknar varje år 1 500–2 000 personer i psykos. Cirka 30 000–40 000 personer i Sverige har diagnosen schizofreni, och inom den slutna psykiatriska vården har en stor andel av patienterna som vårdas denna sjukdom (Socialstyrelsen, 2017). Schizofreni listas som en av de tio mest funktionshindrande sjukdomarna i västvärlden (Världshälsoorganisationen, 2001) och är den vanligaste diagnosen bland alla psykossjukdomar (Socialstyrelsen, 2017). Forskning har visat att medellivslängden för personer med schizofreni är 15 år kortare för män och 12 år kortare för kvinnor jämfört med övriga befolkningen (Socialstyrelsen, 2017). Detta kan förklaras med att personer med psykisk ohälsa tycks få sämre somatisk vård än befolkningen i övrigt (Socialstyrelsen, 2013). Svårigheterna vid schizofreni leder ofta till försämrade funktionsförmåga. Exempelvis är det vanligt att personer med schizofreni eller schizofreniliknande tillstånd har svårt att klara av ett självständigt hemliv (Socialstyrelsen, 2017). Tillståndet har i tidigare studier kopplats samman med begränsningar i ADL-förmåga (Godbout et al., 2007; Hsieh et al., 2011; Viertiö et al., 2011). I en nationell utvärdering av vård och insatser vid depression, ångest och schizofreni (Socialstyrelsen, 2013) identifieras förbättringsområden, bland annat i att kommunerna bör förbättra möjligheterna till stöd i eget boende och öka den arbetsinriktade rehabiliteringen.

Det är särskilt viktigt att ADL-instrument som används kliniskt och som kan påverka personens fortsatta vård och rehabilitering lämpar sig utifrån sensitivitet och specificitet, har hög interbedömarreliabilitet, är lättadministrerat och etiskt försvarbart (Sonn & Törnquist, 2016). För alla tester som avser mäta beteenden är det viktigt att undersöka huruvida testresultat är tillförlitligt. Ju större konsekvenser bedömningarna utifrån ett testresultat har för individen, desto högre krav ställs det på testets tillförlitlighet. Testresultat som används för att fatta betydelsefulla beslut för individen kallas för högriskbedömningar och kräver sålunda högre tillförlitlighet än lågriskbedömningar (EFPA, 2013). För att undersöka ett tests tillförlitlighet kan det utvärderas psykometriskt, det vill säga en utvärdering av testets reliabilitet och validitet. Det kan göras med både kvalitativ som kvantitativ metod. I en psykometrisk undersökning samlas bevis för reliabilitet och validitet för att visa på testets tillförlitlighet. Med reliabilitet avses konsistens och precision i resultatet av mätprocessen. För att ett test ska anses reliabelt bör det därför vara tillräckligt konsistent och tillräckligt fritt från mätfel (precision) för att vara användbart (Urbina, 2014). En typ av reliabilitetsbevis är överensstämmelse i testresultat mellan bedömare som kallas interbedömarreliabilitet. Denna tar hänsyn till mätfel på grund av skillnader i resultat mellan bedömare på grund av subjektivitet i bedömningen. Vissa skattningar påverkas av subjektivitet i bedömningarna trots att testet har explicita och detaljerade instruktioner till skattningen. Den vanligaste metoden för att uppskatta denna typ av mätfel är att låta två olika bedömare skatta samma test på samma testperson. Om bedömarena tränats i administrering av instrumentet av skaparen till det överskattas troligtvis interbedömarreliabiliteten jämfört med om bedömare varit nybörjare på testet (Urbina, 2014). Interbedömarreliabilitet påverkas av känsligheten i en variabel i ett instrument. Om en variabel endast har två skilda alternativ, och alternativen är grovt skilda, förväntas en hög interbedömarreliabilitet (McHugh, 2012).

Med validitet avses graden av bevis och teori som stödjer tolkningen av ett testresultat utifrån testets föreslagna användningsområde (Urbina, 2014). I en valideringsprocess samlas bevis för validitet. En typ av bevis för validitet är konvergent validitet, det vill säga bevis avseende likheten till andra instrument eller skalor som avser mäta samma konstrukt. En vanlig metod är undersöka korrelationer mellan testet i fokus och andra instrument som avser mäta samma konstrukt. Även korrelationer mot delskalor i andra instrument kan användas för ett sådant bevis. I en testgranskningsmodell från European Federation of Psychologists' Associations (EFPA) tillhandahålls kriterier för granskning av psykologiska bedömningsinstrument, skalor och formulär. Testgranskningsmodellen används bland annat vid

bedömning av instruments validitet samt reliabilitet. Enligt modellen anses en urvalsstorlek (n) på minst 100 som adekvat för en psykometrisk granskning av interbedömarreliabilitet och konvergent validitet. Modellen utgår från en femgradig skala beskriven med ord och poäng, från "ej tillämplig information" (0 poäng) till utmärkt (4 poäng). Enligt EFPA bör test som får 1 poäng (inadekvat) eller lägre endast användas mycket varsamt och av testadministratörer med hög kompetens inom området (EFPA, 2013).

Min Förmåga är det instrument som ska utvärderas i denna uppsats. Instrumentet avser mäta utförandeförmåga i ADL hos ineliggande patienter genom strukturerat samtal och observation. Instrumentet kan även användas i boendeformer där personal ser personen utföra ADL. Syftet med instrumentet är bland annat att ge en bild av personens ADL-förmåga, att personer som arbetar nära patienten ska bli medveten om vilka behov personen har, att personalen ska nå en ökad samsyn i det praktiska arbetet och i bemötandet samt att patienten ska få djupare kunskap och förståelse om sin ADL-förmåga. Instrumentet är indelat i tre delar. Den första delen består av en strukturerad intervju med patienten och en närstående eller personal, i den andra delen (del 2) observeras brukaren i personlig ADL, instrumentell ADL och social förmåga och i den tredje delen ska ett ställningstagande göras angående om behov av åtgärd föreligger i patientens nuvarande boendeform, utifrån de strukturerade samtalen och observationerna i instrumentet. Om testledaren bedömer att det föreligger ett behov av åtgärd uppmanas att detta lyfts och eventuellt utreds vidare. Slutligen ska testledaren diskutera resultatet med övrig personal för att kunna förbättra bemötande, träning/anpassning samt habilitering/rehabilitering (Wikeby, 2016).

Patienter med schizofreni eller schizofreniliknande tillstånd har ofta mycket begränsad funktionsförmåga (Världshälsoorganisationen, 2001; Godbout et al., 2007; Hsieh et al., 2011; Viertiö et al., 2011). Syftet med instrumentet Min Förmåga är att möjliggöra för skötare på avdelning att på ett enkelt sätt göra bedömningar av funktionsförmågan hos ineliggande psykospatienter. Instrument som avser mäta psykologiska beteenden bör vara tillförlitliga, och detta undersöks i psykometriska undersökningar (Urbina, 2014). Mot bakgrund av detta formuleras syftet med denna rapport.

Det övergripande syftet med denna rapport är att bidra till att öka förståelsen för ADL-förmåga vid psykosjukdomar genom att utvärdera ett instrument som använts kliniskt på den aktuella patientgruppen. Detta genom att undersöka reliabiliteten och validiteten i instrumentet Min Förmåga. För att besvara syftet formulerades följande frågeställningar:

1. Hur är interbedömarreliabiliteten i Del 2 och i item 3b i Del 3 i Min Förmåga?
2. Hur är den konvergenta validiteten i Del 2 i Min Förmåga?

Frågeställningarna har avgränsats till skattningar gjorda med observationer på ineliggande patienter med psykosjukdom utförda av skötare som har utbildats och handletts i instrumentet.

## Metod

### Deltagare

Deltagarna var 36 patienter som diagnostiserats med psykosjukdom. Vid mättillfället var de ineliggande vid en psykiatrisk slutenvårdsavdelning vid Sahlgrenska universitetssjukhuset i Göteborg. Urvalet av patienterna gjordes baserat på personalens bedömning om patienten skulle bli ineliggande mer än en vecka. Detta då skaparen av instrumentet Min Förmåga uppskattat att det tar ungefär en vecka att administrera.

## Material

Del 2 i Min Förmåga är utformad med två svarsalternativ per item: Utan anmärkning (UA) och Anmärkning (ANM). Del 2 består av totalt 18 item, grupperade i Personlig ADL, Instrumentell ADL och Social förmåga. Testledaren observerar och skattar UA eller ANM och ska därtill motivera vad anmärkningen gäller i fritext vid svarsalternativet ANM. UA kan skattas om patienten initierar och utför uppgiften självständigt, under rimlig tid och tillräckligt ofta. ANM skattas om patienten behöver stöd eller påminnelse för att genomföra aktiviteten eller om utförandet har andra brister. I Personlig ADL ska testledaren observera patientens förmåga att sköta sin personliga hygien: dusch, hårvård, tandvård, händer, fötter, hud, klädsel, sömnhygien och sömn. I Instrumentell ADL observeras patientens förmåga att sköta och ta vara på saker i sin närmiljö. Testledaren ska titta på sängplats, garderob/sängbord, matsal, dusch/toalett och besök på kiosk. I Social förmåga observeras hur patienten interagerar med andra, om personen kan föra enklare samtal samt om personen har förmåga att be om hjälp och ta emot information (Wikeby, 2016). Del 3 i Min Förmåga innehåller två item. Item 3a innehåller flervalsalternativ för notering av nuvarande boendeform, och item 3b (Åtgärd) består av de två fasta svarsalternativ "Ingen åtgärd. Nuvarande boende/stödinsatser är tillräckliga. Brukarens förmåga i aktivitetsutförande överensstämmer med de krav som ställs i nuvarande boende." respektive "Behov av åtgärd. Nuvarande boende/stödinsatser är otillräckliga. Patientens förmåga i aktivitetsutförande överensstämmer INTE med de krav som ställs i nuvarande boende."

För att besvara den andra frågeställningen skattades patienterna även med instrumentet Functional Independence Measure (FIM) och General Assessment of Functioning (GAF). FIM syftar till att åskådliggöra graden av funktionsnedsättning och därmed hur mycket stöd en person behöver för att kunna utföra grundläggande aktiviteter i det dagliga livet (Världshälsoorganisationen, 2015; Uniform Data Set for Medical Rehabilitation, 1996). FIM:s validitet och reliabilitet har undersökts fått goda resultat (McDowell, 2006). FIM är indelat i en motorisk skala och en kognitiv skala, vilket är en indelning som har stöd i faktoranalyser gjorda på instrumentet (Stineman, Shea, Jette, Tassoni, Ottenbacher, Fiedler, & Granger, 1996). Den motoriska skalan kan jämföras med de bästa inom ADL-instrument, däremot anses den kognitiva skalan (Communication and Social Cognition) ha låg sensitivitet (Davidoff, Roth, Haughton & Ardner, 1990), det vill säga att skalan har svag träffsäkerhet i att peka ut de som har de egenskaper som avses mätas. Delen som avser mäta Social Cognition har även visats ha lägst interbedömarreliabilitet (Brosseau & Wolfson, 1994). Instrumentet omfattar 18 områden, uppdelat i sex större domäner: Personlig vård, sfinkterkontroll, kortare förflyttningar, längre förflyttningar, kommunikation, social- och intellektuell funktion. Instrumentet kan administreras av all sjukvårdspersonal, men om ett område är för svårt för några bland sjukvårdspersonalen att skatta kan bedömningen delas mellan sjukvårdspersonal med olika yrken för att skapa bättre förutsättningar för bedömningen. Exempelvis kan en sjukgymnast utvärdera förflyttningar, och en logoped kan bedöma kommunikation (Uniform Data Set for Medical Rehabilitation, 1996). FIM utvecklades för att kunna administreras på vuxna rehabiliteringspatienter inom särskilt och ordinärt boende, men kan även användas på andra patientgrupper. Underlaget till bedömningen kan med fördel inhämtas genom direktobservation av personens förmåga, däremot kan trovärdiga rapporter om vad personen vanligtvis gör även inhämtas från patienten själv, annan personal, familj och vänner, samt från patientjournaler. Hur aktiviteterna utförs graderas sedan från 1–7, där 1 är helt beroende av hjälp och 7 helt oberoende av hjälp. Helt oberoende innebär att personen kan utföra aktiviteten självständigt, på ett säkert sätt och inom rimlig tid. Det är vad patienten faktiskt *gör* som bedöms i instrumentet, inte vad patienten borde kunna göra eller skulle kunna göra om omständigheterna vore

annorlunda (Världshälsoorganisationen, 2015; Uniform Data Set for Medical Rehabilitation, 1996).

GAF är ett instrument som används för att bedöma psykologisk-, social- och arbetsrelaterad funktionsförmåga. Den består av en skala från 1–100, där 1 utgörs av den hypotetiskt sjukaste personen och 100 den hypotetiskt friskaste personen. Skalan i GAF är indelad i tio steg med beskrivningar för alla delar och inkluderar både beskrivningar av symtom och funktion. I en utvärdering av GAFs psykometriska egenskaper på patienter med schizofreni fick instrumentet höga resultat i den psykometriska granskningen förutom om GAF administrerades under aktiv psykos hos patienterna (Startup, Jackson, & Bendix, 2002).

## Design

För att kunna besvara uppsatsens frågeställningar undersöktes de psykometriska egenskaperna av instrumentet Min Förmåga genom en tvärsektionell design med olika bedömare.

## Procedur

ADL-förmågan hos 36 ineliggande patienter med psykossjukdom bedömdes med instrumentet Min Förmåga. Varje patient bedömdes av två skötare som gjorde sina skattningar under samma tidsperiod (en vecka). Syftet med att de genomfördes under samma tidsperiod var att minska risken att patientens ADL-förmåga skulle förändras. Skattningarna gjordes separat och utan kommunikation mellan skötarna om observationerna. Totalt utförde ungefär 10 skötare skattningar med Min Förmåga. De hade fått samma utbildning i instrumentet och hade möjlighet till handledning under hela administreringen. Direkt efter avslutad skattning med Min Förmåga gjorde en arbetsterapeut en bedömning av patientens funktionsförmåga med FIM och en sjuksköterska gjorde bedömningar med GAF. GAF skattades både utifrån funktionsförmåga (GAF Funktion), symtom (GAF Symtom) och det sammanlagda resultatet (GAF Total). Analysen av interbedömarreliabilitet delades upp i en statistisk analys och en kvalitativ analys då svarsalternativen i Del 2 utgörs av fasta svarsalternativ samt motiveringar i fritext.

**Statistisk analys av interbedömarreliabilitet.** För att besvara frågeställning 1 analyserades instrumentet Min Förmågas interbedömarreliabilitet. Interbedömarreliabilitet innebär att samstämmigheten mellan olika bedömares skattningar undersöks på itemnivå och/eller totalpoäng (Urbina, 2014). Varje patient observerades av två bedömare och för analyserna grupperades (slumpvis) en av skattningarna per patient till skattning A och den andra till skattning B.

För analysen av interbedömarreliabilitet i de fasta svarsalternativen i del 2 och i del 3 mellan skattning A och skattning B användes Cohens kapp eftersom data var kategorisk samt då måttet tar hänsyn till slump (Pallant, 2010). Interbedömarreliabiliteten analyserades med Cohens kapp på itemnivå för del 2 och del 3 och totalpoäng för del 2. Enligt EFPAs testgranskningsmodell bedöms kappvärden  $r < 0,60$  som inadekvat,  $0,60 \leq r < 0,70$  som adekvat,  $0,70 \leq r < 0,80$  som god och  $r \geq 0,80$  som utmärkt (EFPA, 2013). Signifikansnivån för de statistiska analyserna bestämdes till  $p < 0,05$ .

**Kvalitativ analys av interbedömarreliabilitet.** I ett andra steg i analysen av interbedömarreliabilitet bedömdes motiveringarna i fritext till de fall där båda skötarna hade skattat ANM. I ett fall (patient 24) drogs slutsatsen efter samtal med skaparen av instrumentet att motiveringarna i fritext var ogiltiga på grund av feladministrering. De togs därför inte med i analysen.

Fritextsvar finns ofta i kvantitativa enkäter för att kunna fånga kvalitativa aspekter (Ong, Dunn, & Croft, 2006). Ellonen, Fagerlund, och Pösö (2018) ställde frågan om det är etiskt

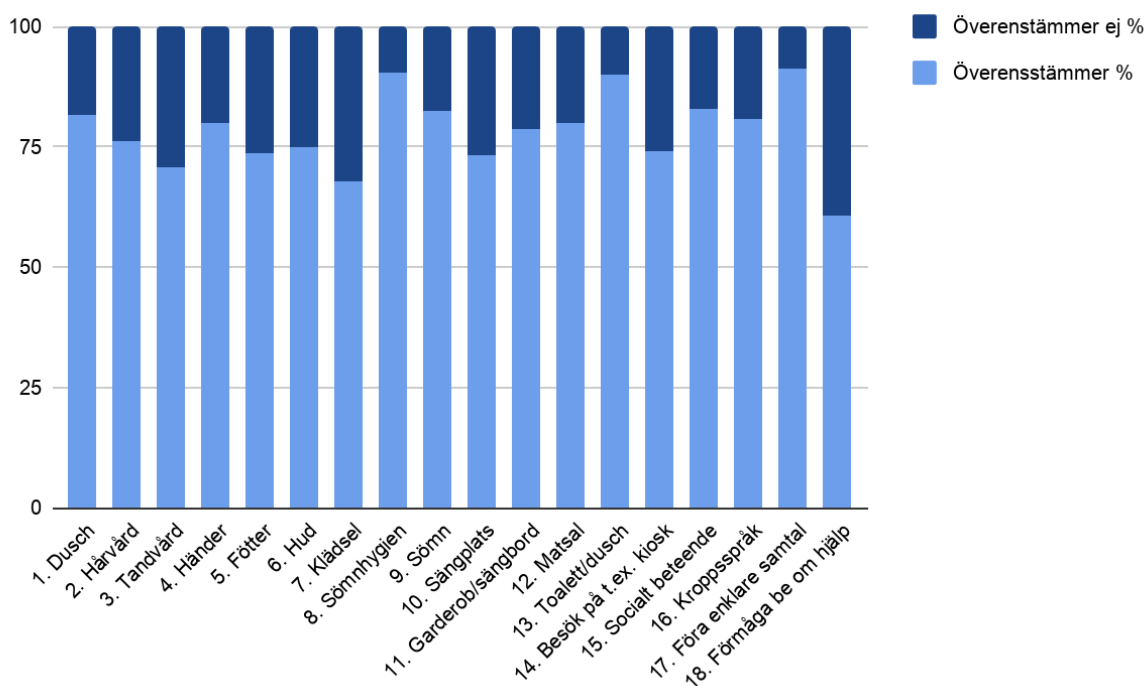


försvarbart att samla in data i form av fritextsvar om de inte avses analyseras. Deras studie visade att fritextsvar kan vara ett värdefullt verktyg för att utveckla och förbättra självskattningsformulär. Deras undersökning av fritextsvar genomfördes med induktiv tematisk analys. Två forskare kodade materialet enskilt och skapade därefter gemensamma teman utifrån koderna (Ellonen et al., 2018). Dibble, Baird, Swenson, och Healey (2017) genomförde en psykometrisk analys i kombination med en kvalitativ utvärdering av en enkät som avsåg mäta patienters nöjdhet med olika radiologiska undersökningar. Den kvalitativa undersökningen innebar att fritextsvaren i enkäten kategoriserades genom innehållsanalys. Forskarna grupperade svaren i teman (positiva-, negativa- och neutrala kommentarer) och jämförde dem sedan med den typ av radiologisk undersökning patienten hade genomgått. Den kvalitativa analysen i denna utvärdering tog utgångspunkt i ovan presenterade förfaranden och utformades därför som en kombination av tematisk analys enligt Braun och Clarke (2006) och ett mekaniskt klassificerande som närmade sig innehållsanalys (Hsieh & Shannon, 2005). Tematisk analys är en grundmetod i kvalitativ analys som används för att identifiera, analysera och rapportera mönster (teman) i ett kvalitativt datamaterial (Braun & Clarke, 2006). Innehållsanalys är en annan form av kvalitativ analys, vilket innebär att forskaren utgår från textdata och tolkar och skapar mening av den. Innehållsanalys kan variera mellan att vara strukturerad och rigid (kvantitativ) till tolkande och intuitiv (kvalitativ) (Hsieh & Shannon, 2005). Det går att dela in kvalitativ innehållsanalys i tre inriktningar: Konventionell, styrd eller summativ innehållsanalys. I summativ innehållsanalys ingår räknande och jämförelser och därefter tolkningar av den underliggande innebörden (Hsieh & Shannon, 2005). Att analysen låg i gränslandet mellan innehållsanalys och tematisk analys berodde på att kvalitativa data både kvantifierades (innehållsanalys) och tolkades och kategoriserades (tematisk analys) (Braun & Clarke, 2006; Hsieh & Shannon, 2005). Fokus för den kvalitativa analysen av interbedömarreliabilitet bestämdes till graden av överensstämmelse mellan motiveringarna, som för enkelhetens skull i rapporten kallas del A, respektive på vilket sätt motiveringarna var lika eller olika, som i rapporten kallas del B.

Inledningsvis kodades par av motiveringar i ett slumpmässigt stickprov utifrån på vilket sätt de var lika eller olika. Koderna var textnära och flexibla och gjordes efter diskussioner till mer snäva och abstrakta kategorier. Därefter, för att systematisera data, togs ett rutnät fram utifrån kategorierna som diskuterats fram. Rutnätet fick även tre huvudkategorier för att bedöma graden av överensstämmelse: Överensstämmer helt, överensstämmer delvis, överensstämmer ej. Med överensstämmer helt avsågs motiveringar där skattning A och skattning B beskriver samma sak, och med Överensstämmer delvis menas motiveringar där skattningarna till viss del beskriver samma sak, men där en del eller aspekt i motiveringen skiljer sig mellan motiveringarna. Överensstämmer ej avsåg par av motiveringar som talade om olika saker. Rutnätet fick även en fjärde huvudkategori för att kunna bedöma motiveringar som not applicable (N/A) då det i den initiala kodningen i stickprovet framkom motiveringar som bedömdes som detta. Huvudkategorin N/A innefattade dels par av motiveringar som saknade information eller hade oläslig handstil, dels motiveringar som bedömdes som icke-relevanta utifrån manualens beskrivning. Även de motiveringar där en bedömare hade skattat UA och den andra ANM analyserades utifrån relevans och inkluderades i del B för att få en bild av relevansen i samtliga motiveringar.

I rutnätet placerades huvudkategorierna i kolumnerna överst, med tillhörande underkategorier under dessa. I raderna placerades itemen. Därefter tolkades motiveringarna och patient-id skrevs in i cellerna i rutnätet efter bedömningen av i första hand graden av överensstämmelse (huvudkategorierna) och i andra hand på vilket sätt motiveringarna var lika eller olika (underkategorierna). Om ett par av motiveringar inte tycktes ha en passande underkategori bedömdes endast graden av överensstämmelse och bedömning av huvudkategori diskuterades senare, då underkategorierna modifierades löpande vid behov under hela denna

process likt tematisk analys (Braun & Clarke, 2006). Huvudkategorierna var däremot fasta då resultatet från dessa avsågs kvantifieras likt innehållsanalys (Hsieh & Shannon, 2005). Bedömningen utifrån rutnätet gjordes först separat av vardera av de två uppsatsförfattarna. Detta för att kunna undersöka tillförlitligheten i bedömningen av interbedömarreliabilitet i denna rapport. Överensstämmelsen mellan uppsatsförfattarnas separata bedömningar undersöktes genom att beräkna hur många procent av bedömningarna vad gäller huvudkategorierna som överensstämde respektive ej överensstämde (se Figur 1). Resultatet från den undersökningen varierade mellan 61–91 procent på itemnivå, och mättes till 78 procent på totalnivå. I de fall där uppsatsförfattarna hade gjort olika bedömningar diskuterades en gemensam slutgiltig bedömning fram. Även slutgiltiga underkategorier för att fånga upp på vilket sätt motiveringar var lika eller olika diskuterades fram. Analysen av interbedömarreliabilitet för motiveringarna i fritext mellan skattning A och skattning B gjordes därefter utifrån den gemensamma slutgiltiga bedömningen som diskuterats fram. För att likt innehållsanalys (Hsieh & Shannon, 2005) kvantifiera våra kvalitativa analyser av graden av överensstämmelse (del A) räknades fördelningen över huvudkategorierna ut i procent. Utgångspunkten var att ju högre grad av överensstämmelse, desto högre grad av interbedömarreliabilitet mellan motiveringarna. För instrument som används till bedömning inom hälso- och sjukvården rekommenderas en överensstämmelse i procent på minst 80 procent (McHugh, 2012).



Figur 1. Uppsatsförfattarnas överensstämmelse i bedömningarna av motiveringar i fritext.

**Statistisk analys av konvergent validitet.** För att besvara frågeställning 2 analyserades instrumentet Min Förmågas konvergenta validitet. Konvergent validitet innebär att korrelationen mellan två instrument eller delskalor som avser mäta liknande begrepp undersöks (Urbina, 2014). Den konvergenta validiteten i skattningarna av Min Förmåga undersöktes mot instrumenten FIM och GAF då dessa också avser mäta ADL-Förmåga/funktion. Delskalorna i Min Förmåga matchades mot delskalorna i FIM genom att undersöka instrumentens respektive manualer. Utifrån en bedömning att skalorna avsåg mäta liknande begrepp matchades delskalan *Personlig ADL* mot Self Care (skalor från Min Förmåga i kursivt) samt *Social förmåga* mot Communication och Social Cognition. En matchning

gjordes även på itemnivå efter en bedömning utifrån båda instrumentens manualer. Det resulterade i denna matchning (item från Min Förmåga i kursivt): *Dusch* mot *Dusch/Bad*; *Hårvård*, *Tandvård* och *Händer* mot Övre toalett; *Matsal* mot Äta/Dricka; *Besök på t.ex. kiosk* mot Problemlösande och *Föra enklare samtal* mot Socialt samspel. Genom lottning slumpades skattning A från Min Förmåga fram att göra mätningarna mot. Därefter undersöktes den konvergenta validiteten mellan Min Förmåga (skattning A) och FIM på totalpoäng, skalnivå och itemnivå med Spearmans rangkorrelation. Även den konvergenta validiteten mellan totalpoäng på Min Förmåga (skattning A) och GAF Funktion, GAF Symtom och GAF Total analyserades med Spearmans rangkorrelation. Spearmans rangkorrelation användes då data var icke-parametrisk (Pallant, 2010). Enligt EFPAs testgranskningsmodell bedöms korrelationer  $r < 0,55$  som inadekvat,  $0,55 \leq r < 0,65$  som adekvat,  $0,65 \leq r < 0,75$  som god och  $r \geq 0,75$  som utmärkt (EFPA, 2013). För instrument som är mycket lika varandra och där datainsamlingen med instrumenten skett vid samma tillfälle anses korrelationer på 0,60 eller mer adekvata. Signifikansnivån för de statistiska analyserna bestämdes till  $p < 0,05$ .

### **Etiska överväganden**

Instrumentet Min Förmåga utarbetades som en del i verksamheten vid Psykiatri Psykos och Kompetenscentrum för Schizofreni vid Sahlgrenska universitetssjukhuset. Reliabilitets- och validitetsundersökningen av Min Förmåga var en del i ett pågående forskningsprojekt. Data samlades som en del av den ordinarie vården och journalfördes. Datafilen och papperskopior avidentifierades innan författarna fick tillgång till dem för att säkerställa patienternas anonymitet. Författarna har varken träffat patienterna eller haft tillgång till deras journaler.

## Resultat

### Statistisk analys av interbedömarreliabilitet

I Tabell 1 presenteras resultatet från de statistiska analyserna av interbedömarreliabilitet för de fasta svarsalternativen UA och ANM mellan skattning A och skattning B i Min Förmåga.

Tabell 1

*Resultat från korrelationsanalys av interbedömarreliabilitet för de fasta svarsalternativen UA och ANM mellan skattning A och B.*

Min Förmåga	Överensstämmelse mellan skattning A och B, Kappa Koefficient	n
<u>Del 2</u>		
1. Dusch	1,00**	36
2. Hårvård	1,00**	36
3. Tandvård	1,00**	34
4. Händer	0,94**	36
5. Fötter	1,00**	36
6. Hud	0,94**	36
7. Klädsel	0,93**	36
8. Sömnhygien	0,93**	35
9. Sömn	1,00**	35
10. Sängplats	1,00**	36
11. Garderob/sängbord	1,00**	36
12. Matsal	1,00**	35
13. Toalett/dusch	0,94**	35
14. Besök på t ex. kiosk	1,00**	35
15. Adekvat socialt beteende	0,87**	36
16. Adekvat kroppsspråk	0,93**	36
17. Föra ett enklare samtal	0,70**	36
18. Förmåga att be om hjälp	0,87**	36
Total	0,83**	33
<u>Del 3</u>		
3b. Åtgärd	0,78**	24

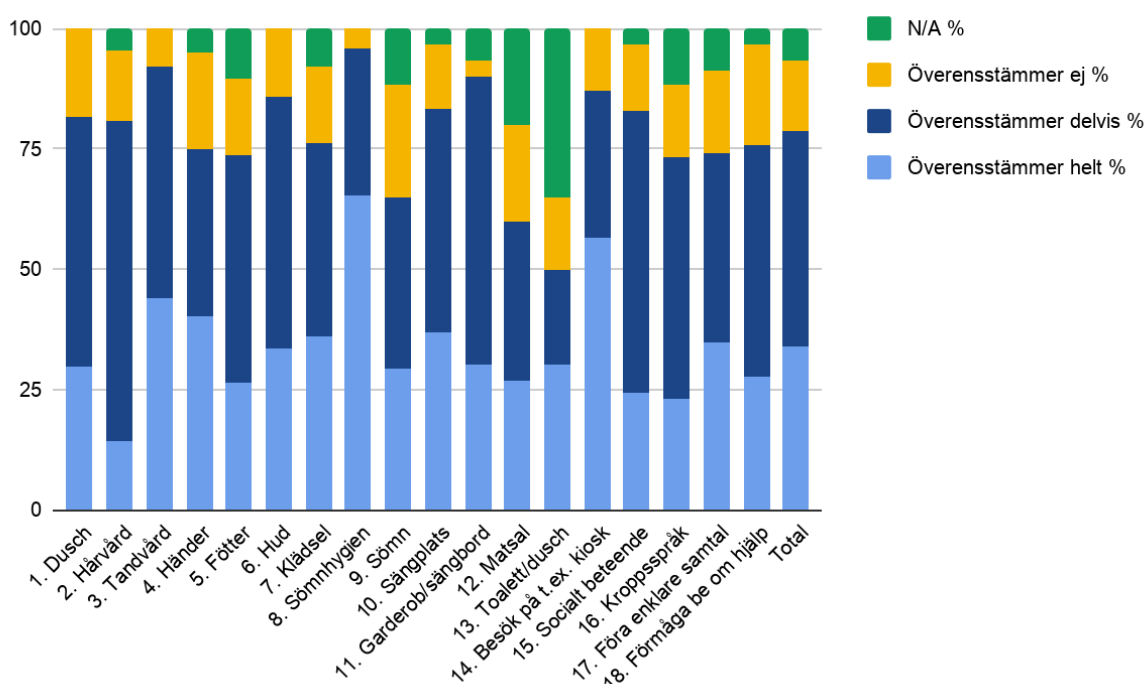
\* $p < 0,05$ , \*\* $p < 0,001$

Värdet på Cohens kappa varierar mellan 0,70 till 1,00 för samtliga item och för totalpoäng (del 2) mättes r till 0,83.

### Kvalitativa analyser av interbedömarreliabilitet

I Figur 2 nedan presenteras resultatet från del A av de kvalitativa analyserna av interbedömarreliabilitet i motiveringarna i fritext från Min Förmåga Del 2. Resultatet av analyserna presenteras i procent. Totalt för alla item var överensstämmelsen 33,7 procent för Överensstämmer helt, 44,9 procent för Överensstämmer delvis, 14,8 procent för

Överensstämmer ej, 6,6 procent för N/A samt 78,6 procent för summan av Överensstämmer helt och Överensstämmer delvis, det vill säga par av motiveringar som antingen beskrev samma sak, eller delvis beskrev samma sak. På itemnivå sticker item Sömnhygien ut då knappt två tredjedelar av motiveringarna (65,2%) hamnade under Överensstämmer helt och nästan alla (95,7%) par av motiveringarna i antingen Överensstämmer helt eller Överensstämmer delvis. Även itemet Besök på t.ex. kiosk sticker ut med resultatet 56,5 procent Överensstämmer helt och 87,0 procent Överensstämmer helt eller Överensstämmer delvis. Item Sömn sticker ut med högst procent för Överensstämmer ej, knappt en fjärdedel av paren av motiveringar (23,5%). För exempel på motiveringar tillhörande varje grad av överensstämmelse och mer utförliga beskrivningar av varje underkategori, se den tematiska kartan i Tabell 2. Antal par av motiveringar i fritext varierade mellan 3–15 par för Överensstämmer helt, 4–18 par för Överensstämmer delvis, 1–6 par för Överensstämmer ej samt 0–7 par för N/A.



Figur 2. Resultat från kvalitativ analys av interbedömarreliabilitet för motiveringarna i fritext mellan skattning A och skattning B.

Fokus i del A av den kvalitativa analysen var grad av överensstämmelse. Detta kombinerades med del B som undersökte på vilket sätt motiveringarna överensstämde eller inte, det vill säga likheter och skillnader mellan motiveringarna i fritext. I Tabell 2 presenteras en översikt av resultatet från del B med exempel på motiveringar tillhörande varje underkategori. Tabell 2 följs av en ingående beskrivning av resultat från del B.

Tabell 2

Tematisk karta över likheter och skillnader mellan motiveringarna i fritext mellan skattning A och skattning B.

Huvudkategori	Underkategorier	Exempel på motiveringar
<b>Överensstämmer helt</b>	Samma på samma sätt	(10, Sömnhygien) Sover på golvet; Sover oftast på golvet
	Samma på olika sätt	(22, Sängplats) Obäddat; Bäddar ej
<b>Överensstämmer delvis</b>	En mer i detalj	(15, Garderob/sängbord) Rörigt; Lite rörigt, kläder och koppar
	En bedömare ger ytterligare information	(6, Adekvat kroppsspråk) Hasande gång; Hasande gång, viftar
	Olika fokus	(7, Dusch) Behöver brytas; Duschar väldigt länge
	Behov respektive förmåga	(19, Sömn) Behöver sömntablett?; Är uppe på nätterna, tänder lampor
<b>Överensstämmer ej</b>	Olika – behov respektive förmåga	(21, Hud) Torr; Vill inte ha hjälp
	Olika motiveringar	(22, Klädsel) Kläder ut och in eller bak och fram; Många tröjor på sig samtidigt, smutsiga
N/A	En eller båda ej relevanta	(9, Toalett/dusch) Behöver hjälp vid duschning; Handduk på golv, städar inte, spolar inte, slänger inte blöja
	En eller båda går ej att tyda	(5, Matsal) <i>[oläsligt]</i> ...bordsskick; Utan påstötning går det för det mesta
	En saknar motivering	(5, Händer) <i>[motivering saknas]</i> ; Påminnas

En stor del av de par av motiveringar i den kvalitativa analysen som klassificerades till huvudkategorin Överensstämmer helt hade endast skillnader av pragmatisk karaktär, det vill säga skillnader som beror på att ord och uttryck bör uppfattas utifrån kontexten (Braun & Clarke, 2006). Sålunda beskriver underkategorin Samma på olika sätt par av motiveringar där endast *sättet att uttrycka sig* skiljer sig men där innebörden av motiveringarna var densamma. Tillhörande huvudkategorin Överensstämmer helt var också underkategorin Samma på samma sätt som avser par av motiveringar där både sättet att uttrycka sig och innebörden av motiveringarna var detsamma, exempelvis motiveringarna “Sover på golvet” och “Sover oftast på golvet”.

Kategorin En mer i detalj som tillhör huvudkategorin Överensstämmer delvis beskriver par av motiveringar där skillnaden i överensstämmelse beror på att den ena bedömaren ger en mer detaljerad beskrivning av anmärkningen än den andra bedömaren. Ofta berodde det på att

bedömaren som var mer detaljerad beskrev anmärkningen mer konkret än den andra bedömaren, eller gav fler exempel (se exempel i Tabell 2).

Då de två motiveringarna innehöll en del som överensstämde helt, men där den ena motiveringen även innehöll ytterligare information, kategoriserades det som "En bedömare ger ytterligare information" tillhörande huvudkategorin Överensstämmer delvis. Motiveringarna hade delvis samma innebörd beträffande ett/flera fenomen i anmärkningen, men skiljde sig delvis på grund av att den ena motiveringen även beskrivit ett/flera *andra* fenomen i anmärkningen.

Generellt i både huvudkategorin Överensstämmer delvis och Överensstämmer ej återfanns skillnaden ofta i *fokus*. I de fall då par av motiveringar någorlunda beskriver samma problematik i förmåga men väljer att fokusera på olika saker kategoriserades detta som Olika fokus under huvudkategorin Överensstämmer delvis. Exempelvis, där ena bedömaren beskriver "Sover i stort sett större delen av dygnet" och den andre "Sover på dagen".

En annan typ av skillnad i fokus som visade sig vara vanligt förekommande var att ena bedömaren fokuserade på bristande förmåga i beskrivningen, och den andra på behov (av stöd eller hjälp). Dock talade de om samma problematik men ur olika synvinklar. Då det förekom ofta har det blivit en egen underkategori, Behov respektive förmåga, under huvudkategorin Överensstämmer delvis. För exempel se Tabell 2.

Par av motiveringar som hamnade under huvudkategorin Överensstämmer ej var sådana där motiveringarna inte talade om samma problematik eller aspekt (av ADL-området). Hit hör underkategorin Olika motiveringar, det vill säga motiveringarna som kort och gott beskrev olika, ibland även motstridiga, saker. Även i dessa par av motiveringar var underkategorin Behov respektive förmåga vanlig. Paren av motiveringar skiljde sig inte endast i att de talade om olika problematik, utan det fanns också en skillnad i vad som fokuserades på för respektive problem. Och återigen var skillnaden i fokus mellan fokus på behov respektive fokus på bristande förmåga.

Huvudkategorin N/A fick tre underkategorier: (1) en eller båda motiveringar ej relevanta, (2) en eller båda går ej att tyda och (3) en eller båda saknar motivering. Underkategorin som handlade om motiveringarnas relevans innehöll motiveringar som inte var relevanta enligt beskrivningen i manualen. Detta kunde bero på att de inte gav någon ytterligare information än att det är en anmärkning (t ex motiveringen "Fungerar ej") eller att motiveringar inte handlade om det som det specifika itemet avsåg mäta, där item Toalett/Dusch sticker ut. Se Tabell 2 för exempel.

### **Statistisk analys av konvergent validitet**

Korrelationerna mellan totalpoäng på Min Förmåga och respektive totalpoäng på FIM, GAF Total, GAF Funktion och GAF Symtom varierade mellan 0,65–0,77. Nedan, i Tabell 3, presenteras resultatet från korrelationsanalys av konvergent validitet mellan de fasta svarsalternativen i Min Förmåga och skattningarna gjorda med FIM.

Tabell 3

Resultat från korrelationsanalys av konvergent validitet mellan Min Förmåga och FIM.

Min Förmåga	FIM	Spearman's r	n
Totalpoäng	Totalpoäng	0,76**	31
Personlig ADL	Self Care	0,79**	31
Social Förmåga	Communication/social cogn.	0,45*	36
Dusch	Dusch/Bad	0,71**	36
Hårvård	Övre toalett	0,63**	36
Tandvård	Övre toalett	0,78**	34
Händer	Övre toalett	0,41*	36
Matsal	Äta/Dricka	0,48*	35
Besök på t.ex. Kiosk	Problemlösande	0,37*	35
Föra Enklare Samtal	Socialt Samspel	0,38*	36

\* $p < ,05$ , \*\* $p < ,001$

Resultatet från korrelationerna mellan Min Förmåga och GAF Total var  $r(31) = 0,68$ ,  $p < 0,001$ , mellan Min Förmåga och GAF Funktion  $r(31) = 0,77$ ,  $p < 0,001$  och slutligen mellan Min Förmåga och GAF Symtom  $r(31) = 0,65$ ,  $p < 0,001$ .

### Diskussion

Syftet med uppsatsen var att bidra till att öka förståelsen för ADL-förmåga vid psykosjukdomar genom att undersöka instrumentet Min Förmåga som använts kliniskt på den aktuella patientgruppen. Följande frågeställningar formulerades:

1. Hur är interbedömarreliabiliteten i Del 2 och i item 3b i Del 3 i Min Förmåga?
2. Hur är den konvergenta validiteten i Del 2 i Min Förmåga?

När det gäller den första frågeställningen indikerar resultatet från analysen med Cohens kappa på hög interbedömarreliabilitet för del 2 och del 3 i instrumentet Min Förmåga. Enligt EFPA's testgranskningsmodell bedöms kappavärden  $r < 0,60$  som inadekvat,  $0,60 \leq r < 0,70$  som adekvat,  $0,70 \leq r < 0,80$  som god och  $r \geq 0,80$  som utmärkt (EFPA, 2013). Utifrån detta bedöms resultatet på totalnivå (del 2) som utmärkt och på itemnivå (del 2 och del 3) som utmärkt för samtliga item med undantag för item 17 (Föra enklare samtal) och item 3b i del 3 (Åtgärd) som bedöms som goda. Däremot indikerar det låga n-värdet i resultatet för item 3b i del 3 ( $n = 24$ ) på att detta var ett item utelämnats av flera skötare, vilket skulle kunna bero på att bedömning om åtgärd uppfattats som svår att göra.

Resultatet från del A i den kvalitativa analysen av interbedömarreliabilitet tyder på att i 78,6 procent av paren av motiveringar i fritext beskrevs samma sak eller delvis samma sak. Detta kan tolkas som en god överensstämmelse mellan skattning A och B med tanke på att svaren var i fritext. För kvantitativa data rekommenderas en överensstämmelse på minst 80 procent för instrument som används till bedömning inom hälso- och sjukvården (McHugh, 2012). Fritextsvar lämnar större utrymme för variation i skattningarna jämfört med kvantitativa data och kan därför förväntas ha mindre överensstämmelse mellan bedömare. Resultatet tyder på att en dryg tredjedel av samtliga par av motiveringar överensstämmer helt, nära hälften av paren av motiveringarna överensstämmer delvis och knappt en sjättedel av paren av motiveringar har ingen överensstämmelse. På itemnivå tycks item Sömn ha lägre



interbedömarreliabilitet jämfört med övriga item, vilket kan bero på att observationen avser sömnrutiner, något som kan vara svårt att observera om skötaren arbetar under dagtid.

Del B i den kvalitativa analysen av interbedömarreliabilitet ämnade undersöka *på vilket sätt* skattning A och B överensstämde eller inte överensstämde. Resultatet från undersökningen visade på skillnader i språkbruk, så kallade pragmatiska skillnader som beror på att ord och uttryck påverkas av kontexten (Braun & Clarke, 2006). Skillnader i språkbruk kan ses som en naturlig del bland en grupp skötare som antagligen kommer från olika kulturer och med olika utbildning och erfarenheter. Skillnader i språkbruk hos skötare medför att tolkning av fritextsvar i Min Förmåga ställer höga krav på förmågan att förstå innebörden i svaren hos mottagaren.

Resultatet visade även på att skillnader i abstraktionsnivå, där ena bedömaren beskrev mer konkret och den andra mer abstrakt, samt skillnader i mängd information och hur detaljerad beskrivningen var. Detta kan tolkas som skillnad i subjektivitet. Manualen och instruktionerna till instrumentet ger förslag på konkreta motiveringar respektive vad som ska tas i beaktning vid skattning, men utrymmet för fritextsvaret är mycket begränsat. Detta kan förklara varför vissa skattare valt att skriva abstrakt istället för att ge konkreta exempel samt varför vissa beskrivningar innehåller mindre information. Resultatet visade även att vissa motiveringar inte gav ytterligare information än att det var en anmärkning. Detta kan också bero på upplevs platsbrist för fritextsvaret. Resultatet visade att item Toalett/Dusch ofta skattats med icke-relevanta svar då motiveringarna inte handlade om städning av toalett och dusch som itemet avsåg mäta. De icke-relevanta motiveringarna handlade istället om personlig hygien vad gäller toalettbesök och duschning, det vill säga personlig ADL istället för städning som har med instrumentell ADL att göra (McDowell, 2006).

Resultatet visade på skillnader mellan motiveringar i fokus på behov respektive fokus på bristande förmåga. Detta kan förklaras med att syftet med manualen är att både ge en bild av personens *ADL-förmåga* samt att personer som arbetar nära patienten ska bli medveten om vilka *behov* personen har. I instruktionerna till del 2 uppmuntras skattaren att observera vad personen gör, men också om det finns behov av "praktisk hjälp eller påminnelser" (Wikeby, 2016). Denna dubbelhet i syfte och instruktioner tycks ge upphov till subjektivitet i bedömningar. Slutligen visade resultatet också på att olika bedömare beskriver helt olika anmärkningar i fritext. Detta är förväntat då ju fritextsvar lämnar stort utrymme för variation och subjektivitet. Snarare kan det sammantagna resultatet från del A och del B tolkas som god interbedömarreliabilitet i motiveringarna i fritext då överensstämmelse mellan motiveringarna totalt sett noterats i stor del av bedömningarna (78,6%).

Interbedömarreliabilitet kan bero på kvaliteten på instrumentet och subjektivitet i bedömningarna (Urbina, 2014). En hög interbedömarreliabilitet kan bero på att bedömarna följt instruktionerna noga och att instrumentet har god kvalitet. Sammantaget, utifrån den bild som de kvalitativa och kvantitativa analyserna ger, tycks del 2 och del 3 i instrumentet ha hög interbedömarreliabilitet. Dock kan det i fritextsvaren i del 2 finnas risk för mätfel på grund av subjektivitet, något som diskuteras ovan. Den höga interbedömarreliabiliteten kan bero på att bedömarna tränats i administreringen av skaparen av instrumentet, haft möjlighet till handledning av samma person under observationsperioden, samt att de fasta svarsalternativen varit grovt skilda, faktorer som tenderar resultera i högre interbedömarreliabilitet (McHugh, 2006; Urbina, 2014). Det skulle också kunna bero på att bedömarna arbetade tillsammans i den dagliga omvårdnaden av patienterna, som skulle kunna innebära att de kommunicerade med varandra om patienternas mående och funktionsförmåga. Dock råder liknande förutsättningar på alla kliniker.

När det gäller frågeställningen om instrumentets konvergenta validitet visade resultatet sammanfattningsvis hög konvergent validitet i de statistiska analyserna. Som presenterat i introduktionen innebär konvergent validitet grad av överensstämmelse mellan olika instrument

eller delskalor som avser mäta samma begrepp (Urbina, 2014). Enligt EFPA (2013) är korrelationer på 0,60 eller mer adekvata om instrumenten är mycket lika varandra och datainsamlingen har skett vid samma tillfälle, vilket gäller för instrumenten i undersökningen. Korrelationer mellan 0,65 och 0,75 bedöms vara goda och korrelationer högre än 0,75 bedöms vara utmärkta (EFPA, 2013). Korrelationer som är lägre än 0,55 bedöms vara inadekvata enligt EFPA (2013).

Korrelationerna mellan totalpoängen från Min Förmåga och de liknande instrumenten FIM, GAF Total, GAF Funktion och GAF Symtom var goda till utmärkta (EFPA, 2013). Det mest betydelsefulla resultatet var korrelationerna mellan totalpoängen från Min Förmåga och totalpoängen från FIM samt GAF Funktion som var utmärkta (EFPA, 2013). Även korrelationerna mellan delskalorna Personlig ADL i Min Förmåga och Self care i FIM var utmärkta, medan korrelationerna mellan Social förmåga i Min Förmåga och Communication samt Social cognition i FIM var inadekvata (EFPA, 2013). För att få fördjupad förståelse för specifika item i Min Förmåga korrelerades de mot liknande item i FIM. Tandvård från Min Förmåga korrelerade utmärkt mot Övre toalett i FIM enligt EFPAs kriterier, medan korrelationen mellan Dusch i Min Förmåga och Dusch/bad i FIM var god enligt EFPA (2013). Korrelationen mellan Hårvård i Min Förmåga och Övre toalett i FIM var adekvat (EFPA, 2013). Slutligen påvisades inadekvata korrelationer mellan Händer i Min Förmåga och Övre toalett i FIM; Matsal i Min Förmåga och Äta/dricka i FIM; Kiosk i Min Förmåga och Problemlösande i FIM; samt Enklare samtal i Min Förmåga och Socialt samspel i FIM (EFPA, 2013). Däremot förväntades inte höga korrelationer mellan dessa item då beskrivningarna till itemen i respektive instrument inte matchar helt.

Den konvergenta validiteten för totalpoängen för instrumenten Min Förmåga mot FIM och GAF Funktion kan anses vara hög då korrelationerna var utmärkta enligt EFPA (2013). Detta går att förstå eftersom det är instrument som avser mäta liknande begrepp – ADL och funktionsförmåga. Att korrelationerna var något lägre mellan totalpoängen för Min Förmåga och GAF Symtom samt GAF Total går att förstå eftersom GAF Symtom avser mäta symtom och inte funktionsförmåga. Följaktligen blir även korrelationen mellan Min Förmåga och GAF Total lägre eftersom den inkluderar både GAF Funktion och GAF Symtom. På skalnivå var den konvergenta validiteten mer ojämn då den varierade från utmärkt till inadekvat beroende på vilka skalor som korrelerades mot varandra. Skalorna Personlig ADL i Min Förmåga och Self care i FIM hade utmärkt korrelation enligt EFPA, vilket skulle kunna gå att förstå utifrån att den delen av ADL-begreppet relativt sett är enklare att observera. Lägst konvergent validitet på skalnivå fick skalan Social förmåga i Min Förmåga när den korrelerades mot Communication och Social cognition i FIM. Orsaken till detta kan diskuteras utifrån flera aspekter. En tänkbar anledning kan vara att patientgruppen generellt har låg social förmåga (Socialstyrelsen, 2017), vilket skulle kunna göra det svårt att observera den sociala förmågan. Observationen görs dessutom av skötare som inte har någon specifik utbildning i att göra bedömningar av social förmåga. De andra itemen i Min Förmåga är dessutom inriktade på mer konkreta aspekter av ADL (exempelvis dusch, tandvård, klädsel och ordning på sängbord), vilket skulle kunna göra dem enklare att observera. Social förmåga innefattas dessutom inte i de vanliga definitionerna av ADL, även om FIM också har skalor som enligt manualen går att förstå avser mäta liknande begrepp. Enligt ADL-taxonomi bör kommunikation (vilket kan ses som en del i social förmåga) inkluderas i ADL-begreppet då det ofta ligger till grund för andra aktiviteter (Sonn & Törnquist, 2016). Dessutom anses delskalorna Communication och Social cognition i FIM ha låg sensitivitet (Davidoff et al., 1990) samt delskalan Social Cognition ha lägst interbedömarreabilitet av skalorna i FIM (Brosseau & Wolfson, 1994) något som kan förklara varför Social Förmåga korrelerar lägre till dessa och som ger ytterligare en bild av att de förmågor som avses mätas tycks vara svårare att mäta än exempelvis förmågor i personlig ADL.

Beträffande begränsningar och styrkor finns det i alla studier möjliga felkällor som kan uppkomma under arbetets gång, vilket gäller även den här utvärderingen. I detta fall har risken för felkällor främst gällt datainsamlingen och analyserna. Datainsamlingen inför denna utvärdering genomfördes i den kliniska miljön som en del av den ordinarie vården innan uppsatsförfattarna var en del av projektet. Skaparen av instrumentet, Wikeby, instruerade personalen i förfarandet och fanns tillgänglig för handledning och frågor, men hade inte total kontroll över hela datainsamlingen. En av de skötare som gjorde flest bedömningar medverkade under stora delar av datainsamlingen och hade löpande kontakt med Wikeby. Uppsatsförfattarna har vid flera tillfällen kommunicerat med både Wikeby och den skötaren för att utforska om datainsamlingen utfördes på ett tillförlitligt sätt, vilket verkar vara fallet.

När det gäller möjliga felkällor i analyserna har det funnits en stor utmaning i att utvärdera denna kategori av instrument som är mer anpassat för den kliniska vardagen snarare än för forskning och utvärdering. För att kunna analysera Min Förmåga har en kombination av etablerade analysmetoder använts – både kvantitativa och kvalitativa. Att skapa en lösning för detta ansågs vara av vikt eftersom denna kategori av kliniska instrument också behöver utvärderas. Dels då kliniskt användbara instrument kan vara utformade på detta sätt, dels då enkäter generellt ofta efterfrågar fritextsvar. Av metodologiska skäl är det vanligt att de sedan inte analyseras, vilket är problematiskt ur ett etiskt perspektiv (Ellonen et al., 2018; Ong et al., 2006). Att det inte fanns en färdig lösning innebar att uppsatsförfattarna tog fram en skraddarsydd kombination av analysmetoder och det är inte säkert att den kombinationen var den mest optimala.

I de kvalitativa analyserna användes procent överensstämmelse och till skillnad från Cohens kappas tar det inte hänsyn till slump (McHugh, 2012). En annan nackdel med att presentera procentuell överensstämmelse är att det är ett resultat från uppsatsförfattarnas subjektiva bedömningar, men att de kan ge sken av att vara objektiva mått. En annan begränsning i utvärderingen var att antalet deltagare som endast var 36 personer. Det är ett för litet antal deltagare för att kunna dra några egentliga slutsatser utifrån de statistiska beräkningarna. Utifrån EFPA:s (2013) testgranskningsmodell anses en urvalsstorlek på 36 deltagare inadekvat för en psykometrisk studie. Detta är därmed inte en slutgiltig psykometrisk utvärdering av Min Förmåga, utan i framtiden behövs mer forskning. Den här preliminära utvärderingen är dock av kliniskt värde då studiens resultat kan ge god information för vidare utveckling av instrumentet.

En metodologisk styrka med denna utvärdering av Min Förmåga är att instrumentet både utvecklades och utprovades i den kliniska miljön vilket ger hög ekologisk validitet, det vill säga tillämpbarhet i det verkliga livet (Borg & Westerlund, 2012). En annan styrka är att eftersom uppsatsförfattarna själva utvecklade kombinationen av analysmetoder utvärderades de löpande vilket medförde att det fanns utrymme för att anpassa dem till just detta instrument. Stor vikt lades vid noggrannhet i analyserna, däribland undersökningen av uppsatsförfattarnas överensstämmelse i bedömningarna av motiveringarna i fritext.

Vissa slutsatser kan dras om Min Förmåga baserat på resultaten. Forskning på ADL-instrument har visat att det är viktigt att instrument i klinisk verksamhet väljs utifrån sitt syfte, då resultatet ofta påverkar personens fortsatta vård och rehabilitering. Särskilt viktigt är att instrumentet lämpar sig utifrån sensitivitet och specificitet, har hög interbedömarreliabilitet, är lättadministrerat och etiskt försvarbart (Sonn & Törnquist, 2016). Denna undersöknings resultat tyder på att Min Förmåga har hög interbedömarreliabilitet och hög konvergent validitet för den aktuella patientgruppen i den aktuella kontexten. Utifrån undersökningens resultat kan dock ett antal förbättringsförslag ges. Att item 3b har utelämnats av flertalet bedömare är något som instrumentets utvecklare bör ta i beaktande. Det ifrågasätter nämligen itemets tillförlitlighet och huruvida det ska behållas som item i instrumentet. Att det utelämnats kan bero på att ställningstagandet om åtgärd eller ingen åtgärd i nuvarande boendesituation kan vara en alltför

svår bedömning att göra för någon som inte är socionom. Det kan också bero på att skötarna som gjort bedömningen inte har tillräckligt med information om det exakta stödet i det nuvarande boendet. I manualen till instrumentet Min Förmåga står det att om testledaren bedömer att det föreligger ett behov av åtgärd i boendet behöver detta lyftas och eventuellt utredas vidare. Detta kan tolkas som att beslut utifrån resultatet är en så kallad högriskbedömning och sålunda har högre krav på tillförlitlighet (EFPA, 2013).

Ytterligare en aspekt att fundera över är huruvida det är problematiskt utifrån instrumentets syfte att motiveringarna i fritext har en tendens att skilja sig mellan fokus på behov respektive fokus på bristande ADL-förmåga. Om syftet med instrumentet fortsatt ska vara att fånga båda aspekter bör instrumentets utformning ses över. En eventuell förklaring till skillnad i fokus skulle kunna vara begränsad plats för motiveringarna i fritext, som gör att testledaren inte noterar allt som ska noteras enligt instruktionerna. Sålunda vore en möjlig åtgärd i utformningen att öka utrymmet för motiveringarna i fritext. Åtgärden skulle också kunna tänkas få bukt med problematiken i interbedömarreliabilitet i fritextsvaren som berodde på att bedömarna gett olika mycket information eller beskrivit ett fenomen olika detaljerat.

Item 17 (Föra enklare samtal) under skalan Social förmåga tycktes ha lägst interbedömarreliabilitet i hela instrumentet. Även lägst konvergent validitet uppmättes för Skalan Social förmåga mot liknande skalor i FIM. Detta resultat kan ifrågasätta inkluderandet av Social Förmåga i instrumentet, dock menar McDowell (2006) att en viktig del i en persons rehabilitering och anpassning till samhället är just social förmåga, något vi uppsatsförfattare instämmer i.

Vad gäller att item Toalett/Dusch tycks ha missuppfattats av flertalet skötare som personlig ADL och inte instrumentell ADL är något som bör åtgärdas. Eventuellt skulle det behövas en tydligare beskrivning i instrumentets instruktioner för just detta item.

Slutligen vore ett förbättringsförslag för denna typ av instrument att utformningen av instrumentet tar hänsyn till metodologiska aspekter av psykometriska utvärderingar. En styrka i denna undersökning är som redan nämnt den ekologiska validiteten och att undersökningen utformats för att kunna utvärdera även fritextsvarens tillförlitlighet. Dock innebär kvalitativ analys av psykometriska egenskaper ett större arbete än kvantitativ analys och det är något som instrumentets utvecklare kan ta i beaktande. Ett förslag vore flervalsoalternativ på vanliga motiveringar istället för motiveringar i fritext.

Resultatet från denna psykometriska undersökning är värdefullt då det ger en helhetsbild utav instrumentets kvalitet. Undersökningen har även innefattat kvalitativ analys och analys på itemnivå, två viktiga delar i en utvärdering och förbättring av instrument. Tack vare detta har undersökningen kunnat belysa kvalitativa aspekter avseende exempelvis relevans i svaren och risk för subjektivitet i instrumentet, vilket är viktiga aspekter för instrumentets kvalitet. De kvantitativa analyserna har kunnat lyfta styrkor och svagheter i instrumentets olika delar, vilket är viktiga fynd för utvecklandet av instrumentet. Sammantaget är undersökningens resultat av stor vikt för den kliniska vardagen för patientgruppen då undersökningen bidrar till förbättringen av ett instrument som tillåter lättadministrerade bedömningar utförda av skötare. Då den aktuella undersökningens deltagarantal bedöms som bristfälligt enligt EFPA (2013) och undersökningen var den första att utvärdera instrumentet Min Förmågas psykometriska egenskaper bör framtida forskning fokusera på att göra fler utvärderingar och med ett större antal deltagare.

## Referenser

- Borg, E., & Westerlund, J. (2012). *Statistik för beteendevetare, (3d Edition)*. Malmö: Liber.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77-101.
- Brosseau, L., & Wolfson, C. (1994). The inter-rater reliability and construct validity of the Functional Independence Measure for multiple sclerosis subjects. *Clinical Rehabilitation, 8*(2), 107-115.
- EFPA (2013). *EFPA review model for the description and evaluation of psychological and educational tests: Test review form and notes for reviewers*. Brussels: European Federation of Psychologists' Association.
- Ellonen, N., Fagerlund, M., & Pösö, T. (2018). Free-text comments as a tool for developing the self-report method: Parents' responses to a survey on violence against children. *Australian & New Zealand Journal of Criminology, 51*(1), 58-75.
- Davidoff, G., Roth, E., Haughton, J., & Ardner, M. (1990). Cognitive dysfunction in spinal cord injury patients: Sensitivity of the Functional Independence Measure subscales vs neuropsychologic assessment. *Archives of Physical Medicine and Rehabilitation, 71*(5), 326-9
- Dibble, E. H., Baird, G. L., Swenson, D. W., & Healey, T. T. (2017). Psychometric Analysis and Qualitative Review of an Outpatient Radiology-Specific Patient Satisfaction Survey: A Call for Collaboration in Validating a Survey Instrument. *Journal of the American College of Radiology, 14*(10), 1291-1297.
- Godbout, L., Limoges, F., Allard, I., Braun, C., & Stip, E. (2007). Neuropsychological and activity of daily living script performance in patients with positive or negative schizophrenia. *Comprehensive Psychiatry, 48*(3), 293-302.
- Granger, C. V., Hamilton, B. B., Linacre, M. J., Heinemann, W. A., & Wright, D. B. (1993). Performance profiles of the functional independence measure. *American Journal of Physical Medicine and Rehabilitation, 72*(2), 84-89.
- Hsieh, H-F., & Shannon, S. E. (2005). Three Approaches to Qualitative Content Analysis. *Qualitative Health Research, 15*(9), 1277-1288.
- Hsieh, P. C., Huang, H. Y., Wang, H. C., Liu, Y. C., Bai, Y. M., Chen, K. C., & Yang, Y. K. (2011). Intercorrelations between the personal and social performance scale, cognitive function, and activities of daily living. *Journal of Nervous and Mental Disease, 199*(7), 513-515.
- McDowell, I. (2006). *Measuring health: A guide to rating scales and questionnaires* (3rd ed.). New York: Oxford University Press.
- McHugh, M. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica, 22*(3), 276-82.
- Ong, B. N., Dunn, K. M., & Croft, P. R. (2006). "Since you're asking...": Free text commentaries in an Epidemiological study of low back pain consulters in primary care. *Quality and Quantity, 40*(4), 651-659.
- Pallant, J. (2010). *SPSS survival manual: A step by step guide to data analysis using IBM SPSS* (4. uppl.). Maidenhead: Open University Press.
- Spector, W. D., & Fleishman, J. A. (1998). Combining activities of daily living with instrumental activities of daily living to measure functional disability. *The Journals of Gerontology, Series B, 53*(1), S46-57.
- Socialstyrelsen, 2017. *Nationella riktlinjer för vård och stöd vid schizofreni och schizofreniliknande tillstånd – Stöd för styrning och ledning – Remissversion*. Hämtad 2018-01-24, från <http://www.socialstyrelsen.se/Lists/Artikelkatalog/Attachments/20735/2017-10-34.pdf>

- Socialstyrelsen, 2013. *Nationell utvärdering 2013: vård och insatser vid depression, ångest och schizofreni: rekommendationer, bedömningar och sammanfattning*. Hämtad 2018-01-24, från <http://www.socialstyrelsen.se/Lists/Artikelkatalog/Attachments/19149/2013-6-7.pdf>
- Sonn, U., & Törnquist, K. (2016). *ADL-Taxonomi - en bedömning av aktivitetsförmåga*. Nacka: Sveriges Arbetsterapeuter.
- Startup, M., Jackson, M. C., & Bendix, S. (2002). The concurrent validity of the Global Assessment of Functioning (GAF). *British Journal of Clinical Psychology*, 41(4), 417-422.
- Stineman, M. G., Shea, J. A., Jette, A., Tassoni, C. J., Ottenbacher, K. J., Fiedler, R., Granger, C. V. (1996). The functional independence measure: Tests of scaling assumptions, structure, and reliability across 20 diverse impairment categories. *Archives of Physical Medicine and Rehabilitation*, 77(11), 1101-1108.
- Uniform Data Set for Medical Rehabilitation (1996). *Functional independence measure*. Buffalo, NY: State University of New York.
- Urbina, S. (2014). *Essentials of psychological testing* (2. uppl.). Hoboken, NJ: John Wiley & Sons.
- Viertö, S., Tuulio-Henriksson, A., Perälä, J., Saarni, S. I., Koskinen, S., Sihvonen, M., Lönnqvist, J., & Suvisaari, J. (2011). Activities of daily living, social functioning and their determinants in persons with psychotic disorder. *European Psychiatry*, 27(6), 409-415.
- Världshälsoorganisationen (2015). *Mätning av hälsa och funktionshinder: Manual till WHO:s formulär för bedömning av funktionshinder "WHO Disability Assessment Schedule" (WHODAS 2.0)*. Hämtad 2018-01-22, från <https://www.socialstyrelsen.se/Lists/Artikelkatalog/Attachments/19807/2015-5-1.pdf>
- Wikeby, M. (2016). *Min Förmåga. Aktivitetsutförande – Strukturerat samtal och observation*. Göteborg: Psykiatri Psykos, Sahlgrenska Universitetssjukhuset.
- Världshälsoorganisationen (2001). *The world health report 2001 - Mental health: New understanding, new hope*. Genève: Världshälsoorganisationen.