

**SAMMANFATTNING**

Så väl anamnes och samtal som tester och mätningar ligger till grund för fysioterapeutens funktionsdiagnos och val av åtgärder samt uppföljning. En del i detta är att kunna mäta subjektiva parametrar såsom förmågor, symtom och upplevelser på ett objektivt och standardiserat sätt för att få kunskap om patientens status. Det finns dock en rad utmaningar med denna typ av mätningar. Denna artikel belyser de svårigheter som finns med mätningar av förmågor, symtom och upplevelser samt visar på de möjligheter som finns att säkerställa mätkvalitet för tillförlitliga och jämförbara mätunderlag. Grundläggande mätprinciper introduceras via exempel på mätningar hämtade från fysioterapeutens vardag.

# Metrologi för att säkerställa kvalitet i fysioterapeutiska mätningar

**JEANETTE MELIN**

Leg. sjukgymnast, med.dr,  
forskare i mätteknik,  
Research Institutes of  
Sweden (RISE)

**MÄTNINGAR AV FÖRMÅGOR**, symtom och upplevelser är för de flesta en självklar del i fysioterapeutens kliniska arbete. Likaså får mätningar en allt större roll inom forskning, kvalitetsregister, nationella patientenkäten och brukarundersökningar. Med ett ökat tryck på personcentering och patientdelaktighet har kravet på att mäta patientens förmågor, symtom och upplevelser ökat markant sedan mitten av 90-talet (1, 2). Utvecklingen påbörjades dock långt innan dess, framför allt med mätningar inom psykologin (3).

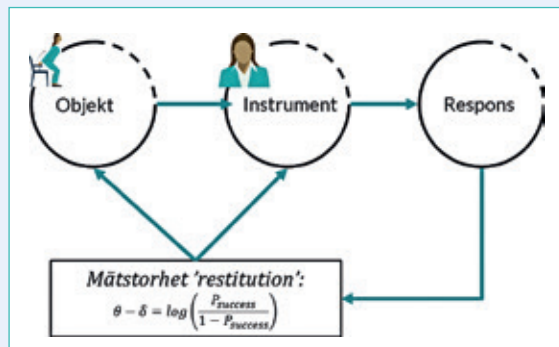
Det finns i dag ett stort antal frågeformulär för patienters egna utvärderingar. Exempel på sådana frågeformulär är *Patient Reported Outcome Measures* (PROM) och *Patient Reported Experience Measures* (PREM) (4). PROM innehåller exempelvis självskattningar av fysisk funktion, symtom samt emotionellt och socialt välbefinnande medan PREM är inriktat på erfarenheter av och syn på vårdande, rehabilitering samt hälso- och sjukvårdsinsatser. Primärt har syftet med PROM och PREM varit att fånga patientens egna perspektiv för att säkerställa att rätt insatser kan ges och att få ett objektivt mått på att hens hälsotillstånd förbättras (5). Tillsammans med PROM och PREM finns även de bedömningar som görs av fysioterapeuten eller annan vårdpersonal samt anhöriga (6).

Dessa typer av mätningar benämns överordnat kategoriskt baserade mätningar. Därtill inkluderar kategoriskt baserade mätningar även mått som visar om målvärden har uppfyllts på verksamhetsnivå och andra skattningar av erfarenheter utanför hälso- och sjukvårdens sammanhang (7).

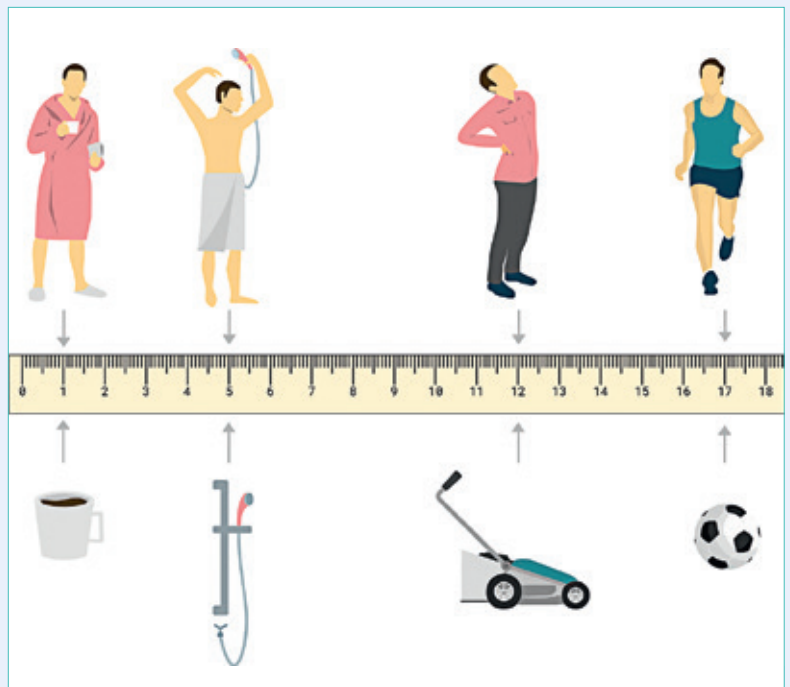
**Utmaningar med kategoriskt baserade mätningar**

En kategoriskt baserad mätning är en mätning som görs med ett begränsat antal möjliga resultat och det ger så kallad ordinal data (8). Exempel på ordinal data är utbildningsnivå, hur väl man utfört en aktivitet eller enkätsvar om subjektiva parametrar (exempelvis symptom eller vårdpersonalens bemötande). Det innebär att data kan rangordnas, men det finns inget tydligt numeriskt värde på alternativen eller skillnaderna dem emellan. Vid kategoriskt baserade mätningar ger man istället "klassificeringstal" för skattningar där man exempelvis låter 0 (noll) motsvara *håller inte alls med* och 4 (fyra) motsvara *håller med fullt ut*. Enligt den klassiska testteorin antar man att skalstegen mellan olika frågor i en enkät är lika stora (linjära) och av samma slag, vilket leder till att man summerar klassificeringstalen. Detta är dock felaktigt och problematiskt eftersom dessa i själva verket är icke-linjära data som varken

Figur 1a



Figur 1b



Figur 1a–b. I Figur 1a illustreras människan som mätinstrument och hur Rasch-formeln ger separata mätvärden för uppgiftens svårighetsgrad (objektet),  $\delta$ , och personens förmåga,  $\theta$ , (instrumentet) på en linjär skala (1a). I Figur 1b illustreras underst hur uppgifter, exempelvis duscha eller klippa gräset, har olika svårighetsgrader på en linjär skala, och överst hur olika personers förmågor kan mätas mot dessa referenser (1b).

lämpar sig för enklare beräkningar av summor eller medelvärden eller statistiska analyser.

Att samma frågor/enkät används över tid är en viktig faktor för att kunna följa hur en patients status utvecklas eller för att göra jämförelser mellan olika grupper. Detta är dock inte tillräckligt för att på ett tillförlitligt sätt kunna göra dessa jämförelser – här behövs även så kallade metrologiska referenser (9, 10). Det finns många globala initiativ som syftar till att utveckla metoder för att standardisera kategoriskt baserade mätningar. Tyvärr har de flesta inte tagit hänsyn till grundläggande metrologi och strukturer för att säkerställa tillförlitlighet och jämförbarhet (11).

De två huvudsakliga utmaningarna som nämns ovan är till stor del relaterade till analysarbetet, och där vet vi att tillförlitliga mätningar kräver att man tar hänsyn till grundläggande metrologiska principer (12, 13). Det finns givetvis ytterligare utmaningar, inte minst relaterade till den bias som kan uppstå vid självrapportering. Exempel på bias är att man över- respektive underrapporterar önskvärda beteenden eller att man bortser från att kvalitet ofta är en följd av flera verksamheters insatser (14). Dessa utmaningar adresseras dock inte i denna artikel.

### Metrologi – läran om mätningar

Metrologi är läran om mätning (15), inte att förväxla med meteorologi som är läran om vädret. Mätningar har funnits med oss sedan byggandet av tempel och pyramider och har sedan haft en avgörande betydelse när handeln utvecklades och därefter inom industrialiseringen. I dag har nästan allt runt om oss en koppling till mätningar, många gånger utan att vi är medvetna om det, så som i affärer, avlopp, vägar, elnät och transporter. Likaså har mätningar en central roll inom hälso- och sjukvården. Att ha tillförlitliga mätvärden spelar en (livs-)viktig roll i allt från screening och diagnostisering till behandlingsplanering och uppföljning.

För att säkerställa tillförlitlighet i mätningar finns två centrala begrepp, *metrologisk spårbarhet* och *mätosäkerheter*. Att mätvärden är jämförbara oberoende av vem som utför mätningarna och oavsett vilka metoder eller vilken utrusning som används är en grundförutsättning för att kunna följa en och samma patient över tid eller när man vill studera skillnader mellan grupper. För att erhålla fullt ut jämförbara mätvärden krävs spårbara metrologiska referenser (16). I och med SI-enheterna har vi i dag internationella referenser →

## Från ett mättekniskt perspektiv är det dock människan själv som ska ses som själva instrumentet.

→ för mätningar av längd, massa, tid, elektrisk strömstyrka, temperatur, substansmängd och ljusstyrka. Från SI-enheterna finns dessutom så kallade härledda enheter, exempelvis kraft, energi eller effekt. Det finns ännu inte motsvarande referenser för spårbara mätningar för kategoriskt baserade mätningar.

Det finns inte en enda mätning som är helt exakt, utan det finns alltid en gråzon kring mätvärdet (17). Mätosäkerhet är således ett mått på säkerheten i ett mätvärde. Exempelvis kan en patients kroppstemperatur vara  $37,5 \pm 0,1^\circ\text{C}$ , vilket något förenklat sett innebär att temperaturen med 95 procents sannolikhet ligger någonstans mellan  $37,4^\circ\text{C}$  och  $37,6^\circ\text{C}$ . I klinik uppmärksammas sällan mätvärdens osäkerheter i den utsträckning de borde. Detta gäller ibland även i forskning och kunskapsbildning, inklusive statistiska sammanställningar. Alltför ofta kan man se eller höra om signifikanta skillnader samtidigt som det inte är någon detekterbar skillnad då de individuella mätosäkerheterna runt respektive mätvärde överlappar varandra.

### Metrologi vid kategoriskt baserade mätningar

För att kvalitetssäkra kategoriskt baserade mätningar och säkerställa att vi får in tillförlitliga och jämförbara märesultat även vid självskattningar måste vi använda grundläggande metrologiska principer. Detta bör göras genom att kombinera modellen *Människan som mätinstrument* och *Rasch Measurement Theory* (13, 18).

#### Människan som mätinstrument

Att se människan som mätinstrument är inte bara illustrativt i linje med en personcentrerad vård utan också en förutsättning för att erhålla mätvärden för varje unik patient. Många gånger säger man att ses exempelvis en enkät är ett instrument för att mäta en patients förmågor, symtom eller upplevelser. Från ett mättekniskt perspektiv är det dock människan själv som ska ses som själva instrumentet (9). På samma sätt som det är vägen som ger en respons baserat på objektet som vägs är det människan (patienten) som responderar på en uppgift [figur 1a, se s. 27]. Ett exempel från fysioterapi skulle kunna vara när man använder

*Functional Independent Measure* (FIM) där patientens förmåga bedöms för 18 olika uppgifter på en skala mellan 1 och 7. Detta genererar klassificeringstal som behöver omvandlas till mätvärden för patientens förmåga. Denna omvandling görs med hjälp av Rasch-formeln.

#### Rasch Measurement Theory

På 1960-talet formulerade den danska matematikern och statistikern Georg Rasch en mätmodell för kategoriskt baserade mätningar, *Rasch Measurement Theory* (19). Med hjälp av denna teori kan ordinal data omvandlas till mätvärden enligt grundläggande mätkriterier. I sitt enklaste format är Rasch-formeln en logistisk regressionsfunktion:

$$\theta - \delta = \log \left( \frac{P_{\text{success}}}{1 - P_{\text{success}}} \right)$$

Formeln ger separata värden för personens förmåga,  $\theta$ , respektive uppgiftens svårighetsgrad,  $\delta$ , på samma linjära skala [figur 1b, se s. 27]. När man definierar olika uppgifters svårighetsgrad genom en Rasch-analys kan man erhålla metrologiska referenser, och på så sätt möjliggörs att patientens förmåga blir jämförbar, exempelvis över tid eller mot andra (11). Olika uppgifters svårighetsgrad,  $\delta$ , kan likställas med en kalibrerad viktsats (9). På samma sätt som man har vikter för 10 g, 50 g, 100 g etc. kan olika uppgifters svårighetsgrad definieras på en linjär skala i förhållande till varandra. Uppgifters svårighetsgrad definieras omvänt från personers förmåga, det vill säga, genom att olika personer får skatta uppgifterna kan respektive uppgifts svårighetsgrad estimeras med hjälp av Rasch-formeln.

Med hjälp av Rasch-analysen utvärderas huruvida kriterier för om en unidimensionell mätning kan erhållas, det vill säga *invariants* i mätningar (19–21). *Invariants* innebär att till exempel en linjal fungerar likadant, oavsett vem som använder den, var den används eller var på linjalen man mäter eller att en köksvåg och en industrivåg responderar likvärdigt på samma kalibrerade vikt. För kategoriskt baserade mätningar, när man likställer ett frågeformulär med ett set av kalibrerade vikter, innebär det att en fråga i ett frågeformulär fungerar lika exempelvis för olika grupper. Det motsatta skulle exempelvis kunna vara att magont betyder olika saker för pojkar och flickor (vilket visats med hjälp av *Differential Item Functioning* (DIF) och sedan kunnat härledas till flickors menstruation) (22). Tar man inte hänsyn till detta finns det en risk att förståelsen för personens situation blir missvisande och

behandlingen felaktig. Givetvis gäller detta oavsett om det handlar om könsskillnader eller andra demografiska eller socio-ekonomiska skillnader.

Vidare, att personförmågor och uppgifter mäts på samma skala ger dessutom en viktig direkt klinisk relevans. Som visas i *figur 1b*, en person som har en förmåga motsvarande "5" klarar aktiviteter som att dricka och duscha, men däremot inte att klippa gräset eller springa. Detta kan i sin tur ge fysioterapeuten och annan hälso- och sjukvårdspersonal en tydlig förståelse för en patients förmåga för att därefter rikta behandlingsåtgärder.

#### Mätosäkerheter vid kategoriskt baserade mätningar

Som nämnts ovan finns det inte en enda mätning som är helt precis. Hur säker man kan vara på ett mätresultat beror på mängden information man har tillgång till. Det vill säga, fler observationer eller frågor innebär att man får mer information om patienten och samtidigt minskar mätosäkerheterna (23–25). Patienter med den lägsta respektive högsta förmågan, det vill säga de som antingen klarar väldigt lite (golfeffekter) eller nästan allt (takeffekter), kommer därför alltid att ha större

mätosäkerheter än de som befinner sig i mitten av skalan (24). Det ska tilläggas att det behöver vara "rätt" uppgifter för patienten. Om en persons förmåga motsvarar att kunna duscha, men inte klippa gräset eller springa, då hjälper det inte att ställa ytterligare frågor om hans förmåga att utföra tunga eller hårt fysiskt krävande uppgifter som hen inte klarar av. Det ger inte mer information om hans förmåga, och påverkar således inte mätosäkerheterna.

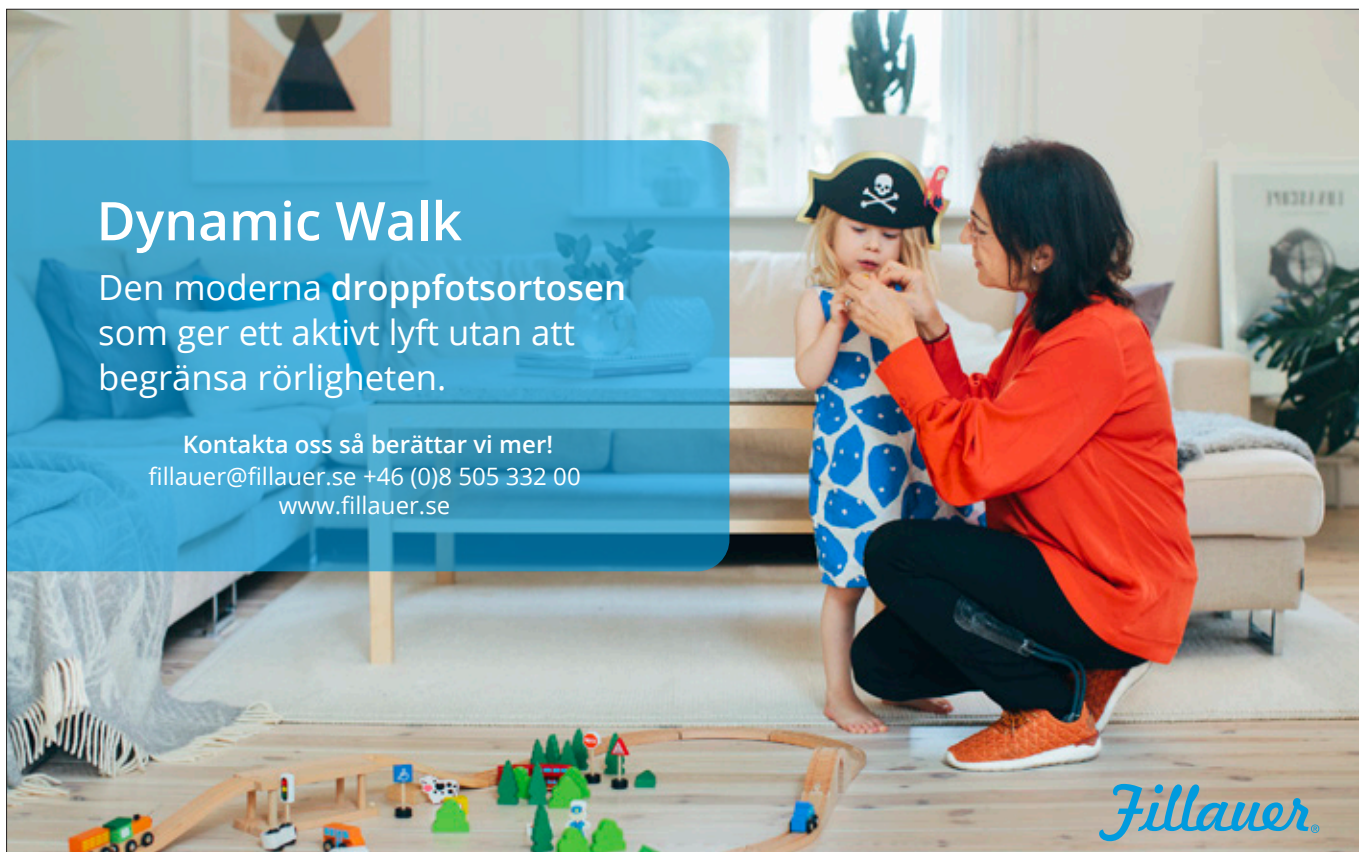
#### Rasch Measurement Theory vid utveckling av kategoriskt baserade mätmetoder

Tillämpning av Rasch Measurement Theory har på senare år fått en allt större roll i forskning och utveckling av kategoriskt baserade mätmetoder (26), även om den klassiska testteorin fortfarande dominerar. Rasch Measurement Theory ingår i den moderna testteorin, men skiljer sig från *Item Response Theory (IRT)* (20, 27). Med Rasch Measurement Theory testas om grundläggande mätkriterier kan uppfyllas utifrån givna data, medan *Item Response Theory* använder statistiska modeller för att förklara data. Utifrån givna grundläggande →

## Dynamic Walk

Den moderna droppfotsortosen som ger ett aktivt lyft utan att begränsa rörligheten.

Kontakta oss så berättar vi mer!  
fillauer@fillauer.se +46 (0)8 505 332 00  
www.fillauer.se



Fillauer.

→ mätkriterier för spårbarhet och invariants ger en Rasch-analys därför information om hur olika frågor i en enkät fungerar tillsammans och är således ett viktigt verktyg för att utvärdera hur väl en enkät kan/bör användas (20, 21, 28–30).

För att skapa ett frågeformulär eller en skala för bedömningar måste variabeln man avser att mäta definieras så att man säkerställer dess innehållsvaliditet (31). Exempelvis, *vad är fysisk funktion* eller *vad är smärta* och vilka påståenden eller frågor kan användas för att operationalisera detta? För att definiera variabeln används vanligen en kombination av patientintervjuer, tidigare forskning och expertutlåtanden (6). Här finns dock en viktig distinktion mellan den klassiska och den moderna testteorin; den klassiska testteorin avser att formulera påståenden eller frågor som ”ligger nära varandra” medan den moderna testteorin avser att ”fånga ett kontinuum”. Det senare är en förutsättning för att inte få för stora mätosäkerheter utan att med större tillförlitlighet kunna mäta en patients förändring över tid eller jämföra olika grupperns förmågor, symtom eller upplevelser. Inom Rasch Measurement Theory framhålls därför

vikten av en ordinal begreppsteori, det vill säga att man på förhand har en teori om hur olika påståenden eller frågor hierarkiskt ska förhålla sig till varandra (32–34).

### Exempel från fysioterapi

Även om metrologiska metoder är sparsamt använda och den klassiska testteorin dominerar så finns det flera goda exempel inom fysioterapi där Rasch Measurement Theory har tillämpats. Nedan följer tre exempel på förmåga, symtom respektive upplevelser.

#### Att mäta förmåga: fysisk funktionsförmåga

Fysisk funktionsförmåga kan vara ett av de tydligaste och mest studerade exemplen på hur olika uppgifter har olika svårighetsgrad. När olika skalor för fysisk funktionsförmåga utvecklats har det oftast funnits en ordinal teori, det vill säga att man har haft en teori om hur olika påståenden/frågor skulle förhålla sig till varandra. Genom Rasch-analyser på flera olika frågeformulär och skalor för bedömningar har dessa teorier bekräftats. Två exempel är jämförelser av värden för uppgifternas svårighetsgrad,  $\delta$ , för delskalan fysisk funktion PF10 i SF-36 från olika studiepopulationer (35) och samkalibrering av *Barthel Index* (BI) och FIM (36). Att på sådana sätt upprepade gånger bevisa hierarkin mellan uppgifter som är lättare att utföra respektive svårare styrker validiteten, det vill säga att man mäter det man avser att mäta.

#### Att mäta symtom: smärta och fatiuge

Ett av de vanligaste verktygen för att för att bedöma grad av symtom, till exempel vid en smärtbedömning är en *Visuel Analog Skala* (VAS). Skattningen behandlas dock oftast som ett linjärt mått, vilket är felaktigt, då avstånden mellan skalstegen är större i mitten av skalan än vid de låga respektive höga skattningarna. Det får bland annat konsekvenser för bedömningar av hur många poängs skillnad som klassas som en ”kliniskt relevant skillnad” (37). Vidare är mätosäkerheter större när en eller få frågor ställs.

Ett annat vanligt symtom i fysioterapeutens kliniska vardag är fatigue. För att bedöma fatigue används en rad olika frågeformulär och även dessa har utvärderats med hjälp av Rasch Measurement Theory (38). Testerna inkluderade 40-item *Fatigue Impact Scale* (FIS) samt kortare versioner av FIS såsom 21-item *Modified Fatigue Impact Scale* (MFIS) och 8-item *Daily Fatigue Impact Scale* (D-FIS). Efter analysen av hur väl de olika varianterna uppfyller



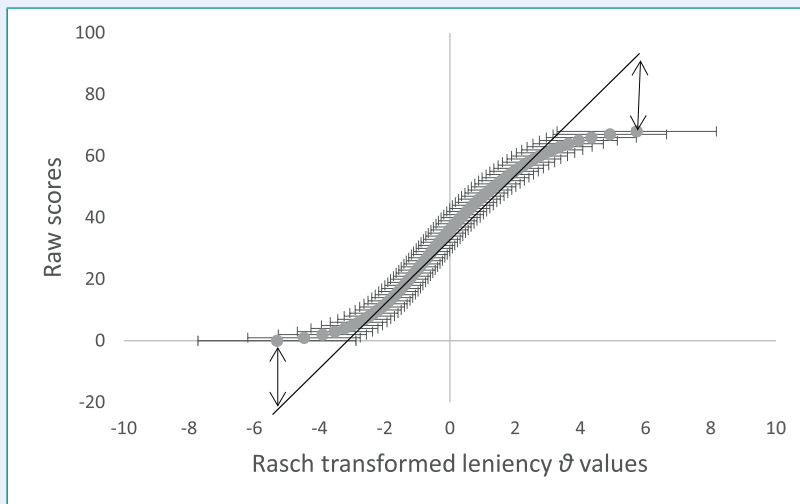
 **GIH** THE SWEDISH SCHOOL OF SPORT AND HEALTH SCIENCES

## Masterprogrammet i idrottsvetenskap

Vid GIH kan idrottsvetenskap omfatta studier i såväl fysiologi, biomekanik, pedagogik, psykologi, historia som medicin.

Sök innan 15 april på [antagning.se](http://antagning.se)

Gymnastik- och idrottshögskolan vid Stockholms Stadion  
[gih.se](http://gih.se)



**Figur 2.**

Jämförelse mellan klassiska testteorins summa poäng (raw scores, y-axeln) och Rasch-värde (x-axeln). Den svarta heldragna linjen indikerar en perfekt korrelation och de streckade pilarna i ändarna visar hur CTT över- respektive underskattar patientens erfarenheter (39).

de grundläggande mätkriterierna drog man slutsatsen att både FIS och MFIS kan användas men att man behöver göra justering för *differential item functioning* (DIF) för att säkerställa att resultaten inte blir missvisande (38).

#### Att mäta upplevelser: erfarenheter av patientdelaktighet

Det tredje exemplet, att mäta upplevelser av delaktighet, är det svåraste avseende att tänka i en hierarkisk struktur mellan olika påståenden eller frågor i en enkät. Även om samma grundläggande mätprinciper är nödvändiga även här, kan det vara det område där det hittills är minst tillämpat.

Vid mätningar av patienters upplevelser av delaktighet i sin rehabilitering har man sett fördelar med att tillämpa Rasch Measurement Theory (39). Frågeformuläret *Patient Participation in Rehabilitation Questionnaire* (PPRQ) utvecklades initialt enligt principer för den klassiska testteorin, men har senare även validerats med Rasch Measurement Theory och uppvisar acceptabla

mättegenskaper. Jämför man en patients summa poäng med Rasch-värdena så ser man en tydlig S-kurva som beror på icke-linjära ordinaldata [figur 2]. Likaså, på grund av takeffekter, ser man stora mätosäkerheter för de som skattar högst. Det finns därför en stor risk att man underskattar dessa patienters erfarenheter om man endast summerar skattningarna.

#### Konklusion

Att tillämpa grundläggande metrologi är en förutsättning för att säkerställa och möjliggöra tillförlitliga och jämförbara mätresultat av kategoriskt baserade mätningar. Det finns i dag stark evidens för att tillämpa modellen *Människan som mätinstrument* och *Rasch Measurement Theory* vid mätningar av förmågor, symtom och upplevelser. Det innebär att metrologin nu bör få en plats inom fysioterapins forskning och klinik för att säkerställa kvaliteten i professionens mätningar, vilket i sin tur kan möjliggöra korrekta funktionsdiagnoser och val av åtgärder samt uppföljningar. ■

## REFERENSER

- 1. Rothman ML, Beltran P, Cappelleri JC, Lipscomb J, Teschendorf B, Mayo FDAP-ROCMG. Patient-reported outcomes: conceptual issues. *Value Health*. 2007;10 Suppl 2:S66-75.
- 2. de Silva D. *Measuring patient experience*. London, UK; 2013.
- 3. *History of Psychometrics*. Wiley StatsRef: Statistics Reference Online.
- 4. Kingsley C, Patel S. Patient-reported outcome measures and patient-reported experience measures. *BJA Education*. 2017;17(4):137-44.
- 5. Black N. Patient reported outcome measures could help transform healthcare. *BMJ* : British Medical Journal. 2013;346:f167.
- 6. Administration USDoHaHSFaD, (CDER) CfDEaR, (CBER) CfBEaR, (CDRH) CfDaRH. *Guidance for Industry Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims* 2009.
- 7. Pendrill L, Melin J. *Measuring counted fractions in healthcare*. *TMQ Techniques, Methodologies and Quality*. 2019;NÚMERO ESPECIAL:60-9.
- 8. Stevens SS. *On the Theory of Scales of Measurement*. *Science*. 1946;103(2684):677-80.
- 9. Pendrill L. *Man as a Measurement Instrument*. *NCSLI Measure*. 2014;9(4):24-35.
- 10. Pendrill L. *Assuring measurement quality in person-centred healthcare*. *Meas Sci Technol*. 2018;29:034003.
- 11. Cano SJ, Pendrill LR, Melin J, Fisher WP. *Towards consensus measurement standards for patient-centered outcomes*. *Measurement*. 2019;141:62-9.
- 12. McGrane J. *Stevens' forgotten crossroads: the divergent measurement traditions in the physical and psychological sciences from the mid-twentieth century*. *Frontiers in Psychology*. 2015;6:Article 431.
- 13. Pendrill L. *Quality Assured Measurement, Unification across Social and Physical Sciences*: Springer International Publishing; 2019.
- 14. *Kvalitet i välfärden - bättre upphandling och uppföljning*. Stockholm: Wolters Kluwer; 2017.
- 15. *Mesures BidPe. Worldwide metrology 2019* [Available from: <https://www.bipm.org/en/worldwide-metrology/>].

## REFERENSER

---

- 16. Metrology JcFgi. International Vocabulary of Metrology – Basic and General Concepts and Associated Terms 2012.
- 17. Metrology JcFgi. Evaluation of measurement data – Guide to the expression of uncertainty in measurement. 2008.
- 18. Cano S, Pendrill L, Barbic S, Fisher W. Patient-Centred Outcome Metrology for Healthcare Decision-Making. *Journal of Physics: Conference Series*. 2018;1044:012057.
- 19. Rasch G. Probabilistic models for some intelligence and attainment tests. Copenhagen: Danish Institute for Education Research (Expanded edition (1980) with foreword and afterword by B.D. Wright, Chicago: The University of Chicago Press, 1980. Reprinted Chicago: MESA Press, 1993. Available from [www.rasch.org/books.htm](http://www.rasch.org/books.htm)); 1960.
- 20. Andrich D. Implications and applications of modern test theory in the context of outcomes based education. *Studies in Educational Evaluation*. 2002;28(2):103-21.
- 21. Hagquist C, Bruce M, Gustavsson JP. Using the Rasch model in nursing research: an introduction and illustrative example. *Int J Nurs Stud*. 2009;46(3):380-93.
- 22. Hagquist C, Andrich D. Recent advances in analysis of differential item functioning in health research using the Rasch model. *Health and Quality of Life Outcomes*. 2017;15(1):181.
- 23. Wright B. Error, variances and correlations. *Rasch Measurement Transactions*. 1991;5:2:147.
- 24. Standard Errors and Reliabilities: Rasch and Raw Score. *Rasch Measurement Transactions*. 2007;20(4):1086.
- 25. Hobart JC, Cano SJ, Thompson AJ. Effect sizes can be misleading: is it time to change the way we measure change? *Journal of Neurology, Neurosurgery & Psychiatry*. 2010;81(9):1044-8.
- 26. Leung YY, Png ME, Conaghan P, Tennant A. A systematic literature review on the application of Rasch analysis in musculoskeletal disease -- a special interest group report of OMERACT 11. *J Rheumatol*. 2014;41(1):159-64.
- 27. Cano SJ, Hobart JC. The problem with health measurement. *Patient Prefer Adherence*. 2011;5:279-90.
- 28. McClimans L, Browne J, Cano S. Clinical outcome measurement: Models, theory, psychometrics and practice. *Stud Hist Philos Sci*. 2017;65-66:67-73.
- 29. Hobart JC, Cano SJ, Zajicek JP, Thompson AJ. Rating scales as outcome measures for clinical trials in neurology: problems, solutions, and recommendations. *Lancet Neurol*. 2007;6(12):1094-105.
- 30. Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis and rheumatism*. 2007;57(8):1358-62.
- 31. Morel T, Cano SJ. Measuring what matters to rare disease patients – reflections on the work by the IRDiRC taskforce on patient-centered outcome measures. *Orphanet Journal of Rare Diseases*. 2017;12:171.
- 32. Fisher WP, Wilson M. Building a Productive Trading Zone in Educational Assessment Research and Practice. *Pensamiento Educativo: Revista de Investigacion Educativa Latinoamericana*. 2015;52(2).
- 33. Salzberger T. Thomas Salzberger - Six decades of measurement using the Rasch model in the social sciences. *PM Health; Kristianstad, Sweden* 2019.
- 34. Stenner AJ, Smith M, Burdick D. Towards a theory of construct definition. *Journal of Educational Measurement*. 1983;20(4):305-16.
- 35. Fisher WP, Jr. Foundations for health status metrology: the stability of MOS SF-36 PF-10 calibrations across samples. *J La State Med Soc*. 1999;151(11):566-78.
- 36. Hobart JC, Cano SJ, Thompson AJ. Effect sizes can be misleading: is it time to change the way we measure change? *J Neurol Neurosurg Psychiatry*. 2010;81(9):1044-8.
- 37. Kersten P, White PJ, Tennant A. Is the pain visual analogue scale linear and responsive to change? An exploration using Rasch analysis. *PLoS One*. 2014;9(6):e99485.
- 38. Lundgren-Nilsson A, Tennant A, Jakobsson S, Simren M, Taft C, Dencker A. Validation of Fatigue Impact Scale with various item sets – a Rasch analysis. *Disabil Rehabil*. 2019;41(7):840-6.
- 39. Melin J, Fornazar R, Spångfors M, Pendrill L. Rasch analysis of the Patient Participation in Rehabilitation Questionnaire (PPRQ). *Journal of Evaluation in Clinical Practice*. 2019;0(0).