



**INSTITUTIONEN FÖR FILOSOFI, LINGVISTIK  
VETENSKAPSTEORI**

# **DEN UTOPISKA VISIONEN – ETABLERING AV EKVIVALENS I ANPASSNING AV PSYKOLOGISKA TEST**

**Nalika Tjärnberg**

---

Uppsats/Examensarbete:	30 hp
Program och/eller kurs:	Masterprogram i evidensbaserad forskning
Nivå:	Avancerad nivå
Termin/år:	Vårterminen 2017
Handledare:	Morten Sager
Examinator:	Ingemar Bohlin

## Abstract

Uppsats/Examensarbete:	30 hp
Program och/eller kurs:	Magisterprogram i evidensbaserad
Nivå:	Avancerad nivå
Termin/år:	Vt/2017
Handledare:	Morten Sager
Examinator:	Ingemar Bohlin
Nyckelord:	Test adaption, instrument adaption, test translation, test standards, cross-cultural test adaption

- 
- Syfte:** Det övergripande syftet med uppsatsen är att bidra med en samlad kunskapsbild och en fördjupad förståelse för teoretiska och praktiska villkor för testadaption för att stödja riktlinjeförfattare, testutvecklande forskare, HR- och organisationskonsulter, policyproducenter och testförmedlare. Det innefattar att undersöka olika kunskapsperspektiv och synliggöra huvudaktörer och intressenter i systemet för testadaption i syfte att identifiera källor till spänningar och problem.
- Teori:** Utgångspunkten för uppsatsen är aktörnätverksteori (ANT) vilket ger tillgång till en generellt symmetrisk hållning i studiet av hur olika riktlinjer har utformats och olika aktörers agerande inom testadaption. Viktiga analytiska resurser hämtas också från Porters (1996) objektivitetsbegrepp och standardiseringens sociologi (Timmermans och Berg, 2003)
- Metod:** De grundläggande tillvägagångssätten i uppsatsen är beskrivande och kartläggande, jämförande och analytiskt perspektiverande. Ett urval av akademiskt publicerade standarder för anpassningsprocesser gällande psykologiska test beskrivs. Ramverk och riktlinjer jämförs, argument för dessa identifieras, objektivitetssyn och kunskapsglapp analyseras liksom eventuella spänningar eller hinder för praktisk tillämpning av riktlinjerna. Undersökningens datainsamlade tillvägagångssätt utgörs av en kombination av litteratursökning och intervjuer (telefon- och skypeintervju samt digital korrespondens) och bygger på kvalitativa data (icke-numeriska data).
- Resultat:**
- Riktlinjernas ”ankdamm”: Utifrån de standarder som granskats framstår det som att en mindre krets av forskare och associerade organisationer är tongivande för publicering, produktion, spridning och diskussion av riktlinjer och standarder. Vissa är även själva testutvecklare.
  - Visionen om ekvivalens: Överlag har den uppsättning standarder som presenteras övervägande gemensamma synsätt som kan sammanfattas i strävan efter funktionell ekvivalens, universell design och rättvisa.

Ett antal spänningar mellan olika sätt att argumentera för riktlinjers kvalitet blir tydliga, spänningar som riktlinjeproducenter och testutvecklare antagligen bör reflektera över:

- Mellan procedurer och människor: I kampen för ekvivalens förlitar sig aktörerna (riktlinjeförfattarna) mestadels på vad som kan beskrivas som en mekanisk objektivitet, alltså en tillit till procedurer, även om disciplinär objektivitet, alltså tillit till professionellt omdöme, ges visst utrymme.
- Hårdare disciplinering: Inslagen av disciplinär objektivitet skiljer sig på avgörande sätt från generella hänvisningar till standardiseringens saliggörande effekter inom delar av evidensrörelsen. Proceduralisering följs åt av nästan omöjliga krav på expertis vid testutveckling. Det handlar inte enbart om en öppning för bedömning, utan krav på vem som ska få genomföra bedömningarna.
- Begränsad transparens: Trots att många ställer krav på dokumentation i riktlinjebaserad testadaptation verkar inte tyst expertkunskap inkluderas i sådana krav på transparens.
- Hopplös kamp för ekvivalens? En springande fråga synes vara i vilken utsträckning kulturella jämförelser kan mekaniseras: Kan disciplinär objektivitet omvandlas till (mekaniska) procedurer utan att kulturellt känsliga element reduceras bort för mycket? I förhållande till detta presenteras idéer om flexibel standardisering liksom pragmatisk evidensinformerad praktik.
- Riktlinjernas irrelevans: Förslaget för mer flexibel standardisering riskerar dock att bli irrelevant. Testutvecklarens faktiska arbete verkar i vissa drag bortse från riktlinjeförfattarnas kamp för att uppnå ekvivalens i översättning av riktlinjer, främst utifrån resursfrågor.

Tjärnberg, Nalika (2017). Masteruppsats. *Den utopiska visionen: Etablering av ekvivalens i anpassning av psykologiska test*. Institutionen för filosofi, lingvistik och vetenskapsteori, Göteborgs Universitet.

## Innehållsförteckning

1.	Inledning	1
1.1.	Bakgrund	1
1.2.	Tester och datainsamlande i ett globaliserat samhälle	1
1.3.	Standarder för anpassning av tester – teori och praktik	2
1.4.	Uppsatsens struktur	4
2.	Syfte och frågeställningar	5
2.1.	Syfte	5
2.2.	Frågeställningar	5
3.	Metod	8
3.1.	Introduktion	8
3.2.	Litteratursökning	9
3.2.1.	Urval av riktlinjer och standarder för denna uppsats	10
3.3.	Intervjuer	10
4.	Teori	12
4.1.	Symmetrisk ansats	12
4.2.	Actor Network Theory	13
4.2.1.	Latours blomma – Vetenskapens cirkulationssystem	13
4.3.	Objektivitet – vetenskapens klassiska ideal	16
4.3.1.	Mekanisk och disciplinär objektivitet	17
4.3.2.	Sammanfattande slutsatser	18
4.4.	Standardisering och standarder	19
4.4.1.	Bakgrund – standardiseringens politik	19
4.4.2.	Vad är standardisering och standarder? Definition och karaktäristika	20
4.4.3.	Flexibel standardisering	20
4.4.4.	Sammanfattning	21
5.	Ramverk och riktlinjer: kartläggning och jämförelse	22
5.1.	Introduktion	22
5.2.	Standards for Educational and Psychological Testing	22
5.3.	International Test Commission (ITC)	23
5.4.	Guidelines for the process of Cross-Cultural Adaption of Self-Report Measures	25
5.5.	Translating and Adapting Measurement Instruments for Cross-Linguistic and Cross-Cultural Research: A Guide for Practitioners	27
6.	Gemensamma teman i riktlinjernas innehåll	29
6.1.	Introduktion till testadaption med riktlinjeförfattarnas perspektiv	29
6.2.	Ekvivalens	30
6.3.	Universell design	31

6.4.	Psykometri	31
6.5.	Jämförelse: variationer på samma kamp för ekvivalens	32
6.6.	Bristande kontakt med användarna	33
7.	Testutvecklare: riktlinjernas irrelevans	35
7.1.	Testernas marknad	35
7.2.	Resurskrav	36
7.3.	Sammanfattande diskussion	37
8.	Analys av ramverk och riktlinjer – Objektivitetssyn och standardisering	39
8.1.	Introduktion	39
8.2.	Epistemologisk förståelse: (o)giltigheten i kulturell ekvivalens	40
8.3.	Universell design och rättvisa	43
8.4.	Dokumentation	44
8.5.	Psykometrisk utvärdering	45
9.	Diskussion och slutsatser	47
10.	Referenslista	52
11.	Bilagor	59
11.1.	Bilaga 1: Litteratursökning.	59
11.1.1.	Avsikt med litteratursökning	59
11.1.2.	Inklusionskriterier	59
11.1.3.	Exklusionskriterier	60
11.1.4.	Litteratursökningsperiod augusti 2015 - november 2016	60
11.1.5.	Ramverk och riktlinjer	64

# 1. Inledning

## 1.1. Bakgrund

Användandet av test och instrument är globalt sett mycket spritt. Olika typer av test, instrument och frågeformulär används i ett stort antal länder, för alla åldrar för att tillhandahålla kunskap, information och vägledning för en mängd relevanta och aktuella samhällsfrågor. Det är en etablerad socialvetenskaplig metod (Behling och Law, 2000). En variation av tester/instrument av psykologisk och/eller pedagogisk karaktär har kontinuerligt utvecklats under de senaste dryga hundra åren här i väst.

Olika typer av tester och instrument nyttjas i samhället idag som underlag för att fatta ytterst viktiga och avgörande beslut för individer och program av varierande slag, som kan sträcka sig från att avgöra passande psykologisk behandling, till urval och rekrytering vid utbildningar eller anställningar (Oakland, Pootinga, Schlegel och Hambleton, 2001). Olika former av tester och instrument används även på nationell och mellannationell (populations)nivå som beslutsunderlag för till exempel samhällsinstitutioners policyprogram eller för olika samhällsorienterade och jämförande attitydundersökningar (Behling & Law, 2000; Plake & Wise, 2014).

Termerna *skala*, *inventorium* eller *frågeformulär* används vanligtvis för mätningar av sådant som attityder, intressen, och dispositioner. För enkelhetens skull, i texten som följer får benämningen *test* representera alla typer av (ut)värderande psykologiska instrument (så som inventorium, frågeformulär och enkäter).

## 1.2. Tester och datainsamlande i ett globaliserat samhälle

Som en spegling av en pågående rörelse mot en allt mer globaliserad och mångkulturell värld ser vi inte minst inom det psykologiska kunskapsfältet att det växer fram ett ökat behov av flerspråkiga versioner av pedagogiska och psykologiska tester – tester som till exempel mäter förmåga, begåvning och personlighet men även sådant som attitydundersökande och självskattningsbaserade enkäter och surveys.

Dagens ökande behov av testadaption (översatta och anpassade tester) drivs på bland annat utifrån att a) internationella utbyten av tester blir allt vanligare, b) allt fler (akademiska) examina utgår från internationella poängsystem och c) intresset för tvärkulturell forskning ständigt växer (Hambleton, Merenda och Spielberger, 2005). Detta sker i takt med att våra samhällen blivit allt mer flerspråkiga och mångkulturella (Erickson, 2002; Hambleton och de Jong, 2003) och därför har även det psykologiska och pedagogiska vetenskapsfältet behövt bli allt mer internationellt över de senaste decennierna (American Educational Research Association (AERA), American Psychological Association (APA), National Council on Measurement i Education (NCME), 2014; Hambleton, et al., 2005). Globaliseringen av psykologi som vetenskap märks kanske mest uppenbart genom just det internationella nyttjandet av tester och andra former av datainsamlade instrument (Oakland, 2005).

För att kunna jämföra och/eller överföra forskningsresultat från olika länder, mellan olika språk och kulturer, så är det förstas viktigt att främja och sträva efter att säkerställa tillförlitlig jämförbarhet inom ramen för de bedömningsmetoder som används (Alonso-Arbo och van de Vijver, 2010). Det räcker dock inte med att översättningen resulterar i bibehållen god reliabilitet och validitet. Olika författare framhåller att det som ska undersökas eller mätas också behöver framstå som konceptuellt meningsfullt inom ramen för dess nya kulturella kontext (se till exempel Beaton, Bombardier och Guillemin, 2000; eller Hambleton et al., 2005). Ett av uppsatsens genomsyrande teman är analytiska frågor som berör vad som är överförbart från en kultur till en annan.

Ett vanligt förekommande antagande är att både bedömningsskalor och ingående operationaliserade koncept eller begrepp – inom testpsykologi så kallade *konstrukt* – som testet är designat att mäta opererar likvärdigt för respondenter över de olika intressegrupperna (Byrne, 2016). Att antaga att en

bedömningskala som utvecklats för att användas för ett språk och ett land per automatik skulle mäta samma fenomen på exakt samma sätt när det översätts till ett annat språk för att användas i ett annat land och kultur är inte särskilt rimligt. Det kan röra sig om skillnader i språkbruk som förändrar betydelsen av specifika *items* (items är ett samlingsbegrepp för de frågor eller påståenden respondenter ska svara på i test och instrument), kulturella skillnader i beteenden som antas representera underliggande konstrukt, erfarenhetsmässiga skillnader för respondenterna vad gäller exponering för specifika bedömningsformat (text, bilder, symboler) och så vidare. Det innebär att när tvärnationella jämförelser och tvärkulturell forskning ska bedrivas så blir det nödvändigt att forskare och andra testutvecklare har förståelse för komplexiteten i att anpassa tester och instrument. Detta gäller även om ett test anpassas från ett språk och kultur till ett annat språk och kultur för att isolerat användas i den nya målkulturen, det vill säga utan att det avses användas för tvärkulturell jämförelse. Testanpassning är minst sagt en intrikat process. Behovet av systematiskt, vägledande stöd för de som vill ta sig an den grannliga uppgiften att mer noggrant översätta och anpassa ett befintligt test till andra förhållanden än originalversionen, har uppmärksammats av en rad författare.

### 1.3. Standarder för anpassning av tester – teori och praktik

För att kvalitetssäkra testutveckling och vägleda för testutvecklare har standarder och riktlinjer för testutveckling publicerats åtminstone sedan 1950 talet (Plake & Wise, 2014). Däremot tycks mer sammansatta, substantiella standarder och riktlinjer med specifikt fokus på testrelaterat översättnings- och anpassningsarbete ha vuxit fram främst under de senaste cirka tjugo-trettio åren, som till exempel de senaste två versionerna av American Educational Research Association et al. (1999 och 2014) och International Test Commission (ITC, 2005 och 2017). Dessa standarder och riktlinjer avser enligt egen utsaga att reflektera de bästa professionella bedömningarna kring hur man kan designa, utveckla och använda test som förväntas ge återkopplande resultat som är rättvisande och som präglas av giltighet för respondenterna i den nya målpopulationen (som kan tänkas representera en signifikant annorlunda kulturell kontext än originalversionens målpopulation). Riktlinjerna som nämnts ovan vänder sig till testutvecklare forskare men också till andra intressenter, så som policyproducenter och företag vars affärsmodell bygger på att ta fram och utveckla tester och instrument. Fler riktlinjer presenteras längre fram i texten.

Enligt vissa författare är dock denna typ av ramverk och riktlinjer relativt okända för testutvecklare och testanvändare och nyttjas inte i önskvärd omfattning (Oakland et al., 2001). Hambleton et al. (2005) hänvisar till forskare inom det tvärkulturella fältet som anser att en ansevärd mängd av deras egna tidigare forskningsresultat idag framstår som skeva och bristfälliga, och därmed icke tillförlitliga på grund av att de i sin tidigare forskning utgått från slarvigt översatta och anpassade test (se även resonemang i Banville, Desrosiers, & Genet-Volet, 2000; Sireci, Yang, Harter & Ehrlich, 2006). Om vi utgår från en rad experter inom testadaptionsfältet och deras syn på saken, så framstår det som att testutvecklare och andra praktiker åtminstone tidigare (oroande) sällan låtit sig vägledas av de metodologiska rekommendationer som sammanställts och publicerats (Hambleton, 2005; Hulin, 1987; van de Vijver & Leung, 1997).

Resonemanget så långt leder fram till några olika indikationer och möjliga slutsatser. Dels framstår det som att det tidigare, överlag tycks ha rått en bristande systematisk kunskap om det praktiska hantverksarbetet, vad gäller att anpassa tester till andra målgrupper och fler miljöer än vad de ursprungligen togs fram och avsågs för. Det kanske till och med i vissa lägen – trots kunskap om – förekommit en alltför lättvindig attityd till uppgiften och därför en nonchalans i att beakta komplexitet och noggrannhet i anpassningsarbetet (med risk för dyrbara konsekvenser så som de tvärkulturellt inriktade forskarna ovan vittnar om och ger exempel på). Dels tycks också de hittills producerade riktlinjerna inte alltid ha haft önskad genomslagskraft, sett ur riktlinjeförfattarnas perspektiv. En möjlig förklaring kan vara att forskarsamhället tidigare överlag har arbetat och kommunicerat förhållandevis isolerat, med spretig dokumentation om dessa kunskapsfrågor. Detta inte bara ut mot omvärlden men kanske också gentemot kollegor, såväl inom det egna som gentemot andra discipliner



(detta resonemang stöds av till exempel Hambleton et al., 2005 och Kristjansson, Desrochers och Zumbo, 2003). Mer eller mindre omfattande rekommendationer och riktlinjer har producerats parallellt inom och mellan olika disciplinära domäner.

Att de riktlinjer som arbetats fram ändå generellt inte nått någon vidare läs- och användarkrets utanför den akademiska sfären, indikerar att de i så fall helt enkelt inte varit lättillgängliga för andra möjliga intressenter. Kanske har externa intressenter, utanför forskarkretsarna, inte heller alltid vetat var de ska söka information och vägledning. Rekommendationer gällande riktlinjer och standarder för testanpassning förmedlas formellt enligt min kännedom – förutom via forskningsartiklar och ett fåtal böcker – av ett antal granskande och intressebaserade nationella och internationella organisationer vars (del)syfte är att förmedla och upprätthålla kvalitetssäkring av standarder för psykologiska tester. Exempel på sådana organisationer är The British Psychological Society's (PBS) och dess systerorganisation Psychological Testing Centre (PTC) och det svenska Psykologförbundet. Kännedom om dessa organisationers existens och om deras arbete inom testkvalitet varierar sannolikt bland testutvecklare och testanvändare.

Ytterligare ett perspektiv på utvecklingen av riktlinjer kommer från studier av evidensrörelsen. För att förstå riktlinjearbetet är det till stor hjälp att se det som en del av den vidare evidensrörelsens strävan att uppnå mer jämlik och standardiserad vård. Begreppen ”evidens” eller ”evidensbaserad praktik” används inte explicit i den testadaptionslitteratur som jag studerat men strävan mot implementering eller användning av evidensbaserade procedurer är genomgående.

Evidensbaserad och evidensbaserad praktik har fått stor genomslagskraft i Sverige och internationellt. Begreppet evidensbaserad praktik har sina rötter i det medicinska forskningsfältet från början av 1990-talet (Bohlin & Sager, 2011). Vikten av att arbeta på basis av forskning och evidens framhålls idag av många aktörer, så som statliga myndigheter, professionella organisationer och nätverk av olika slag (Nilsen, 2014). Genom ambitionen att erbjuda bästa möjliga vård, utbildning eller ledarskap – som bygger på bästa tillgängliga kunskap – är det många som vill eftersträva att överbrygga *the know-do gap*, det vill säga gapet mellan produktion och användning av forskningsresultat – alltså precis det problem som syns inom testadaptionsfältet. Att förverkliga en vision om evidensbaserad praktik *i praktiken* inom olika områden har dock ofta visat sig vara svårare att realisera än kanske förväntat (se till exempel Bohlin och Sager, 2011; Levinsson, 2013 och Nilsen, 2014).

Tidigare forskning har framgångsrikt nyttjat begreppsbyggnader som tematiserar objektivitet och standardiseringar för att fånga strävan inom evidensrörelsen (Porter, 1996 och Timmermans och Berg, 2003). Objektivitet berör uppfattningar om hur giltighet etableras inom olika fält, såsom tillit till procedurer (mekanisk objektivitet) respektive tillit till professionella gemenskaper (disciplinär objektivitet). Standardiseringssociologin följer hur standarder både skapas av och bidrar till formeringen av vetenskapliga och samhällliga processer.

Evidensrörelsen kan i sig betraktas som en rörelse mot ökad standardisering och ökad tillit till mekaniska procedurer. Initialt verkar testadaptionsfältet uppvisa liknande drag men begreppen väcker intressanta frågor: Hur bidrar standarder med förflyttningen av relevansen i ett psykologiskt test från ett språk och kultur till ett annat språk och kultur? I vilken utsträckning kan kulturell kontext ”fångas in” utan att kulturellt känsliga element reduceras bort för mycket i översättnings- och anpassningsprocessen? I denna högst intrikata utmaning blir standarder förstärkt attraktiva i att de ger signaler om stabilitet, myndighet och ändamålsenlighet (Timmermans, 2015, s. 81) på vägen att skapa tester som är giltiga, både där och här. Vid betoning på tillit till standarder devalveras dock oundvikligt andra alternativa källor till tillit (Porter, 1996). Rådande syn på objektivitet definierar vilken kunskap som värdesätts. Sådana frågor om objektivitetssyn kan belysa riktlinjeproduktion och vetenskaplighet med en uppfriskande distans, för att bättre förstå praxisnära frågor och problem. Analytiskt inriktade frågor om objektivitetssyn hjälper i strävan att identifiera spänningar som finns mellan olika aktörer, utan att direkt ”välja sida”. Vi har sannolikt alla våra vetenskapliga preferenser. Förhoppningen är att reflektioner kring vetenskaplighet och syn på objektivitet kan bidra till att anlägga ett epistemologiskt helikopterperspektiv för frågor om vad som står på spel och vad som

skulle kunna vara annorlunda? Dessa frågor leder oss över till uppsatsens syfte och frågeställningar, efter en presentation av uppsatsens struktur.

## 1.4. Uppsatsens struktur

Uppsatsen har inlett med att i avsnitt 1 ge en bakgrund till tester och testadaptionsfältet som sådant samt några inledande kritiskt reflekterande frågor. För att underlätta för dig som läsare ges en övergripande bild av uppsatsens struktur framåt med de kommande avsnitten.

I avsnitt 2, som är det nästa, presenteras uppsatsens syfte och frågeställningar. Detta kommer vara vägledande för det som undersöks i uppsatsen. Frågeställningarna är uppdelade i beskrivande och analytisk karaktär för att mynna ut i de lärdomar som kan dras av denna studie – vilken samlad kunskapsbild växer fram? Därpå följer i avsnitt 3 en kort redogörelse för metod innan jag i avsnitt 4 redogör för uppsatsens teori och ansats med tillhörande vetenskapsteoretiska analysverktyg. I denna tid då ramverk och riktlinjer för testanpassning med målet att uppnå kulturell ekvivalens växer fram, vilka ansatser är då möjliga? Jag har tagit fasta på den symmetriska logiken inom aktörnätverksteori (ANT) enligt Bruno Latour (2005) samt Theodore Porters (1996) reflexioner över objektivitetsideal samt standardiseringens sociologi (Timmermans och Berg, 2003) som problematiserar och sätter in standarder som sådana i ett större politiskt perspektiv.

Därefter i avsnitt 5 presenteras i korthet de ramverk och riktlinjer som ingår i uppsatsen. Detta för att ge en kort inblick i respektive uppsättning standarder. Därpå följer en introduktion till teman som riktlinjeförfattarna betonar i ramverk och riktlinjer, detta i avsnitt 6. Dessa teman med tillhörande centrala begrepp belyses för att visa på vad riktlinjeförfattarna fokuserar på och underlätta för dig som läsare att följa med i de analytiska resonemangen allt eftersom.

I avsnitt 7 lämnar uppsatsen riktlinjerna och deras författare för att undersöka hur riktlinjerna används av testutvecklare. Här blir det tydligt att användarna har ett helt annat synsätt på riktlinjerna än producenterna.

Med detta tillryggalagt övergår avsnitt 8 till en analys av ramverk och riktlinjer där frågan är vilken syn på objektivitet riktlinjeförfattarna ger uttryck för och förhållandet till praktisk tillämpbarhet. Därpå följer i avsnitt 9 diskussion och slutsats för att sammanfatta och samla de lärdomar som kunnat dras.

Du som läsare kommer att få en inblick i vetenskapsteoretiska teorier och metoder som berör syn på vad vetenskap är, perspektiv på objektivitet och standardiseringens sociologi (avsnitt 3). Dessa inledande vetenskapsteoretiska presentationer lägger grund för att kunna belysa riktlinjeproduktion och vetenskaplighet med en välgörande distans och ta ett epistemologiskt helikopterperspektiv, för att bättre förstå praxisnära frågor och problem.

Härnäst följer presentation av uppsatsens syfte och frågeställningar.

## 2. Syfte och frågeställningar

### 2.1. Syfte

Det övergripande syftet med uppsatsen är att bidra med en samlad kunskapsbild och en fördjupad förståelse för teoretiska och praktiska villkor för testadaptation för att stödja riktlinjeförfattare, testutvecklande forskare, HR- och organisationskonsulter, policyproducenter och testförmedlare. Det innefattar att undersöka olika kunskapsperspektiv och synliggöra huvudaktörer och intressenter i systemet för testadaptation i syfte att identifiera källor till spänningar och problem.

Uppsatsen redogör för ett urval av akademiskt publicerade standarder för anpassningsprocesser gällande psykologiska test. Uppsatsen avser jämföra ramverk och riktlinjer; identifiera argument för dessa; kunskapsglapp eller hinder för praktisk tillämpning av dessa. Förutom att bidra med kunskap till testadaptionsfältet som sådant förväntas även att kunna ge vägledning till de som avser att anpassa eller använda anpassade test. Studien gör detta via att analytiskt fånga och kartlägga spänningar i testadaptionsfältet. Jag vill i mina slutsatser påvisa spänningar som aktörerna bör beakta. Det analytiska syftet är således att djupare förstå spänningsfälten, snarare än att presentera en lista med punkter på ”gör så här”.

### 2.2. Frågeställningar

För att fördjupat kunna förstå spänningsfälten inom testadaptation – på teoretiska och praktiska plan – behöver vi ställa och söka svar på ett antal frågor. Vi behöver ta reda på vad som rör sig – och inte – i testadaptationens cirkulerande blodomlopp, genom att packa upp möjliga ”svarta boxar”. Svarta boxar utgör i detta sammanhang icke synliggjorda och för givet antagna förhållningssätt, procedurer och kunskapsyn, något som i förlängningen påverkar både riktlinjeproduktion och praktisk tillämpning.

För att uppnå syftet delas uppsatsens frågeställningar in i:

- klarläggande och beskrivande frågor med inriktning att undersöka befintlig kunskap och ta reda på ”hur det ligger till” (frågeställning nummer 1 - 2).
- frågor som är analytiskt/kritiskt inriktade och som berör möjliga – och för uppsatsens syfte – intressanta glapp eller spänningar i den vetenskapliga cirkulationen (frågeställning nummer 3 - 4).
- Slutligen, givet det som ”packats upp” under klarläggande och granskande fas – Vilka lärdomar kan vi dra (frågeställning nummer 5)?

En inledande fas av uppsatsen är inriktad mot undersökande frågeställningar, med klarläggande och beskrivande inriktning som berör *produktion* av standarder och riktlinjer. Detta ger underlag till de mer analytiska frågeställningarna. I detta skede vill jag ta reda på:

1. Vilken vetenskapligt eller akademiskt producerad kunskap i form av standarder och riktlinjer finns att tillgå för testadaptation gällande psykologiska test? Var sprids och förmedlas de?

I anknytning till denna första frågeställning vill jag även ta reda på vilka som producerar dessa standarder och riktlinjer, till exempel om det är organisationer eller enskilda riktlinjeförfattare.

2. Hur är de olika standarderna och riktlinjerna utformade?

Här avser jag att undersöka likheter och skillnader i sådant som omfattning, struktur och tematiska innehåll samt i vilken mån det råder samstämmighet eller inte mellan olika ramverks- och riktlinjeproducenter för hur testadaptation rekommenderas att gå till.

Med grund i det som framkommit från ovan klarläggande och beskrivande frågeställningar, rör vi oss över till undersökande frågor med analytisk/kritisk inriktning. Frågeställningen tar explicit ansats att analysera objektivitetssyn (Porter, 1996) och standardiseringens sociologi (Timmermans och Berg, 2003).

3. Vilka eventuella spänningar finns gällande synsätt på giltig kunskap och vetenskaplighet, inom och sinsemellan ramverk och riktlinjer?

Det kan tänkas att innehållet i de olika standarderna och riktlinjerna bygger på en epistemologisk förförståelse där vissa former av kunskap framhålls framför andra kunskapsperspektiv. Vilka former av kunskap och vetenskaplighet är det i så fall som framhålls? Finns det kunskapsperspektiv som motsäger varandra eller kunskapsperspektiv som saknas utifrån ramverk och riktlinjers syfte och mål?

4. Finns det kunskapsperspektiv som riktlinjeförfattarna skulle behöva beakta och inkludera för att främja ökad tillämpbarhet av standarderna och riktlinjerna? Vilka i så fall?

Frågeställningen adresserar hur de företrädde objektivitets- och kunskapsperspektiven gifter sig med praktiskt genomförande. Med utgångspunkt i de objektivitets- och kunskapsperspektiv som framhålls i publicerade ramverk och riktlinjer, avses att undersöka hur dessa perspektiv matchar praktisk tillämpning. Matchar dessa perspektiv de presenterade avsikterna – att underlätta för testadaptation i praktiken? Finns det några problem med grund i sådant som objektivitets- och kunskapsyn som signifikant påverkar möjlighet till praktisk tillämpning? Tillika, finns det intressenter som glöms bort i testadaptationens system vars intressen och behov ej uppmärksammas? Hur påverkar det i praktiken och vilka följder får det?

När vi kommit så här långt har vi förhoppningsvis nått ”pudelns kärna”. Hela studien syftar till att bidra med en samlad kunskapsbild och en fördjupad förståelse för villkoren för testadaptation. Med grund i analys av de publicerade standarderna och riktlinjerna, via frågeställningarna och svaren på dessa (enligt ovan) eftersträvas att klarlägga: Vilka lärdomar kan vi dra? Utifrån vilka spänningar som har uppdragats, vilka konsekvenser får det för tillämpning av riktlinjerna? Om det finns behov - Hur kan standarder och riktlinjer utformas för att bättre matcha praktisk tillämpbarhet? Detta leder fram till studiens femte och sista frågeställning:

5. Hur kan teoretiska instruktioner behöva förändras för att bli mer användbara?

Efter att ha presenterat uppsatsens syfte och centrala frågeställningar rör vi oss över till en presentation av metod och tillvägagångssätt, som legat till grund för att kunna besvara dessa frågeställningar.

## 3. Metod

### 3.1. Introduktion

Jag vill inleda med några rader om vetenskapligt förhållningssätt och min analytikerroll. Som vetenskaplig analytiker avser jag att rikta likvärdig nyfikenhet till de olika aktörer som ingår i analysen. I detta ligger även att försöka förstå bakomliggande orsakssamband till ”varför det är som det är”. Med ett symmetriskt förhållningssätt i min analys ämnar jag lyfta fram samtliga aktörers situation och perspektiv som likvärdiga, att utifrån deras respektive perspektiv behandla dessa som om de var likvärdigt ”riktiga och sanna”. Jag strävar mot att inte på förhand inta en given föreställning om vad som utgör god vetenskaplig produktion och i förlängningen riktningen för evidensbaserad praktik inom testadaptionfältet. Min avsikt är att eftersträva ett symmetriskt förhållningssätt (beskrivs nedan). Ur ett vetenskapsteoretiskt perspektiv möjliggörs och erbjuds ett sådant analytiskt förhållningssätt till exempel via aktörsnätverk (ANT). Jag har utgått ifrån ANT och det har därför präglat och influerat inriktningen på det metodinriktade och analytiska arbetet. Dock avser jag att kombinera detta förhållningssätt med att utifrån den kunskapsbild som växer fram kunna ge förslag på hur riktlinjer kan behöva förändras för att vara mer användbara i praktiken.

Genom att försöka synliggöra delar av testadaptionens system och centrala aktörer ökar sannolikheten att få grepp om verkande mekanismer och hur detta tar sig uttryck i sammanhanget: vad som ges legitimitet i kunskapsproduktion och vad som förs vidare – och inte – från riktlinjeproduktion till praktiskt genomförande, och omvänt. I utformning och spridning av standarder finns sannolikt olika intressen där aktörer (riktlinjeproducenter) utövar inflytande på sådant som kunskapsynn, tillgänglighet och användbarhet. Olika aktörer har olika intressen som kan påverka sådant som vilka standarder som blir etablerade, var de sprids och/eller hur testutvecklande praktikers perspektiv hanteras.

De riktlinjer jag kände till vid uppstart av uppsatsen var en äldre version av Beaton, Bombarider och Guillemin (2000), som legat till grund för arbetet med svensk anpassning av enkäten Copsoq (Copenhagen Psychosocial Questionnaire). Därmed hade jag knapphändig kunskap om publicerade riktlinjer för testadaption.

Metodarbetet har genomsyrats av ett ANT-perspektiv. Efter att inledningsvis ha läst Latour (se till exempel 1999; 2005) lade jag denna litteratur åt sidan för att helhjärtat ägna mig åt att utforska akademiskt publicerade ramverk och riktlinjer samt angränsande litteratur, spridning av riktlinjer och praxis inom testadaption. Min teorisyn har samtidigt påverkat mitt läsande och sökande arbete, där jag har försökt att följa och synliggöra olika aktörers perspektiv och hur olika aktörer relaterar, eller inte, till varandra.

Den första centrala frågeställningen - vilken vetenskapligt eller akademiskt producerad kunskap i form av standarder, ramverk och riktlinjer finns att tillgå för testadaption gällande psykologiska test? – förde in inledande arbete på att ta reda på om det för det första överhuvudtaget finns fler akademiskt publicerade ramverk och riktlinjer än de jag hade kännedom om. Jag sökte efter forskningsbaserade riktlinjer och ramverk som var inriktade mot praktisk tillämpning. Det ledde fram till att jag identifierade ett antal ramverk och riktlinjer inom ramen för de inklusionskriterier jag satte upp för mitt arbete.

Därefter ville jag ta reda på utformning och innehåll för ett urval av dessa i linje med frågeställning nummer 2 (Hur är de olika standarderna och riktlinjerna utformade?). Jag undersökte likheter och skillnader i sådant som omfattning, struktur och tematiska innehåll samt i vilken mån det (inte) råder samstämmighet mellan olika ramverks- och riktlinjeproducenter för hur testadaption rekommenderas att gå till. Jag uppmärksammade även vilka som genererar dessa standarder (enskilda forskare eller organisationer och eventuella kopplingar dem emellan).

Under arbetets gång tog jag även kontakt via email med organisationer vars uppdrag delvis utgörs av att värna testkvalitet och/eller kvalitetssäkra psykologiska tester. Dessa organisationer var: British Psychological Society (BPS) och systerorganisationen Psychological testing Centre (PTC), Psykologförbundet och European Federation of Psychologists' Association (EFPA). Detta för att ta reda på vilka riktlinjer, om några alls, som de rekommenderar, sprider information och kunskap om. Detta relaterar i första hand till frågeställning nummer 1. Valet av organisationer att kontakta skedde genom att jag antingen fick kännedom om dessa organisationer via litteratursökningen eller genom förslag från de intervjuade informanterna men också genom sökande över internet. Kontakt med de olika organisationerna skedde under vårtermin och hösttermin 2016.

Förfaringsätt för urval av informanter för intervjuer utgjordes av en slags tillgänglighetsprincip. Intervjuerna adresserade olika aspekter av praxis inom testadaptation, som relaterar till frågeställning nummer 3 (Vilka eventuella spänningar finns gällande synsätt på giltig kunskap och vetenskaplighet för de granskade ramverk och riktlinjer? Hur relaterar detta till praktik?) och nummer 4 (Vilka kunskapsperspektiv finns som riktlinjeförfattarna skulle behöva beakta och inkludera för att främja ökad tillämpbarhet av standarderna och riktlinjerna?). Intervjuerna genomfördes efter att jag hade fullföljt den grundläggande litteratursökningen.

Jag har genomgående försökt att läsa och lyssna mig till vad som bedöms fungera väl, vilka problem, spänningar och svårigheter som råder inom området samt identifiera dolda perspektiv och osynliggjord kunskap. Jag har även haft i åtanke och sökt att ta reda på hur de olika aktörerna inom testadaptation samverkar, eller inte samverkar och vilka konsekvenser det tycks leda till.

Jag har försökt identifiera företrädade objektivitets- och kunskapsperspektiv samt effekter av detta. Som ett led i detta har jag fördjupat mig i relationen objektivitetssyn och vetenskaplighet via Porter (1996). Jag har även relaterat till detta studerat standardiseringens sociologi via Timmermans och Berg (2003) men även Timmermans, (2015).

Frågeställning nummer 5 - Hur kan teoretiska instruktioner behöva förändras för att bli mer användbara? - följer av strävan att, utifrån uppsatsens omfång och i görligaste mån, bidra med en mer samlad kunskapsbild över testadaptationens system och en fördjupad förståelse för villkoren för testadaptation. I kontrast till en betoning på standarder och etablerade normer har därför även andra objektivitets- och kunskapsperspektiv belysts, inklusive att praxisorienterad kunskap synliggjorts.

Nedan redogörs för en mycket nedkortad version av hur metodarbetet gått till.

## 3.2. Litteratursökning

Avsikten med litteratursökningen har varit att hitta, jämföra och sammanställa riktlinjer för testadaptation samt relevant angränsande kunskapsbärande litteratur. Jag sökte efter publicerade riktlinjer och standarder samt vägledning i systematiska tillvägagångssätt för testadaptation. Litteratursökning genomfördes via databaser tillgängliga via de båda universitetsbiblioteken vid Umeå Universitet och Göteborgs Universitet samt databasen Google Scholar. Sökning har genomförts vid olika tillfällen från augusti 2015 till oktober 2016 samt vid ett tillfälle juli 2017. Inklusionskriterier, sökvägar och så kallade söksträngar redovisas i bilaga 1, tillsammans med utförlig beskrivning av litteratursökningen i sin helhet.

Jag har arbetat med systematiska litteratursökningar i kombination med så kallad snöbollsmetod. Snöbollsmetoden har i detta fall inneburit att följa upp ett urval av referenser som framkom i litteraturen i de systematiska sökningarna. Jag sökte vidare i dessa spår och på så vis hittades ibland ny relevant litteratur, som i sin tur gav vägledning vidare till ytterligare litteratur.

Den systematiska litteratursökningens karaktär skiftade under arbetets och tidens gång. Övergripande rörde den sig från att innefatta ett vidare, explorativt sökljus till ett allt mer avsmalnat fält utifrån de fynd jag gjorde och bedömd relevans av dessa. Parallellt med att litteratur valdes bort så

bidrog snöbollsmetoden å andra sidan med nya sökvägar och fynd som resulterade i ökad tillgång till intressant litteratur. Tillsammans bidrog dessa metoder till att en viss mättnad så småningom uppstod, då flertalet artiklar och böcker hänvisade till en och samma samling av författare. Graden av mättnad bedömdes utifrån rimlighet i relation till uppsatsens omfång.

Jag arbetade med litteraturen genom att läsa och jämföra den sinsemellan, detta utifrån innehåll och omfattning, likheter och skillnader, eventuella särdrag och betoningar i rekommendationerna.

### 3.2.1. Urval av riktlinjer och standarder för denna uppsats

De ramverk och riktlinjer för testadaption som slutligen kom att ingå i uppsatsen är:

- American Educational Research Association, American Psychological Association, National Council on Measurement in Education. (2014). *Standards for Educational and Psychological Testing*. Published by American Educational Research Association, US, Washington, DC.
- International Test Commission. (2005 och 2017, inledningsvis 2005 och senare 2017). International Guidelines on Test Adaptation. Hämtad 15 juni 2016 respektive 14 juli 2017 från International test Commission: <https://www.intestcom.org>.
- Kristjansson, E. A., Desrochers, A. &, Zumbo, B. (2003). Translating and Adapting Measurement Instruments for Cross-Linguistic and Cross-Cultural Research: A Guide for Practitioners. *Canadian Journal of Nursing Research*, 35 (2), 127 -142.
- Beaton, D., Bombardier, C. &, Guillemin, F. (2000). Guidelines for the Process of Cross-Cultural Adaption of Self-Report Measures. *SPINE*, 25 (24), 3186-3191.

### 3.3. Intervjuer

Uppsatsen har frågeställningar inriktade mot att ta reda på om det finns kunskapsperspektiv som riktlinjeförfattarna skulle behöva beakta och inkludera för att främja ökad tillämpbarhet av standarderna och riktlinjerna. Som ett led i att besvara detta har praxisbaserad kunskap inkluderats genom ett par intervjuer. De intervjuade har båda gedigen erfarenhet av testutveckling och praktisk testadaption.

Hanne Berthelsen arbetar vid Centrum för tillämpad arbetslivsforskning och utvärdering (CTA), samt vid Odontologiska fakulteten vid Malmö Högskola. Berthelsen är även svensk och internationell koordinator för det internationella Copsoq-nätverket. Copsoq (Copenhagen Psychosocial Questionnaire) är ett instrument som undersöker organisatorisk och social arbetsmiljö. Copsoq finns i olika reviderade utgåvor och har översatts till cirka tjugofem språk (för mer information se <http://www.copsoq.se>). Instrumentet har utvecklats av det danska nationella forskningscentret för arbetsmiljö (<http://www.arbejdsmiljoforskning.dk>), som kan sägas vara en motsvarighet till Arbetsmiljöverket i Sverige. Berthelsen har varit del i det team som översatt och anpassat Copsoq till svenska. I Stressforskningsrapport nummer 326 (Berthelsen, Westerlund och Söndergård, 2014) återfinns en utförlig beskrivning av de olika stegen och procedurerna under den senaste svenska anpassningsprocessen, där forskarna utgick från en äldre version av Beaton et al:s riktlinjer (2000, se redovisade riktlinjer ovan). Intervju med Berthelsen skedde över Skype (160610).



Legitimerad psykolog Sofia Sjöberg (telefonintervju 160819) arbetar på det privata testföretaget Assessio med flerårig erfarenhet av att både utveckla egna tester och anpassa psykologiska tester. I sammanhanget avser *testutveckling* att utveckla och generera nya tester. *Anpassning* avser att översätta och (kulturellt) anpassa befintliga tester. Assessio har en flerårig historia inom testutveckling och anpassning av tester, både kliniska tester och arbetspsykologiska sådana. De arbetar uteslutande med egna tester förutom de så kallade Hogan-instrumenten. Assessio arbetar idag främst med att utveckla och anpassa arbetspsykologiska tester.

Urvalet av intervjupersoner skedde genom att jag redan innan uppsatsarbetets start kände till Copsoq-instrumentet och Berthelsens roll i det internationella utvecklingsarbetet av detta. Jag hade även kännedom om Assessio, som en välmeriterad organisation inom psykologisk testning, med nationella och mellannationella uppdrag.

Jag försökte att få kontakt med ytterligare ett par forskare och praktiker men dessa kontaktförsök gav tyvärr ingen framgång. Jag sökte kontakt med angiven kontaktförfattare (Lars Wallin) till artikeln ”Swedish Translation, Adaptation and Psychometric Evaluation of the Context Assessment Index (CAI)” (Nilsson Kajermo et al., 2013) för en möjlig intervju. Tyvärr har svarsrespons uteblivit trots upprepade försök till kontakt.

Jag sökte även kontakt med artikelförfattare till ”Att översätta och anpassa ett engelskt instrument för bedömning av fysisk miljö på särskilt boende till svenska förhållanden – utmaningar och erfarenheter” av Nordin, Elf, McKee och Wijk (2012). Även i detta fall har svarsrespons uteblivit.

De intervjuer som genomfördes var semistrukturerade och dialogbaserade till sin karaktär. En intervju genomfördes över internet via Skype (Berthelsen) och en intervju över telefon (Sjöberg). Intervjuerna varade mellan cirka en timme till en och en halv timme. Jag eftersträvade att främja ett samtalsklimat som genomsyrades av öppenhet gällande att dela med sig av personliga erfarenheter och reflektioner ur ett praktikerperspektiv. Jag sökte även spinna vidare på – det jag bedömde som – värdefulla, intressanta eller unika trådar som uppkom i samtalet. De strukturerade frågeställningarna och inriktning på resonemangen berörde främst tillämpning av testadaptation, dels i relation till riktlinjer och standarder och dels med inriktning mot alternativ eller kompletterande praxis (se frågeställningarna nummer 3 och 4 ovan). I upplägget av intervjuerna och i anslutning till grundfrågorna fanns även utrymme för andra relaterade frågeställningar, som informanten ville dela med sig av och ansåg relevant i sammanhanget.

Informanterna har gett sitt medgivande till publiceringen i uppsatsen. Resultatet från intervjuerna presenteras löpande i belysande och analyserande text nedan.

Efter en genomgång av uppsatsens metod följer härnäst en redogörelse för uppsatsens teori.

## 4. Teori

### 4.1. Symmetrisk ansats

Denna studie begränsas till att presentera, jämföra och analysera ett antal ramverk och riktlinjer avsedda för testadaptation (översättnings- och anpassningsprocesser) för psykologiska tester, och detta med särskilt fokus på (teoretiska) riktlinjer i förhållande till praktik. Jag avser att belysa epistemologiska perspektiv där objektivitet och standardisering som fenomen relateras till riktlinjeproduktion och tillämpning inom testadaptation. För att studera möjliga spänningar inom testadaptionsfältet avser jag att sträva efter en symmetrisk analys som innefattar en icke normativ syn på kunskap och möjliga kunskapsglapp.

Begreppet *symmetri* introducerades på sjuttioalet av vetenskaps sociologen Bloor (1976) som en reaktion på det han uppfattade som asymmetri i historie- och vetenskapsbeskrivningar. Det innefattar en kritisk teoretisk hållning där Bloor menar att ideologiska förhållanden avgör framställningen av hur saker och ting ligger till, men det skulle likväl kunna framställas och vara annorlunda (beroende på vem som formar narrativet). Symmetri och opartiskhet innebär i detta fall att analysen bedrivs likartat, oavsett om det som studeras betraktas som bevisade fakta eller misstag inom en viss vetenskap, i syfte att komma åt aktörernas faktiska situation (Sager, 2015). Detta är en allmän princip inom teknik- och vetenskapsstudier (STS) och ej specifikt ANT.

För mig betyder ett symmetriskt förhållningssätt att förhålla mig agnostiskt, till exempel i förhållande till vad som kan tänkas anses vara ”rätt” kunskapsperspektiv eller bästa praktik för anpassning av tester. Vetenskaplig produktion är dock som jag ser det förbunden med en uppsättning värderingar i förhållande till egna (disciplinära) institutioner. Mitt undersökande avser att i möjligaste mån spåra rötterna till rådande vetenskapliga argument inom riktlinjeproduktion för testadaptation och visa på dess eventuella relation till sociala och politiska aspekter. Symmetrisk analys kan därför även ge en klarare bild av varför forskarna (riktlinjeförfattarna) har nått fram till sina trosföreställningar och slutsatser. Jag avser även att lyfta in aktörer och perspektiv i den vetenskapliga processen, som med gängse vetenskapliga perspektiv skulle kunna ses som biased eller som icke-vetenskapliga aktörer, och därmed icke hemmahörande i det vetenskapliga arbetet. Med ett symmetriskt perspektiv inkluderas dessa i den vetenskapliga processen. Min och uppsatsens uppgift är att spåra och att öppna en del av de svarta boxar som uppdragas, just för att förhoppningsvis i slutändan kunna framlägga en mer samlad kunskapsbild, som dessutom kan ge nya vägledande perspektiv för olika i sammanhanget aktuella aktörer.

Som analytiker ställer jag mig utanför och kan på distans granska spänningar inom testadaptationens fält. Jag vill framhålla att detta är en analytikers privilegium. Aktörerna (riktlinjeförfattarna) som jag granskar är fullt ut inblandade och verkar utifrån sina intressen – akademiska, epistemiska och sociala – i sitt system för att säkra resurser och främja status för sin forskning (jmf. Sager, 2015). I min analytiska roll kan jag lyfta fram spänningar utan att själv riskera några uppenbara nederlag eller att förlora något medan det för aktörerna står mycket mer på spel. Denna positionsskillnad behöver vägas in som analytiker och som läsare av analysen.

I mitt analytiska arbete utgår jag från vissa centrala begrepp som är användbara i att belysa testadaptionsfältet som sådant – dels Porters (1996) resonemang om olika synsätt på vetenskaplig objektivitet och dels begrepp inom det som kallas standardiseringens sociologi (eng. *The Standardization of Sociology*). Det sistnämnda med stöd främst i Timmermans och Bergs (2003) men även Timmermans (2015) publikationer. Dessa olika perspektiv på vetenskaplighet länkar väl in i varandra för att ta ett symmetriskt grepp på uppsatsens analyser.

Härnäst följer en presentation av dessa tre olika vetenskapsteoretiska perspektiv med tillhörande resonemang. Detta inleds med en kortare introduktion till ANT, med valda delar som jag anser har relevans för denna uppsats. Därefter följer objektivitet och slutligen standardisering.

## 4.2. Actor Network Theory

Actor Network Theory (ANT) är en teori om *hur* fenomen kan studeras eller hur aktörer tillåts att få utrymme att uttrycka sig (Latour, 2005, s.142). Bruno Latour, en av upphovsmännen bakom ANT, ser samtidigt ANT mer som en metod än som en teori; ”*The name of the game is to get back to empiricism*” (Latour, 2005, s.146).

På ett metodologiskt plan karaktäriseras ANT av sin betoning på ”science in action”, det vill säga att följa vetenskap/forskare från produktion till presentation och se hur den gradvisa etableringen av vetenskapliga fakta sker på olika nivåer och genom olika nätverk (Latour, 1999, s. 98). Fokus vilar på att se *aktörerna* (eng. *actors*), följa vad de gjort och gör, spåra associationer vart än de leder (även utanför ens egna disciplinära fält). Tillsammans med aktörer utgör *verkan* (eng. *agency*) ett samverkande, centralt begrepp inom ANT (Latour, 2005, s.46 och s. 74 - 78). Aktörernas agerande och vad som skapar verkan sker i ett intrikat sammanhang där vem eller vad som är ansvarigt för det som sker – det som skapar verkan - är outgrundligt och ej alltid uppenbart synligt (detta kallas aktörsnätverk, Latour, 2005, s. 217).

Inom ANT ses natur och kultur som produkter av en process. De är inte förklarande, men de behöver en förklaring. Hårt draget finns ingen natur att avslöja, med hjälp av våra kulturella instrument. Latours intresse inriktas på att synliggöra en rad transformationer från *world to word*, för att bättre förstå denna process och vad som ”händer på vägen” i forandet av vetenskap (Latour, 1999, s.98).

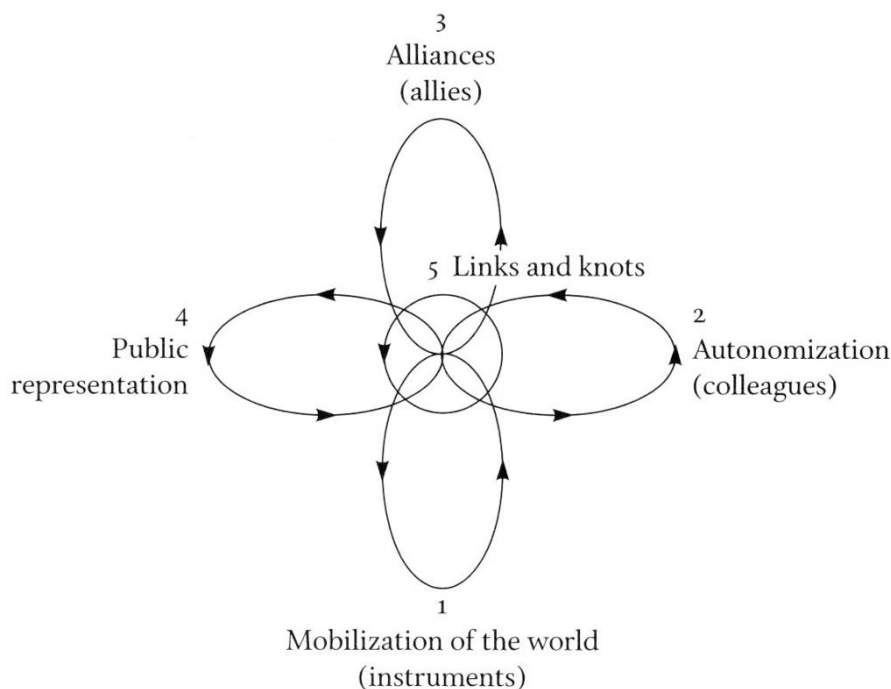
Enligt Latour utgörs vetenskapsstudier av att följa hur fakta cirkulerar – att rekonstruera (genom att packa upp de svarta boxarna) blodåder för blodåder genom hela det vetenskapliga blodomloppssystemet. Där står inte endast det vetenskapligt konceptuella i fokus utan även kontextuella faktorer. Att betrakta det vetenskapliga som något särskilt från resten av samhället framstår som meningslöst menar Latour (1999, s. 91 – 92). Syftet är dock ej att påvisa att det a priori existerar ett samband mellan vetenskap och olika delar av samhället, eftersom förekomsten av dessa former av förbindelser beror på vad aktörerna har gjort eller ej gjort för att etablera dessa samband (Latour, 1999, s. 86). ANT tillhandahåller medel att spåra förbindelser när de finns och existerar – att förstå vetenskap är att förstå denna komplexa (myllrande) väv av sammankopplingar utan att på förhand föreställa sig att det existerar ett givet *state of society* och ett givet *state of science* menar Latour. Detta agnostiska förhållningssätt betonas inom teknik- och vetenskapsstudier (Science and Technology Studies, STS), det fält där många av Latours idéer plockats upp. Ett sätt att se det är att ANT bidrar med att bryta loss analytikern från klassiska kategorier och möjliggör att analytiskt ta sig an och undersöka ett objekt eller fenomen mer symmetriskt.

Efter att ha introducerat ANT och perspektiv på vetenskap så följer en presentation av Latours modell över vetenskapens cirkulationssystem. Modellen som återges ytterst kortfattat, bidrar med ett underliggande synsätt på forandet av vetenskap jag har valt att utgå ifrån. Modellen bidrar med vidgade perspektiv för att symmetriskt undersöka och analysera vetenskaplig produktion.

### 4.2.1. Latours blomma – Vetenskapens cirkulationssystem

Enligt Latour (1999, s. 80) cirkulerar referenser och han anser att hur referenser cirkulerar signifikant påverkar vår förståelse av kopplingarna mellan vetenskapliga discipliner och resten av dess värld. I en modell som innefattar fem stycken olika typer av aktivitetsområden i form av blomblad, eller öglor (eng. *loops*) presenterar Latour (1999, figur 3.3, s. 100) vad han menar att alla

vetenskapsstudier behöver belysa och beskriva om avsikten är att på ett realistiskt sätt försöka förstå det som ligger till grund för varje given vetenskaplig disciplin.<sup>1</sup> De fem öglorna som tillsammans ingår i och utgör ett vetenskapligt vaskulärt cirkulationssystem består av: *mobilisering av världen, autonomi, allianser, offentlig representation* och slutligen det han kallar *länkar och knutar* (författarens översättning). Varje aktivitetssfär av dessa fem är lika viktiga som de övriga menar Latour. Varje ögla eller aktivitetssfär länkar tillbaka till de övriga öglorna eller aktivitetssfärerna. Utan kollegor ingen..., utan allierade ingen... och så vidare. I *figur 1* är de fem öglorna återgivna i form av en blomma. Tillsammans utgör de en karta som kastar ljus på och rekonstruerar cirkulationen av vetenskapliga fakta. Modellen återges i korthet men det finns mycket intressant skrivet och mer att säga om dess innefattande idéer, dock på grund av uppsatsens omfång ges en begränsad beskrivning. För denna uppsats tjänar den som verktyg i att belysa och analysera inte bara det konceptuella innehållet i ramverk och riktlinjer utan också olika aktörers kunskapssyn, intressen, spänningar och förutsättningar för att bidra till en mer samlad kunskapsbild av testadaptionens fält.



Figur 1: The five loops of the circulatory system of science

<sup>1</sup> Jag använder Latours modell för att belysa olika aktörer i olika öglor som tillsammans utgör väsentliga och vitala element för att vetenskapliga koncept ska få fäste och nå ut i samhället. Genom att visa på spänningar mellan olika aktörer i olika öglor kan vi få fatt på förutsättningar och vad som försvårar för att vetenskap blir förankrad, användbar och nyttiggjord även utanför de akademiska kretsarna. Via sin modell undersöker och synliggör Latour transformationen från värld till ord (eng. *from world to word*). Jag vill belysa att Latours (1999) explicita fokus inte ligger på spänningar mellan olika öglor i sin modell, så som presenteras i boken *Pandoras Hope*. Modellen är dock likväl enligt min uppfattning en adekvat analytisk resurs för att belysa glapp eller spänningar mellan olika öglor/aktörer inom testadaptionens fält, för att bidra med en mer samlad kunskapsbild. Jag vill även nämna att på grund av uppsatsens begränsade omfång går jag dock delvis ifrån Latours modell i att rekonstruera cirkulationen av vetenskapliga fakta, detta då jag inte närmare belyser aktörer i den fjärde öglan.

(Latour, 1999). (Medgivande från Bruno Latour 160109)

Inom ramen för Latours modell kan vi exemplifiera genom att placera in olika aktiviteter och/eller aktörer i dessa öglor. Ögla nummer 1 – mobilisering av världen – avser den data som samlas in om (den psykologiska) världen. I vårt fall berör det de instrument som översätts och anpassas med eller utan stöd i riktlinjer. Genom denna mobilisering konverteras världen till argument. I korthet kan man säga att genom att följa denna ögla möjliggörs studerandet av den logistik som är så oumbärlig för formanterna av vetenskapliga logiker. Riktlinjeförfattarna har ett särskilt intresse för denna ögla och med att de är måna om att den data som produceras (för att kunna genomföra tvärkulturella studier) förbättras. I ögla nummer 2 som berör formanterna av vetenskapliga idésystem och bedömningsgrunder, byggandet av akademisk kredibilitet och kollegiala truster (så som egna intressenätverk och branschorganisationer) men även att hantera interna motsättningar, där ingår riktlinjeförfattarna men också de som internt kritiserar deras arbete. Det är också här i den andra ögla som de epistemologiska diskurserna formas – de för testadaption avgörande idéerna om universalitet och möjligheten till ekvivalens mellan kulturer. Detta relaterar i sin tur till att i den kollegiala gemenskapen försöka enas om vad som är en framgångsrik objektivitetssyn och därmed vad som utgör legitim kunskap. Som vi kommer att se pågår det vissa interna diskussioner av sådant som definition av ekvivalens och i vilken grad riktlinjerna bör bli än mer enhetliga med stringentare standardiserade procedurer. Detta är frågor som tangerar epistemologiska perspektiv (objektivitetssyn) och i vilken mån testadaption kan standardiseras eller mekaniseras ytterligare – frågor som visar på spänningsfält mellan olika kollegors sätt att se på detta.

Ögla nummer 3 utgörs av att forma allianser med aktörer som tror på och förmedlar forskarnas konceptuella idéer. Detta är också en del i det vitala vetenskapliga arbetet enligt Latour. Ögla kan i detta fall till exempel utgöras av samspel mellan riktlinjeförfattare och myndigheter och näringsliv som är intresserade av den data som sammanställs i ögla 1 men även externa branschorganisationer, kommersiella testförmedlande företag som stödjer och sprider praktiserandet av riktlinjerna. Ögla 3 utgörs även av till exempel olika testanvändare (så som HR-konsulter, policyskapare utanför akademien med flera) som kan vara mer eller mindre allierade. Testanvändare kan ha mer eller mindre intresse av testadaptionens procedurer och huruvida de är utförda via akademiskt etablerade riktlinjer eller ej. Det påverkar vilka krav dessa aktörer ställer, då deras behov och preferenser sannolikt ser olika ut, vilket påverkar vad de efterfrågar. Även om det finns testutvecklande och testförmedlande företag, som strävar efter och ser vetenskaplig evidensbaserad som en kvalitetsstämpel och därmed ett säljargument gentemot kunder, så tycks samma typ av aktörer men som har andra intressen än de just nämnda, vara vanligt förekommande (testmarknaden är som sagt en lukrativ marknad). Det innebär i så fall en slags motrörelse till de som förespråkar en mer rigorös testadaptionprocess innan tester marknadsförs och används gentemot testmottagare (de som ska svara på testerna). Eftersom praktiserandet av riktlinjer inte fått den genomslagskraft som riktlinjeförfattare hoppats på (Oakland et al., 2001; Hambleton et al., 2005; International Test Commission, 2017) så kan vi ponera att det finns ett glapp i kommunikationen och/eller intressekonflikter med de aktörer som ingår i denna ögla, som behöver vitaliseras för vetenskaplighetens cirkulation och fäste.

Testmottagarna (som kan sägas utgöra medborgarna eller allmänheten) återfinns i ögla nummer 4. Deras intressen återspeglas i denna uppsats främst via riktlinjeförfattarnas och testutvecklarnas uttalade omsorg om dessas möjlighet till rättvisande tester och testningsvillkor. Dessa är dock vanligtvis i en position att behöva anpassa sig till ”det som erbjuds” utifrån den test-situation de befinner sig i. Sannolikt frågar sig flertalet av dessa inte om hur testers adaptionsprocess gått till. Detta är dock mina egna antaganden. Forskare behöver enligt Latour och denna modell engagera sig i relationer till en utomliggande, civil värld (än sina studieobjekt/subjekt): till exempel reportrar och journalister, lärda och olärda på gatan. Hänsyn till allmänhetens åsikter och opinion behöver beaktas seriöst eftersom

information inte endast flödar från de första tre öglorna till den fjärde, utan även från andra hållet och formar en hel del av de föreställningar som råder om forskare och deras olika kunskapsfält. Därför är denna ögla inte på något sätt marginaliserad eller i utkanten för vetenskapligheten. Denna ögla är också del i att paketera de fakta som vävts fram och som ska presenteras offentligt. Huruvida riktlinjeförfattarna bedriver offentliga diskussioner i dagspress och andra forum, om vikten av rigorösa testanpassningsprocedurer (för tester som utgör underlag för till exempel diagnostisering inom DSM-V och PISA-undersökningar eller arbetspsykologisk testning) undersöks inte i uppsatsen.

Den femte öglan utgör det konceptuella innehållet – den ”hårda” kunskapen som presenteras utåt, som får sin form via mer eller mindre förbindelse med de övriga öglorna. Hjärtat (det konceptuella innehållet) är så klart avgörande för att förstå det cirkulerande systemet i den mänskliga kroppen men det hänger ihop med resten och ingår i kroppens hela system. Detsamma gäller för vetenskapsstudier.

Uppmaningen från Latour är att när vi studerar ett vetenskapsesoteriskt innehåll så ska vi inte endast studera den femte öglan (det konceptuella innehållet) utan även titta efter de övriga fyra öglorna. Det konceptuella elementet är placerat i det vetenskapliga navet men utgör mer en slags knutpunkt som binder samman de fyra övriga öglorna (istället för att utgöra en kärna med en omgivande kontext). Det är dessa öglor tillsammans som sätter in det konceptuella innehållet i ett sammanhang och ger det dess mening i en vidare bemärkelse. Ett vetenskapligt koncept blir ej mer vetenskapligt genom att det isoleras eller flyttas längre bort från resten av vad det håller och utgörs av - utan för att det är sammanlänkat med en större repertoar av resurser (Latour, 1999, s.10). Innehållet i en vetenskap är inte något som härbärgeras i en container, det är i sig själv en container. När vi studerar vetenskapsesoteriska innehåll så ska vi även titta efter de övriga fyra öglorna, som ger innehållet dess mening. Det som denna uppsats analyserar, så som angavs ovan, berör dock främst det som pågår i de tre första öglorna samt förstås den femte.

Modellen är ett värdefullt analytiskt verktyg för att symmetriskt närma sig undersökandet av sambandet mellan å ena sidan teoretiska riktlinjer med underliggande objektivitetssyn och å andra sidan det praktiska tillämpandet av desamma. Nu rör vi oss över till att belysa olika perspektiv på objektivitet ur ett större samhälleligt perspektiv och hur det kan relatera till en akademisk underdisciplin som testadaption.

### 4.3. Objektivitet – vetenskapens klassiska ideal

I sin bok *Trust in Numbers* analyserar historikern Theodore Porter (1996), ur både politiska och kulturella perspektiv, den ökade tilliten till kvantifieringar av olika slag. Porter lyfter fram att objektivitet är ett av de klassiska idealen inom vetenskap. Det refererar till ett kluster av attribut: först av allt sanningen om naturen men det handlar också om stabila mänskliga egenskaper (personlighet), rättvisa (eng. *fairness*)<sup>2</sup>, universalitet, och generell immunitet mot allehanda sorters lokala störande faktorer som nationalitet, språk, personliga intressen och fördomar (Porter, 1996, s. 217). Så som begreppet allmänt förstås och används är objektivitet således nära nog synonymt med realism (förmågan att veta saker *så som de är*), medan subjektivitet refereras till som idéer och troföreställningar (som endast existerar i sinnet) (Porter, 1996, s. 3). Objektivitet betyder här kunskap som egentligen inte är för beroende av forskarna/författarna som producerar eller (be)skriver resultaten.

Porter sällar sig dock till ett annat resonemang som motsätter sig den vedertagna, allmänna betydelsen av objektivitet: kunskap icke påverkad av intressen och idiosynkrasier; kunskap som appellerar till verkligheten; eller värderingsfri kunskap. Porters slutsats är att objektivitet inte har någon djupare mening än kunskap producerad i överensstämmelse med rådande standarder inom

<sup>2</sup> Porters (1996) resonemang utifrån begreppet rättvisa (till exempel s. 4, s. 217 – 220) relaterar till olika perspektiv och synsätt på objektivitet. Dessa resonemang har gemensamma drag med hur begreppet diskuteras inom testadaptionlitteraturen men Porter belyser likväl andra aspekter av begreppet.

vetenskaplig praxis, som bestämts av gängse bedömningsgrunder inom det vetenskapliga samhället (1996, s. 216). Porter hävdar vidare att vetenskap är personlig kunskap, att all vetenskap innefattar ett kritiskt element av tyst kunskap, som ej kan artikuleras eller reduceras till nedtecknade regler (1996, s. 13). En slutsats är att böcker och vetenskapliga tidskrifter är ett dåligt sätt att dela och utveckla kunskap då essentiella aspekter ej kommer med i det som nedtecknas, eftersom detta reduceras bort eller inte låter sig fångas.

#### 4.3.1. Mekanisk och disciplinär objektivitet

Porter (1996) problematiserar vad som kan förstås som en inte sällan upplevd spänning mellan *mekanisk* (procedurell, författarens benämning) och *disciplinär* (professionell, expertisbaserad) objektivitet, i relation till standarder och riktlinjer inom en rad olika professioner (s. 90). Mekanisk objektivitet, som syftar till att ingjuta tillförlitlighet hos olika intressenter, gynnas av att följa tekniska regler menar författaren. Alternativa källor till tillit och förtroende – inte endast professionell expertis utan även hierarkisk styrning (chefsled), religiös auktoritet, allmän opinion, tradition eller magi – har fått ge vika för den mekaniska objektiviteten. I detta avsnitt belyses specifikt Porters resonemang för spänningen mellan mekanisk och disciplinär objektivitet.

Tilliten till standarder - i vetenskapliga och tekniska kontexter – har alltmer ersatt den personligt och relationellt grundade tilliten, som är mer typisk för mindre samhällen eller (disciplinära) gemenskaper. Under 1900-talet banades det väg för och skedde en successiv rörelse från förtroende och personlig tillit till experter och forskare, till förmån för kredibilitet baserat på standarder och (numerisk) kvantifiering. Porter benämner det som en rörelse mot mekanisk objektivitet, som uppstod i ett försök att ena ett svagt och pressat vetenskapssamhälle. Pressen kom dels utifrån i form av politiska påtryckningar och dels genom interna, objektivitetsgrundade motsättningar (Porter, 1996, s. 89 - 90; 193; 199 - 200, se även Timmermans, 2015, s. 93). Förlusten av att bygga tillit via personligt omdöme (expertobjektivitet) underminerade samtidigt möjligheten för andra (motståndare) att ifrågasätta vetenskapliga analyser och resultat (Porter, 1996, s. 228). Opersonlig kredibilitet har på så vis kommit att ses som mer trovärdig än personlig sådan.

En intressant iakttagelse som Porter gör är att, kvantifiering och statistiska slutsatser inte i första hand har sipprat ned från toppen av vetenskapens hierarkier, från den prestigefulla naturvetenskapen för att slutligen nå samhälls- och socialvetenskapliga fält. Snarare lades det mekaniska tidigt beslag på av svagare discipliner så som psykologi och viss medicinsk forskning. Standarder och statistik gav – och ger – förtroende där personligt (expert)förtroende saknas. Detta är förklarande argument för psykometrins starka ställning inom psykologin enligt Porter (1996, s. 200). Porter menar att nyligen formade vetenskapliga gemenskaper eller disciplinära fält (som testadaption) har behov av explicita standarder, för att adressera osäkerhet, spretighet och oroande mycket variation. Inom psykologi beskriver Porter amerikanska psykologer som drivande i standardiseringsrörelsen och hänvisar till *Standards for Educational and Psychological Testing* (1996, s. 228. Författarens notering: denna handbok i dess senaste utgåva analyseras i denna uppsats). Omfattningen på denna handbok ser han som ett resultat av en exceptionell standardiseringsiver med uppenbara och kännbara negativa begränsningar för psykologiforskning. Drivkrafterna att skapa standardisering inom psykologi har medfört att detta i sig stundvis ställt sig i vägen för att förstå de fenomen som fältet är intresserat av att studera. Porter menar vidare att psykometri har inom psykologi använts för att markera professionalism och skapa en mer robust forskningsidentitet (dock inte bara inom psykologi).

Mekanisk objektivitet ger privilegier för viss typ av vetenskaplig kunskap (så som psykometri och numerisk kvantifiering) där standarder även fyller en funktion. Standarders makt består i att de i sig har potential att sätta spelregler som andra behöver följa eller att sätta ramar för vilka alternativ som finns att tillgå (se Timmermans, 2015, s.79 med hänvisning till Busch, 2011).

För att ytterligare försöka förstå idésystemet och idealen bakom mekanisk objektivitet och kvantifiering, är att förstå kvantifiering som *en teknologi för att hantera avstånd*. Matematik är strukturerat och högst regelstyrt, ett nära nog universellt språk. Siffror, grafer och tabeller är enligt Porter kommunikativa strategier. De bidrar med att lägga fram världen som förståbar, på ett igenkänningsbart, överblickbart och standardiserat sätt som kan förstås och (mer eller mindre) enkelt förflyttas världen över. Det som är mest avgörande är att, genom att förlita sig på siffror och kvantifiering minimeras behovet av intim kunskap och personligt tillitsskapande. Det ger en form av objektivitet som kämpar med och mot subjektivitet, en objektivitet som utesluter personligt omdöme. Numerisk kvantifiering som metod tar bort mänskliga beslut. Forskaren blir endast en *förmedlare* av resultat (Porter, 1996, s.7 - 8). Det attraktiva i detta är att det syftar till att lösa eller vinna interna konflikter och kontroverser via kvantifiering likväl som att skydda sig mot yttre (politiska) hot. Det definierar likväl en syn på kunskap som i sig ger vissa företräden, för vissa.

Mekanisk objektivitet har under en längre tid setts som vetenskapens rågång. Ett beslut fattat av siffror (eller explicita regler av någon annan sort) har åtminstone skenet av att vara rättvist och opersonligt färgat. Porter säger dock att kvantifiering icke *endast* är en social strategi för politiska problem, men det är likväl det som är den stora vinsten med hans redogörelse: förståelsen för kvantifiering som socialt verktyg.

#### **4.3.2. Sammanfattande slutsatser**

Objektivitet, ordet i sig, frambringar en passionerad tyngd som få andra ord gör (Porter, 1996, s.1). Dess närvaro är bevisligen behövd för basal rättvisa, ärlig myndighetsutövning, och (så) sann kunskap (som möjligt). Men en excess av detsamma säger Porter, underminerar sådant som individuell subjektivitet och minoriteters kulturer, förminsakar artistisk kreativitet och försvårar demokratisk medverkan. Trots denna kritik är dess resonans i stort sett övervägande positiv.

Frågan om mekanisk objektivitet kan ersätta expertbedömningar kan generellt betecknas som en vetenskaplig fråga men rymmer även som vi sett politiska och kulturella aspekter: Är mekanisk objektivitet kapabel att ersätta expertkunskap? Är det ens önskvärt? Porter menar att inget enkelt svar finns på dessa frågor men han lyfter fram behovet av att börja se att idealet av mekanisk objektivitet är just ett ideal (1996, s. 213). Det finns en oartikulerad expertis och personlig involvering i alla vetenskapliga processer och beslut – det pågår så att säga ett socialt liv även i laboratorier. Kan olika objektivitetssyn möjligtvis förenas så att vi rör oss bort från frågan om mekanisk eller disciplinär objektivitet ska råda? Porter menar att dessa differentierade logiker utgör två olika ”världar” medan Timmermans och Epstein (2010) menar att även i en värld som marinerats av standarder växer nya hybrider och former av expertis fram. Dessa författare menar att riktlinjer och mekanisk (proceduere)ll objektivitet ej nödvändigtvis skapar formaliseringskrav. Om de är öppna för bedömningar kan riktlinjer driva mot disciplinär objektivitet, där även kreativitet och inre motivation tillåts främjas (Berg, Horstman, Plass & van Heusden, 2000).

Jag har med resonemangen ovan velat visa hur de ingående begreppen kan användas för att analysera och förstå utvecklingen av exempelvis psykometri och riktlinjer inom testadaptation. Oavsett var inom det vetenskapliga samhället som mekanisk objektivitet ges företräde så representerar det också gemensamma epistemologiska drag i synsätt på kunskap, människor och kultur. Hur det står till med detta inom produktionen av standarder och riktlinjer inom testadaptation analyseras längre fram i uppsatsen.

Nu övergår uppsatsen till att belysa standardiseringens sociologi. Den ökade förekomsten av standarder, eller standardiseringsrörelsen, är ett tydligt exempel på det som Porter benämner mekanisk objektivitet och den politiska kraft som ligger i standarder som sådana.



## 4.4. Standardisering och standarder

I följande avsnitt ges en bakgrundsbild till standardisering och standarder, deras ofta självklara och icke synliggjorda roll som (politiska) verktyg för påverkan i våra samhällen. Det kan vara många olika krafter involverade i framarbetandet av standarder, som innefattar förhandlingar och kompromisser på olika nivåer. Varje uppsättning av standarder har sitt eget ursprung, sina egna syften, de innefattar en sammansättning av möjligheter och begränsningar. Standardiseringsrörelsen bygger generellt på bakomliggande (mekaniska) idéperspektiv, som influerat samhällsutvecklingen och därigenom människors vardag allmänt. Dessa system av idéer påverkar idag utformning av verksamheter och praktiskt arbete inom olika disciplinära fält.

### 4.4.1. Bakgrund – standardiseringens politik

Som en del i uppbyggandet av modernitetens samhälle har olika former av nationell och global standardisering allt eftersom vuxit fram inom de flesta sektorer. Standardisering genomskär idag våra postmoderna samhällen på sätt som vi kanske sällan reflekterar över. Standardiseringar möjliggör och underlättar samarbete inom och över landsgränser, det kan gälla så vitt skilda områden som gemensamma grundläggande måtenheter, utformning och mått på verktyg, voltenhet för elektricitet, storleksangivelse på kläder till procedurer inom medicin, diagnosmanualer och internationella forskningsprotokoll. Med denna standardisering följer en uppsättning tillhörande standarder för att systemen ska fungera smidigt och tillförlitligt över tid och rum. Standarder specificerar hur vi arbetar och hur våra olika teknologier interagerar. På så vis håller de samman våra sociotekniska samhällen.

I deras bok *The Gold Standard* bidrar Timmermans och Berg (2003) med initierade perspektiv på och problematiserar standardisering (vår tillit till standarder) genom att granska och rikta uppmärksamheten mot de *socio-politiska kontexter* i vilka standarder uppstår. Författarnas tes är att *processen* att göra en standard pålitlig och ansluta sig till antaganden om dess fördelaktiga *resultat*, innefattar specifika och inte sällan problematiska konsekvenser som lätt ignoreras på vägen mot att försöka bygga förbättring en masse.

Författarna utmanar det som konstituerar ”den bästa vetenskapen” och undersöker under vilka förutsättningar enhetlighet, uniformitet, objektivitet, demokrati och universalitet uppnås (Timmermans och Berg, 2003, s. 26). De anser att standarder obevekligen är politiska i det att deras konstruktion och applicering transformerar de praktiker inom vilka de ingår i (Timmermans och Berg, 2003, s. 22 och s. 53). De förändrar aktörers positioner; förändrar ansvarsfördelning som betonar eller förändrar rådande hierarkiska ordningar med mera. Det är dessa transformationer i kölvattnet av införandet av standarder som författarna är intresserade av. Timmermans (2015) menar att intressanta nyckelfrågor för vetenskapsstudier och dess forskare är hur - i processen att skapa förbättring eller använda standarder - denna balans mellan det som extraheras fram kontra vad som exkluderas när standarder formas, påverkar och har påtaglig betydelsen för arbetet som ska utföras inom ramen för det som standarden anger. I länken mellan form och funktion finns inbäddat olika val, prioriteringar, makt eller politik om man så vill (för flertalet referenser se Timmermans, 2015, s. 81).

Ur ett större perspektiv är standardisering en produkt av en global ekonomi, en nödvändighet för en ständigt pågående uppskalning av världshandeln och dess stegrande komplexitet. Ett nutida välkänt globalt fenomen i denna standardiseringsanda är förstas McDonalds. Moderniteten kan i detta sammanhang ses som en process som betonar *teknisk* standardisering och eliminering av andra etablerade kulturellt baserade standarder (i linje med Porters resonemang om objektivitetssyn).

#### 4.4.2. Vad är standardisering och standarder? Definition och karaktäristika

Timmermans och Berg (2003, s.24 – 27) definierar specifikt *standardisering* som processen att framställa sakers uniformitet och *standarder* som både medlen för och resultaten av standardisering. I allmänt tal refererar en standard till en måttstock etablerad av auktoriteter, seder, eller generellt samförstånd att utgå ifrån som referenspunkt. En mer övergripande definition av standarder är att de utgör medel för att skapa enhetlighet över tid och rum genom framställning av fastställda tekniska regler. Denna breda definition täcker in sådant som teknisk infrastrukturell design, terminologier, prestationer och procedurer (Timmermans och Berg, 2003; Timmermans, 2015).

Timmermans och Berg (2003, s. 24 – 27) definierar fyra typer av standarder: *design*; *terminologiska*; *prestationsbaserade*; och *procedurella standarder*. Procedurella standarder som analyseras i denna uppsats väcker frågor om människors autonomi, flexibilitet, kreativitet och samarbete, rationalitet och objektivitet. I korthet anger och reflekterar procedurella standarder viktiga kulturella antaganden om hur människor lever och arbetar tillsammans.

En tänkt fördel med standarder – en av flera – är att det både ska underlätta och kvalitetssäkra för användarna. Användare ska dock inte behöva bry sig om själva framtagandet eller skapandeprocessen av standarderna i fråga. En inbyggd auktoritet förväntas attribueras till standarder, något som således skulle ligga i deras rena natur. De antas underförstått representera reliabelt innehåll och användarvänlig karaktäristika, vilket ger andemeningen att arbetet bakom med att formulera och ta fram en standard med tillförlitlighet bygger på en pålitlig och solid process. Detta representerar ett synsätt som bygger på underliggande mekaniska perspektiv på objektivitet (jmf Porter, 1996).

#### 4.4.3. Flexibel standardisering

Förespråkarna av standarder inom forskarsamhället menar att införande av sådana bidrar till att deras disciplinära kunskapsfält och dess forskare kan röra sig mot en mer ”exakt” vetenskaplighet (Porter, 1996; Timmermans och Berg, 2003). Inom evidensbaserad medicin (EBM) skulle det kunna motsvara vägen mot mer demokratisk och jämlik vård. Inom testadaptionfältet (riktlinjer) förväntas det kunna resultera i mer rättvisande testförfaringsätt för testmottagarna, genom ökad systematik och universell design för framarbetandet av standardiserade tester.

Kritiker menar att standarder ofta är formulerade och utformade på ett onyanserat sätt, med förväntningar att följa riktlinjeinstruktioner som påminner om att följa recept ur en kokbok (Timmermans & Berg, 2003, s. 69 – 70). Egen intuition och expertis förväntas ställas åt sidan, något som rimmar illa i miljöer där det finns otalet kulturella och etiska frågor att hantera, så som inom vården (eller inom testadaption). En identifierad grundrädsla från kritiker inför standarder är att dehumaniseras och att det arbete som utförs blir statiskt och livlöst utan någon egen inre drivkraft från de som utför arbetet. Dessa farhågor kan tänkas förekomma inom andra disciplinära fält än de medicinska. Detta återkommer vi till nedan.

Istället för att iklä sig en position av passivisering och maktlöshet vill Timmermans och Berg istället uttala och föreslå en aktiv aktörsroll i att arbeta med standarder, som en aktiv handling av att tillåta sig att bli transformerad samtidigt som man i sin professionella roll transformerar standarden (2003, s.73). Standardisering som kännetecknas av differentiering och som för bort från enkla lösningar utgörs av flexibilitet i implementering och praktisk tillämpning av standarder. Framgången med en sådan design är ytterst beroende av möjligheterna som ges till de professionella att skickligt integrera standardernas krav utifrån lokala förutsättningar och med de verktyg och metoder de har tillgång till i deras arbetspraktik.

Timmermans och Berg summerar i sin bok ett antal förslag på hur få procedurella standarder att fungera i praktiken (2003, se sid. 207 – 216 för fullständig beskrivning). Delar av dessa förslag ser jag möjligen som tillämpbara inom riktlinjeutformning för praktiskt inriktad testadaption. Detta

återkommer vi till under diskussionsdelen nedan men i korthet kan nämnas utformning av det som de kallar *flexibla standarder* – ej mer detaljerade än behövt, ej mer stringenta än nödvändigt, ej mer tvingande än användbart (s. 211). Utifrån syfte med och de behov som följer med anpassning av ett test kan man fråga om det kan finnas flera alternativ för *omfattning* av testadaptnionsprocessens respektive ingående steg? Kan det finnas olika (mer eller mindre resurskrävande) alternativ för de olika (obligatoriska) stegen? För att som riktlinjeförfattare arbeta med sådana frågor kan det tänkas att det finns värde i att involvera fler aktörer och deras perspektiv, till exempel att skapa forum för att involvera testanpassade praktiker i utformning av riktlinjer. Huruvida det finns behov av sådant eller liknande återkommer vi till i analysen nedan.

#### **4.4.4. Sammanfattning**

I grund och botten så är det så att många intressen tenderar krocka eller ställas mot varandra under framarbetandet och implementering av standarder. En fråga som jag ställer mig är: Hur bidrar standarder med förflyttningen av relevansen i ett psykologiskt test från ett språk och kultur till ett annat språk och kultur? I den myriad av språkliga och kulturella aspekter att beakta för att genomföra en sådan övergång blir standarder förståeligt attraktiva verktyg för att erhålla klinisk eller vetenskaplig payoff: de ger stabilitet, auktoritet och funktionalitet (Timmermans, 2015, s. 81). De förmedlar bilden av att kunna undvika fallgropar, spara tid och resurser och skapa tester som är giltiga, både där och här. Vid betoning på tillit till standarder devalveras dock oundvikligt andra alternativa källor till tillit. Samtidigt, med flexibel standardisering kan även den professionella och disciplinära objektiviteten – som motsats till ensidig mekanisk objektivitet – ges utrymme.

Efter att ha redovisat hur standardiseringsrörelsen gått i linje med modernitetens härskande och underliggande mekaniska kunskapsideal (Porter, 1996) samt problematiserat standarder som sådana (Timmermans och Berg, 2003) är det dags att gå över till kartläggning av de ramverk och riktlinjer som ingår i uppsatsen.

## 5. Ramverk och riktlinjer: kartläggning och jämförelse

### 5.1. Introduktion

För att fortsätta besvara uppsatsens frågeställningar (huvudsakligen nummer 1 – 2) som berör att undersöka produktion, utformning och innehåll presenteras i avsnittet nedan en översikt av urvalet av ramverk och riktlinjer. Vi rör oss parallellt i ögla nummer 2 och 5 i Latours modell i detta avsnitt med inriktning på riktlinjernas struktur, beståndsdelar och innehåll (öglan 5) och indirekt riktlinjeförfattarnas kollegiala samhälle (öglan 2). Ambitionen under detta avsnitt är främst att kort beskriva ramverk och riktlinjer. Detta för att senare jämföra ramverk och riktlinjer och ta en analytisk granskande blick på desamma gällande objektivitetssyn och perspektiv på standarder som sådana. För utförligare beskrivning av respektive ramverk och riktlinjer se hänvisningar till referenser.

Någon kanske frågar sig vad skillnaden är mellan främst *ramverk* och *riktlinjer* men kanske också *modeller*. För att undvika begreppsförvirring ger jag förklaringar, giltiga i sammanhanget. Modeller som sådana, ger vanligtvis en visuell presentation eller representation, till exempel i form av en figur eller en bild av ett system, som är tänkt att kunna replikeras. Ramverk däremot innefattar och ger en större konceptuell struktur, med en ram av referenser som även kan innefatta beskrivningar av specifika (teoretiska eller epistemologiska) angreppssätt. Slutligen, riktlinjer är vanligtvis mer specifika till sin karaktär, med målinriktade rekommendationer för professionella praktiska tillvägagångssätt (Bernal & Rodriguez, 2012, s. 23 – 24). Ett sätt att tänka är att modeller skulle kunna ”stå på egna ben” eller ingå som visuellt stöd i ramverk och riktlinjer. Riktlinjers innehåll skulle kunna byggas utifrån ett specifikt ramverk men jag har svårt att tänka mig det omvända. Modeller, ramverk och riktlinjer för testadaptation erbjuder perspektiv som varierar påtagligt i omfattning, bredd och djup. Begreppen ramverk och riktlinjer används uteslutande för de fyra ramverk och riktlinjer som analyseras i denna studie.

Målsättningen i detta avsnitt är att översiktligt summera urvalet av ramverk och riktlinjer för testadaptation. Samtliga ramverk och riktlinjer har - enligt min bedömning - sin grund mer eller mindre inom fältet för praktisk tillämpning av forskning (eng. *research utilization* eller *research use*). I detta fall innebär det att vart och ett syftar till att underlätta planering och genomförande av anpassningsarbetet för testutvecklare, med inriktning att främja genomgående hög testreliabilitet och testvaliditet samt konceptuell relevans. Ramverken och riktlinjerna har sinsemellan vunnit olika grader av genomslagskraft vad gäller tillämpning och spridning. I nästa avsnitt följer en kort introduktion till vart och ett av dessa. Det som återges är sammandrag och utvalda delar av den text som direkt ingår i, eller från anslutande publikationer, för respektive ramverk eller riktlinjesammanställning.

### 5.2. Standards for Educational and Psychological Testing

*Standards for Educational and Psychological Testing* (American Educational Research Association (AERA), American Psychological Association (APA) & National Council on Measurement in Education (NCME), 2014) - fortsättningsvis benämnd *Standards* – kan ses som ett ramverk i form av en handbok, som innefattar fastställda normer för utbildningsorienterade och psykologiska tester och testprogram. Den första versionen av *Standards* publicerades av American Psychological Association (APA) redan 1954. Inför den senaste och därmed den sjätte revisionen lades det bland annat särskild vikt vid att se över och förfinas normerande standarder för testers applicerbarhet för internationell kontext, vilket innefattar validering av och främjande av generaliserbarhet för tester som ska översättas och tvärkulturellt anpassas. Två större koncept som inkluderats i den senaste versionen är *tillgänglighet* (som berör att främja rättvisa gällande rättvisande villkor för alla som ska genomföra och svara på tester) och *universell design* (ett förhållningssätt till testdesign som syftar till att

maximera möjlighet till tillgänglighet för alla avsedda test-målgrupper). För mer övergripande information om de huvudsakliga förändringarna som genomfördes från 1999 års version (det vill säga den föregående upplagan) till nuvarande version, se gärna Plake och Wise (2014).

Syftet med Standards är att främja och stödja sund och etisk användning av tester och att erbjuda en vägledande bas för att bedöma och utvärdera kvaliteten på olika testmetoder och testpraxis.

Målgrupper för detta ramverk är: testutvecklare; förläggare/testutgivare samt; test-användare och skall upplysa och guida vid utveckling, val och användning av instrument (Plake och Wise, 2014, s. 6).

Standards kan inte ses som en generell checklista för testadaptionprocesser (American Educational Research Association et al., 2014, s. 7). Vissa av de ingående, riktgivande normerna ges med mycket breda övergripande penseldrag medan andra är betydligt mer detaljerade till sin karaktär.

Sammanvägd och professionell bedömning som baseras på bland annat beteendevetenskap, psykometri och andra relevanta normer inom det fält som testet är avsett att användas ska beaktas. Tester skiljer sig åt inom ett antal dimensioner och det behöver tas hänsyn till vid arbete med testadaption menar författarna.

Handboken utgår oftast från begreppen *test* och *testning* med betoning mot standardiserade tester, exempelvis begåvnings, prestations- och personlighetstester. De normerande standarder som sammanställts är dock även avsedda för och applicerbara på andra typer av test (författarens notering: alltså olika typer av instrument som diskuterats ovan men som innefattar någon form av psykometri).

Varje kapitel inleds med en introducerande normativ text som ger övergripande bakgrund till de specifika standarder som följer i kapitlet. Till exempel rubriceras det första kapitlet *Validity* och det tredje kapitlet *Fairness*. Dessa övergripande texter utgör ett slags organiserande tema för respektive kapitelns innehåll. Därefter inordnas kapitlets standarder i ett antal integrerade kluster. Till exempel i kapitlet som rubriceras *Fairness*<sup>3</sup> (s. 49) hittar vi först en inledande text på detta tema (*rättvisa*), därefter ett antal kluster som i sig innehåller ett antal numererade standarder. Kluster nummer 3, i samma kapitel: *Accommodations to Remove Construct-Irrelevant Barriers and Support Valid Interpretations of Scores for Their Intended Uses*, innehåller i sig sex separata och numererade standarder, där en av dessa utgörs av en standard som har ett direkt fokus på översättning och testadaption (*Standard 3.12, s. 68*). På detta sätt är boken organiserad genom samtliga kapitel.

Handboken omfattar totalt tretton kapitel och omfattar således ett brett spektrum av stöd och rekommendationer för utveckling, användning och bedömning av olika tester, alltså inte endast med inriktning mot översättning och anpassning av tester. Handboken rekommenderas av BPS/PTC och svenska Psykologförbundet.

### 5.3. International Test Commission (ITC)

ITC:s riktlinjer (International Test Commission, 2017) för adaption av tester har utvecklats genom en internationell kommitté bestående av representanter från ett antal internationella organisationer och forskare. Under 2017 uppdaterades ITC's riktlinjer. Detta är den tredje utgåvan (första utgåvan kom 1999 och den andra 2005). Målsättning har i samtliga versionerna varit att ta fram en detaljerad uppsättning riktlinjer för att anpassa psykologiska och pedagogiska tester för användning i olika språkliga och kulturella sammanhang. ITC:s riktlinjer för adaption är avsedda att tillämpas överallt där tester flyttas från en kulturell miljö till en annan - oavsett om det finns ett behov av lingvistisk översättning, eller om behov av ett tests standardisering föreligger eller ej. De tillhandahåller även riktlinjer med andra syften inom testutveckling.

---

<sup>3</sup> American Psychological Association et al. (2014) viger ett helt kapitel till *fairness* (kap. 3) som jag översatt till *rättvisa*. Begreppet är centralt i samtliga riktlinjesammanställningar som ingår i och analyseras i uppsatsen. Rättvisa som begrepp definieras utifrån sammanhanget och belyses vidare nedan i uppsatsen.

En riktlinje definieras inom ITC's arbete som: "en viktig praxis för att genomföra och utvärdera anpassningen (även ibland kallad lokalisering) eller samtidig utveckling av psykologiska och pedagogiska test för användning i olika populationer" (International Test Commission, 2017, s. 8).

Arton (18) riktlinjer delas in i sex breda områden: *Förutsättningar* (3); *Testutveckling* (5); *Konfirmering* (Empiriska analyser) (4); *Administration* (2); *Skalor och tolkning* (2); *Dokumentation* (2).

Det första avsnittet som belyser *förutsättningar* lyfter fram det faktum att beslut måste fattas, till exempel om ett test lämpar sig för att anpassas innan översättning och anpassningsprocessen någonsin börjar. Den andra sektionen - Riktlinjer för *testutveckling* - är inriktad på den faktiska processen att anpassa ett test. Det tredje avsnittet som berör *konfirmering* via empiriska analyser innehåller de riktlinjer som är kopplade till sammanställning av empiriska bevis för att hantera ekvivalens, tillförlitlighet (reliabilitet) och giltighet (validitet) för ett test som ska användas på flera språk och kulturer. De sista tre delarna är relaterade till *administration; poängskalor och tolkning*; samt *dokumentation*. Dokumentation har varit ett särskilt försummat ämne i testanpassningsinitiativ inom psykologi och utbildning menar ITC och de vill se att tidskriftsredaktörer och testutgivningsföretag kräver mer när det gäller dokumentation av ansvariga för testanpassningsprocessen (International Test Commission, 2017, s. 28).

För varje riktlinje erbjuds en förklaring för och förslag på hur att implementera den i praktiken. För att ge en inblick i innehåll och struktur ges exempel på en av fyra riktlinjer som berör riktlinjeområdet *testutveckling*, med tillhörande förklarande text och praktiska förslag från ITC (2017, s. 11 - 12, betydligt nedkortat på grund av utrymmesskäl. Svensk översättning är författarens):

(TD-1) *Se till att översättning och anpassningsprocesser tar hänsyn till språkliga, psykologiska och kulturella skillnader i de avsedda befolkningarna genom att välja experter med relevant kompetens.*

Förklaring:

Kunskap om kulturen, allmän kunskap om ämnet och testkonstruktionen, har de senaste åren blivit en del av urvalskriterierna för översättare. Denna riktlinje verkar ha varit inflytelserik i att uppmuntra organisationer som översätter och anpassar tester till att använda minst två översättare i olika design (till exempel framåt- och tillbakaöversättning).

Kunskap / expertis i målkulturen uppnås genom att använda översättare som har målspråket som sitt modersmål och lever i målområdet, där det första är nödvändigt och det senare är mycket önskvärt.

Vår (det vill säga ITC, författarens förtydligande) definition av en "expert" är då en person eller ett team med tillräcklig kombinerad kunskap om (1) de berörda språken, (2) kulturen, (3) testets innehåll och (4) allmänna principer för testning, för att producera en professionell kvalitetsöversättning / -anpassning av ett test.

I praktiken kan det vara effektivt att använda team av personer med olika kvalifikationer (till exempel översättare med och utan expertis inom ämnet, en testexpert, etc.) för att identifiera områden som andra kan förbise. I alla fall bör kunskap om allmänna test-principer utöver kunskap om test-innehållet ingå i den utbildning som översättarna får.

Förslag till praktik (Författarens notering: ITC föreslår följande):

- Välj översättare som har sitt modersmål i målspråket och har en djup kännedom om kulturen som testet ska anpassas till, helst bör personen bo i målområdet. Ett vanligt misstag är att identifiera

personer som översättare som känner till språket, men som inte kan kulturen så bra, eftersom en djup kännedom om kulturen ofta är nödvändig för att upprätthålla kulturell ekvivalens. Att ha kulturell kunskap kommer att identifiera kulturella referenser (t.ex. cricket, Eiffeltornet, president Lincoln, känguru etc.), med vilka lokala deltagare kan vara obekanta.

- Översättare med kunskap om testutvecklings-principer kan vara praktiskt taget omöjliga att hitta och det är därför viktigt att tillhandahålla utbildning för översättare för att ge dem kunskap om principerna för de item-format som de ska arbeta med. Utan träningen kommer ibland alltför noggranna översättare att introducera felkällor, vilket kan sänka validiteten för ett översatt test. Till exempel kan en översättare ibland lägga till en klagande anmärkning för att säkerställa att ett riktigt korrekt svar faktiskt är det rätta svaret. På så sätt kan översättaren göra ett item enklare än vad det var avsett, eller det längre korrekta svaret kan ge en ledtråd till det korrekta svaret för erfarna testkandidater.

Flertalet artiklar och kapitel i böcker har allteftersom publicerats parallellt med utvecklingen av dessa riktlinjer (första och andra utgåvan) som mer detaljerat beskriver hur de ska läsas och tillämpas. För fördjupad inläsning om ITC:s riktlinjer se även Hambleton et al. (2005, även om den boken bygger på den första utgåvan av riktlinjerna återfinns flera kapitel som ännu är relevanta menar författarna) samt <https://www.intestcom.org>.

Ett större redigeringsarbete har skett i och med andra utgåvan men det hänvisas till äldre källor som även ingick i första utgåvan (så som Tanzer och Sim, 1999; van de Vijver och Hambleton, 1996; och van de Vijver och Tanzer, 1997). Exempel på några nytillkomna referenser som rekommenderas att läsas som tillägg är: Grégoire och Hambleton (2009) och Rios och Sireci (2014). De flesta artiklarna publiceras i ITC's egen tidskrift. Riktlinjerna rekommenderas av BPS/PTC, EFPA och svenska Psykologförbundet. ITC hänvisar själva återkommande till Standards (American Educational Research Association et al., se ovan) i olika utgåvor.

## 5.4. Guidelines for the process of Cross-Cultural Adaption of Self-Report Measures

*Guidelines for the Process of Cross-Cultural Adaption of Self-Report Measures* (Beaton, Bombarider och Guillemin, 2000) grundar sig på forskning inom medicin, psykologi och sociologi samt på publicerade metodologiska ramverk med inriktning mot tvärkulturella adaptationsprocesser. Riktlinjerna är initialt framarbetade för att användas vid adaptation av hälso- och sjukvårdsrelaterade självskattningsformulär. De används dock i vidare sammanhang (se återgivning av intervju med Berthelsen nedan). Från författarna ligger en betoning på procedurer för att främja rättvisa i testsammanhang för testmottagande grupper med annan etnisk och/eller kulturell bakgrund än värdkulturen (som originalversion av testet är utformat för).

Som initial vägledning för att bedöma testadaptionens lämplighet, systematiseras olika möjliga scenarion som anger när och om, och i så fall vilka former av tvärkulturell adaptation som krävs vid de olika ponerade situationerna (se tabell 1).

Table 1. Possible Scenarios Where Some Form of Cross-Cultural Adaptation is Required

Wanting to use a questionnaire in a new population described as follows:	<u>Results in a Change in . . .</u>			<u>Adaptation Required</u>	
	Culture	Language	Country of Use	Translation	Cultural Adaption
<b>A</b> Use in same population.  No change in culture, language, or country from source	—	—	—	—	—
<b>B</b> Use in established immigrants in source country	√	—	—	—	√
<b>C</b> Use in other country, same language	√	—	√	—	√
<b>D</b> Use in new immigrants, not English-speaking, but in same source country	√	√	—	√	√
<b>E</b> Use in another country and another language	√	√	√	√	√

Beaton et al., 2000, s. 3187

Riktlinjerna som sådana är i den första fasen indelade i sex olika på varandra följande steg:

- Översättning
- Skapa syntes av översättningar
- Tillbaka-översättning
- Granskande expertgrupp (granskning av testanpassningen)
- För-testning
- Löpande granskning (något som även sker mellan samtliga steg ovan)

I sin helhet avser riktlinjerna att utgöra stöd för fullskalig testadaptation. Författarna betonar att säkerställande av ekvivalens gällande *inhåll* mellan original och översättning inte garanterar reliabilitet och validitet på item- och skalnivå (författarna hänvisar även till Herdman, Fox-Rushby och Badia (1998) som myntat begreppet *funktionell ekvivalens*. Begreppet beskrivs under avsnitt 6.2, s.27 nedan). Därför förespråkar Beaton et al. fortsatta tester för de psykometriska egenskaperna hos det anpassade testet *efter* att översättningsprocessen är genomförd (2000, s. 3189). I deras modell utgör ovan beskrivna riktlinjer första fasen i en trestegsprocess. De övriga två faserna omfattar i korthet:

1. Verifiering av skalor (adekvata item/item-prestanda, viktning av item, internkonsistens m.m.)
2. Validering av och etablering av normativa värden för den anpassade versionen (konstrukt-/begreppsvalidering, reliabilitet och responsivitet m.m).



Den process som beskrivs innefattar möjlighet till att byta ut items eller att förändra skalor för att göra ett test relevant för en ny kultur (Beaton et al., 2000, s. 3190).

Ingen av de branschorganisationer jag varit i kontakt med nämner dessa riktlinjer i deras rekommendationer (BPS/PTC, Psykologförbundet, EFPA).

## 5.5. Translating and Adapting Measurement Instruments for Cross-Linguistic and Cross-Cultural Research: A Guide for Practitioners

Kristjansson, Desrochers och Zumbo's (2003) *Translating and Adapting Measurement Instruments for Cross-Linguistic and Cross-Cultural Research: A Guide for Practitioners* är en riktlinjeguide framarbetad för att undvika vanliga fallgropar i anpassnings- och översättningsprocessen. Författarna har sin bakgrund inom psykologi, pedagogik och global hälsa. Med riktlinjerna vänder de sig främst till forskare som planerar att utveckla eller använda översatta instrument inom sitt disciplinära fält. De har sammanställt och syntetiserat rekommendationer från en rad olika forskare, några av dem aktiva inom ITC (se ovan).

Författarna adresserar jämte själva riktlinjerna, metodologiska problem som kan äventyra tillförlitlighet i tolkning av testdata (på item- och skalnivå). Det ultimata målet med att översätta och anpassa ett test är just att intakt bibehålla innehållsbetydelse och teoretiska konstrukt eller koncept resonerar de (i sällskap med övriga riktlinjeproducenter i denna uppsats). De legitimerar riktlinjernas bärighet via flertalet exempel på faktiska problemsituationer, som även rapporterats in för översatta och anpassade tester (för ingående beskrivning av exempel på sådana problem för vedertagna, validerade tester, se Kristjansson et al., 2003, s. 129 - 135).

Några av de vanligaste problemen och fallgroparna vid testutveckling i tvärkulturella sammanhang som kan leda till bias utgörs enligt författarna av tre typer:

- brist på konceptuell likvärdighet
- skillnader i kulturella normer
- brist på semantisk likvärdighet

Författarna menar att just semantiska problem är vanliga och samtidigt förhållandevis enkla att åtgärda.

Riktlinjerna som sådana är indelade i fem och i ordning följande steg:

1. Verifiering av konceptets relevans
2. Översättning av instrumentet och utveckling av preliminära versioner (inklusive tillbaka-översättning)
3. Expertgrupp – översikt och utvärdering av den preliminära versionen
4. För-testning av testet
5. Pilot-test av testet

Den initiala pilottestningen följs upp av en större pilottestning (fas 2) med faktor- och item-analyser. Referenser gällande procedurer för testning av konceptuell – och item – överensstämmelse för båda versionerna av testet liksom etablering av jämförande normgrupper med mera återfinns i riktlinjeartikeln.

Ett intressant tillägg är att författarna förespråkar att initialt undersöka etiska-etiska perspektiv för att avgöra om ett test överhuvudtaget är lämpligt att översätta (s. 136, något som tagits upp av International Test Commission (2017) i den senaste versionen av deras riktlinjer).

Kristjansson et al. (2003) hänvisar frekvent till artikelförfattare med anknytning till ITC. Bruno

Zumbo, som är en av författarna till Kristjansson et al:s riktlinjer, ingick i kommittén för att arbeta fram ITC's andra utgåva av riktlinjer (de gavs ut 2005, Kristjansson et al:s riktlinjer publicerades 2003).

Ingen av de branschorganisationer jag varit i kontakt med nämner Kristjansson et al:s riktlinjer i deras rekommendationer (BPS/PTC, Psykologförbundet och EFPA)

Efter att kort ha presenterat ramverk och riktlinjer som ingår i denna studie följer nu en presentation av några grundläggande teman som ingår i riktlinjernas innehåll och syftessträvan.

## 6. Gemensamma teman i riktlinjernas innehåll

För att få en fortsatt inblick i riktlinjeförfattarnas synsätt på testadaption finns ett antal gemensamma teman och begrepp som behöver belysas: ekvivalens, universell design och psykometri. För riktlinjeförfattarna är ekvivalens överordnad de övriga två. Universell design och psykometri används för målet att uppnå ekvivalens. Dessa begrepp är viktiga att förklara och redogöra för, inför fortsatt analys och resonemang framåt i studien.

I detta avsnitt fokuseras uppsatsens två första frågeställningar: riktlinjerna och deras utformning och innehåll men även den tredje frågeställningen introduceras i den kritik som läggs fram och frågan om hur man kan åstadkomma ekvivalens mellan kulturer. Vi utgår i Latours modell från ögla nummer 5 – riktlinjerna och deras innehåll – och ögla nummer 2 – riktlinjeförfattarnas kollegiala och disciplinära kriterieutformning.

### 6.1. Introduktion till testadaption med riktlinjeförfattarnas perspektiv

Eftersom uppsatsen undersöker och analyserar riktlinjer för testadaption för psykologiska test är det nära till hands att inledningsvis fråga sig: Vad är ett psykologiskt test för någonting? En koncis version presenteras av American Educational Research Association et al. (1999, s. 3): “A test is any evaluative device or procedure in which a sample of examinee’s behaviour in a specified domain is obtained and subsequently evaluated and scored using a standardized process”.

Med psykologiska test strävas efter att mäta egenskaper och tillstånd som inte *direkt* kan observeras, kartläggas och kvantifieras. Exempel på sådana egenskaper kan vara begåvning, personlighetsstruktur, psykopatologi (psykiska störningar och psykiska sjukdomar) och neuropsykologiska störningar och sjukdomar men även attityder, preferenser och olika former av dispositioner.

Benämningen *test* är vanligtvis reserverad för instrument där respondenternas responser värderas utifrån deras riktighet/korrekthet eller kvalitet (American Educational Research Association et al., 2014, s. 2). Termerna *skala*, *inventorium* eller *frågeformulär* används vanligtvis för mätningar av sådant som attityder, intressen, och dispositioner. Så som deklarerades i inledningen, benämningen *test* får i uppsatsen representera alla ovan nämnda typer av (ut)värderande instrument (så som inventorium, frågeformulär och enkäter).

Enligt författare inom testadaptionens fält (American Educational Research Association et al., 2014; Hambleton et al., 2005) så har *standardiserade* psykologiska test, oavsett tillämpningsområde, utvecklats med beaktande av vissa procedurer för att säkerställa att testet mäter vad det ska mäta (att testet har validitet) på ett tillförlitligt sätt (att testet har reliabilitet). Kritiska perspektiv på standardiserade test och karaktären på standarder för att utveckla test läggs fram till exempel av Porter (1996), som menar att detta ingår i en renodlad mekanisk objektivitetssyn. Se även vidare resonemang nedan under avsnittet för psykometri (s. 28).

I testadaptionlitteraturen som ingår i denna studie föredras termen (*tvärkulturell*) *adaption* istället för *översättning* när det refereras till processen att utveckla tester och mätinstrument mellan språk och/eller kulturer. Översättning ses således som endast en del i mer omfattande processteg. Begreppet *tvärkulturell adaption* omfattar således den process som beaktar både språkliga (översättning) och kulturella faktorer för anpassning av befintliga test för att användas i annan miljö än vad den ursprungliga versionen designades för. *Adaption*, *anpassning* och *anpassningsprocesser* får nedan utgöra samlingsbegrepp för ovan resonemang och de anpassningsprocesser som syftar till strävan efter ekvivalens mellan testversioner.

Härnäst redogörs för riktlinjeförfattarnas perspektiv på ekvivalens, universell design och psykometri samt initiala kritiska perspektiv på dessa begrepp i sammanhanget.

## 6.2. Ekvivalens

Ett överordnat begrepp, eller en överordnad strävan, som återfinns inom den teoretiska sidan av testadaption är *ekvivalens*. Likvärdighet mellan de olika testversionerna är sammantaget vad ambitionen med adaption av test syftar till, utifrån riktlinjeförfattarnas perspektiv. För att säkerställa konceptuell relevans och kunna bedriva tvärkulturella jämförelser (till exempel inom forskning) bör testutvecklare således eftersträva att uppnå ekvivalens mellan test. Herdman, Fox-Rushby och Badia (1998) har definierat ekvivalens utifrån olika underkategorier:

- Konceptuell ekvivalens: domäner har samma relevans, mening och betydelse för det koncept som utforskas i de båda kulturerna.
- Item-ekvivalens: Items är lika relevanta och acceptabla i båda kulturerna.
- Semantisk ekvivalens: Betydelsen av items är samma i båda kulturerna.
- Operativ ekvivalens: frågeformuläret kan användas på samma sätt av målgrupperna i båda kulturerna; till exempel ett självrapporterat frågeformulär förutsätter läskunnighet, och ett online-frågeformulär kan vara svårare att använda i vissa områden med låg tillgång till internet, med andra ord låg grad av operationell ekvivalens.
- Mät-ekvivalens: Ingen signifikant skillnad i psykometriska egenskaper (konstrukt-validitet, reliabilitet, responsivitet etcetera) av de två testversionerna.

*Funktionell ekvivalens* är en sammanfattning av de ovan benämnda fem ekvivalens-kategorierna, med betydelsen att båda versionerna av instrumentet ”gör vad det ska göra lika bra”. Dessa beståndsdelar refererar samtliga ramverk och riktlinjeförfattare till i olika utsträckning och på olika sätt.<sup>4</sup>

Sätten, eller möjligheterna, att uppnå ekvivalens beror i första hand på det skede där frågan om detta (att uppnå likvärdighet) uppstår. I början av den tvärkulturella adaptationsprocessen, innan någon översättning påbörjats, bör den potentiella konceptuella ekvivalensen undersökas och utvärderas. Detta rekommenderas av Herdman et al. (1998) men adresseras även av de riktlinjeförfattare som ingår i denna uppsats. Dock tycks denna utvärdering sällan genomföras enligt samma författare.

Konceptet ”ekvivalens mellan språk i tvärkulturell anpassning av test” i litteraturen innefattar dock olika definitioner och bärande ramar enligt Epstein, Miyuki och Guillemain, (2015, författarens notering: Guillemain är en av författarna till Beaton et al:s (2000) riktlinjer). De kritiserar fältet och menar att riktlinjemodellerna överlag både spretar och saknar utvärdering. Dessa hävdar att de endast vilar på ett antagande om att vara funktionella (ekvivalens enligt Herdman et al:s definition, 1998). Epstein et al. (2015) vill se en noggrannare utvärdering och färre, enhetligare proceduriska beskrivningar och menar att nuvarande situation är undermålig. De vill se en utveckling mot mer likriktade standarder, där en ”gyllene modell” eftersträvas.

Med Herdman et al:s (1998) definition av funktionell ekvivalens som utgångspunkt menar Epstein et al. (2015) att det saknas bevis på någon överlägsenhet för någon metod (som återges i publicerade riktlinjer) över en annan. Mer forskning behövs menar de för att tillhandahålla evidens som stöder nuvarande praxis, så som det återges i olika riktlinjemodeller. Omfattande diskussioner har löpande pågått i litteraturen om möjligheter och begränsningar vad gäller översättbarhet.<sup>3</sup> Frågan om huruvida ekvivalens är möjlig att uppnå återkommer vi till längre fram i analysen.

<sup>4</sup> För exempel på detta från respektive ramverk och riktlinjesammanställning: American Educational Research Association et al. (2014) s. 53 – 54 och s. 69; International test Commission (2017) s. 9 och s. 18; Beaton et al. (2000) s. 3186 – 3187; samt Kristjansson et al. (2003) s. 128 – 129 och s. 131.

<sup>3</sup> Caro och Stiles (1997) ser översättning och anpassning av tester som en akt av re-kreation, att rekonstruera och ”bygga om”. Andra har deklarerat att en av de knivigaste uppgifterna genom hela den kosmiska evolutionen kan möjligtvis ha varit översättning (Candell & Hulin, 1987) medan vissa hävdar till och med att detta bör undvikas (Triandis, 1994) utifrån det vanskliga i att tro att kultur låter sig överföras via översatta frågeformulär. En anledning till att denna uppgift anses så komplex är att översättning inte utgörs av en enkel mekanisk insats bestående av att ändra ord från ett språk till ett annat, utan en subtil och invecklad uppgift. Vikten av anpassningsprocessens kvalitet och

### 6.3. Universell design

*Universell design* (eng. *universal design*) är ett förhållningssätt till testdesign och instrumentutveckling som betonas hos riktlinjeförfattarna. Begreppet används främst i det ramverk som benämns Standards (American Educational Research Association et al., 2014) men tankegångarna – om än uttryckta med andra begrepp – är explicita även i de övriga riktlinjesammanställningar som presenteras i denna uppsats.

Med universell design eftersträvas att maximera tillgängligheten för samtliga för testet avsedda målgrupper. Universell design ger emfas till behovet av att utveckla tester som är så användbara som möjligt för alla testmottagare i den avsedda populationen, oavsett karaktäristika som kön, ålder, språk, bakgrund, kultur, socioekonomisk status, eller funktionsnedsättning. Det kräver att testdesigners skapar tydlighet gällande konstrukt (begrepp) och skalor som ska mätas, inkluderande mål och syfte med mätning/undersökning, syftet med respektive testpoäng, samt slutsatserna som kommer att dras från testpoängerna. Designen syftar till att minimera risken för konstrukt-irrelevant varians mellan målpopulationerna (jmf. ekvivalens ovan) enligt riktlinjeförfattarna.

Även om principerna inom universell design förser med användbar och nyttig guidning för testutveckling så vill American Educational Research Association et al. (2014) betona att uppbyggandet av empirisk evidens för designens hållbarhet, för att minimera konstrukt-irrelevant varians, pågår men att fortsatt forskning behövs som stöd för vissa av dessa ingående principer (s. 58). American Educational Research Association et al. menar att alla test inte kan göras tillgängliga för alla via att uppmärksamma förändringar i tillvägagångssätt i själva testdesignen. Författarna till samtliga ramverk och riktlinjer som kartläggs och analyseras i denna uppsats menar att psykologiska koncept inte per automatik är universella.

### 6.4. Psykometri

Psykometri har en framträdande roll inom utveckling av (standardiserade) psykologiska test och därför även inom testadaptation, under processen att utvärdera anpassade test. Psykometri är en statistisk disciplin med syfte att granska kvalitet hos psykologiska tester. Psykometri kommer från grekiskans *psyche* och *metron*, vilket betyder själ och mått (Nationalencyklopedin). Med hjälp av psykometriska metoder och instrument skapas ”mätning av själen”. Psykometri är dock inte en metod i sig för att mäta psyket eller personligheten, utan avser metoder som kan granska och utvärdera tester som säger sig mäta psyket/personligheten.

Att mäta psykiska egenskaper är svårt av naturliga skäl, då tankar, känslor och attityder inte går att mäta eller väga (på samma sätt som en människas längd eller vikten hos ett knippe morötter). Därför utgår forskningen från indirekta mått. Riktlinjeförfattare och förespråkare för psykometrisk utvärdering menar att en anpassad statistisk disciplin behövs för att utvärdera det som observeras (American Educational Research Association et al., 2014; Beaton et al., 2000; Hambleton et al., 2005; ITC, 2017; Kristjansson et al., 2003). Detta eftersom det sker en översättning av humör, motivation, samarbetsförmåga med mera till siffror, så att även beräkningar och jämförelse med normdata låter sig göras.

Testutvecklare rekommenderas av riktlinjeförfattarna att presentera och ge test en kvalitetsreferens genom tillhörande psykometriska analyser, som uttrycks framför allt i olika former av begreppen

---

stråvan efter ekvivalens mellan testversionerna, för att stödja jämförbara slutledningar mellan olika testgrupper lyfts fram av olika författare (till exempel American Educational Research Association et al., 2014; Hambleton et al., 2005). Även om felaktigheter aldrig går att eliminera vid en sådan översättnings- och anpassningsprocess så är tesen att de går att minimera (till exempel Geisinger, 1994; Hambleton et al., 2005; Herdman et al., 1998; Kristjansson, Desrochers & Zumbo, 2003).

*validitet* och *reliabilitet*. För fördjupade resonemang om psykometri i sig, se hänvisningar i originaltext för ramverk och riktlinjer som presenteras i uppsatsen men även till exempel Shadish, Cook & Campbell (2003).

Kortfattat kan sägas att riktlinjeförfattarna anser det eftersträvansvärt att skapa så goda psykometriska värden som möjligt utifrån ett tests avsedda konstrukt (det koncept eller begrepp som ska mätas). Det ger generellt sett ett test en högre kvalitetsstämpel och anses vara mer tillförlitligt och användbart, än tester med sämre sådana värden.

Som återgetts ovan finns kritiska röster gentemot numerisk kvantifiering och psykometri som statistisk disciplin och vad som egentligen kan uppnås utifrån psykometriska utvärderande värden. Kritiken berör således bland annat vikten av att uppmärksamma risken för övertro på bärigheten i psykometriska värden och att dessa snarare tenderar ge *sken av* kausala samband och (icke subjektivt influerad) objektivitet.<sup>5</sup> Inom naturvetenskapen har visserligen numerisk kvantifiering varit en ”framgångssaga”. Porter (1996) frågar sig varför denna väg till framgång och giltighet, varför denna modell för kunskapsbyggande – där man studerar stjärnor, molekyler, kvarkar eller celler – skulle vara detsamma inom studier av samhällsvetenskaplig eller psykologisk och socialorienterad forskning. Metoden i sig banar väg för något, nämligen ett tillstånd av icke-ifrågasättbarhet och utgör på så vis en social strategi (Porter, 1996, s. 199 – 200; 228).

Det framstår som att ekvivalens, universell design och psykometri som tematiskt innehåll går i linje med ett epistemologiskt ideal som tycks ligga till grund för riktlinjeförfattarnas syn på vetenskap och vetenskaplig produktion.

Efter att ha belyst några centrala teman och begrepp inom testadaption så är det dags att göra några sammanfattande och jämförande reflektioner.

## 6.5. Jämförelse: variationer på samma kamp för ekvivalens

De ramverk och riktlinjer som kartlagts ovan är tillsammans avsedda att fungera som stödjande och guidande standarder för processer rörande språkliga och kulturella testadaptioner. För inte allt för länge sedan ansågs tillförlitlig testadaption säkerställas via procedurer för tillbaka-översättning, enligt flera riktlinjeförfattare. Vi kan dock se att mycket vatten har hunnit rinna under broarna sedan dess där pendelrörelsen har förflyttat sig mot att värna kulturellt sensitiv anpassning. Fältet har således utvecklats och utvecklas kontinuerligt i förhållande till vad som definierar kvalitet i fråga. Något som visar på tecken för ökat intresse och akademisk aktivitet är att antalet tidskriftsartiklar som berör översättnings – och anpassningsfrågor har ökat med åtminstone 350% under de senaste två decennierna (sökning som genomförts i databasen PsycINFO, hänvisning i Byrne, 2016).

I de senast publicerade standarderna – mest framträdande i det ramverk som ingår i uppsatsen från American Educational Research Association et al. (2014) och ITC’s riktlinjer (2017) – lyfts dels behovet fram att väga in ökad hantering av webbaserade tester. Det mest framträdande är dock en successiv rörelse mot att definiera komponenter för och väga in universell design i procedurerna men även att inkludera och betona både förfinad och mer avancerad psykometrisk testning (som avslutande steg), för att uppnå funktionell ekvivalens. Övriga riktlinjer har närmat sig Standards (i hänseende av statistiska granskningars framträdande roll) som traditionellt tydligt betonat psykometrisk granskning för att säkerställa testers tillförlitlighet. Standards har å andra sidan kommit att i högre grad väga in flexibilitet och universell design i sina testutvecklande resonemang. Universell design är som sagt ett nytt begrepp i Standards i och med den senaste revisionen. Resonemangen kan sammanfattas i att

---

<sup>5</sup> Se till exempel Åsberg (2001) för ett resonemang om ontologi, epistemologi och metodologi med kritisk genomgång av grundläggande vetenskapsteoretiska begrepp och ansatser, där bland annat förgivet tagna antaganden om psykometrins ”rena objektivitet” problematiseras. Se även Porter (1996) kapitel 8, s. 193 – 216.

ekvivalens är överordnat och att universell design och statistisk granskning (psykometri) är metodologiska angreppssätt för att uppnå kulturell jämförbarhet.

Trots vissa skillnader kan det antas att idag handlar val av riktlinjer mest om preferenser eller logistik. De är övervägande lika i synsätt och innehållsinriktning. De flesta berör eller betonar interdisciplinärt teamarbete och innefattar tillbakaöversättning (även om det inte alltid ses som obligatorisk längre), kommittéer eller fokusgrupper bestående av experter enbart eller i kombination med lekmän, och väl utvecklade förslag på statistisk/psykometrisk testning. Det som skiljer ut är att Standards utgör ett övergripande ramverk för en samlad kunskapssyn och de övriga riktlinjerna ger en mer praktisk, och ibland detaljerad, inriktning (som dock bygger på samma epistemologiska bas). Man kan se det som att samtliga utgör varianter på samma strävan eller kamp för att etablera ekvivalens.

Min bedömning är att ITC:s senaste utgåva (2017) är den mest sammansatta riktlinjesammanställningen som explicit avser vägledning i kvalitetssäkrande testadaptionsprocesser för både standardiserade och icke standardiserade (psykologiska) test. ITC har inkluderat innehåll som ingår i Beaton et al. (2000) och Kristjansson et al. (2003), till exempel att säkerställa rättigheter för att genomföra anpassning av ett test och initialt undersöka kulturell lämplighet. Författare till Beaton et al:s riktlinjer har dessutom ingått i en ITC-kommitté för riktlinjeproduktionsprocess (åtminstone inför första utgåvan) som visar på internt kollegialt samröre. ITC refererar även återkommande till Standards. På så vis kan man se ITC:s riktlinjer som ett uttryck för fältets samlade akademiska kunskap och huvudsakliga frågor som diskuteras i testadaptionslitteraturen. Utifrån ögla 2 i Latours modell framstår det som att det pågått ett aktivt inom-kollegialt utbyte och arbete, även om kritik mot det egna fältet tycks kvarstå (så som från Epstein et al., 2015 och Byrne, 2016).

Tillsammans med Standards ges ITC:s riktlinjer ut och publiceras även utanför akademiska databaser och forskarsamhället. Dessa båda framstår som de mest spridda (citerade) standarderna och rekommenderas utslutande av branschorganisationer som värnar testkvalitet (i alla fall de organisationer jag varit i kontakt med). Detta har troligen en av förklaringarna i att Beaton et al. och Kristjansson et al. publicerats i medicinskt inriktade tidskrifter.

Nästa avsnitt som berör bristande kontakt med användarna av riktlinjer ger en stegvis övergång från ovan fokus på ögla nummer 1 och 2 i Latours modell till ögla nummer 3 (allierade i form av testutvecklare med flera).

## 6.6. Bristande kontakt med användarna

ITC skriver själva i slutordet för sina riktlinjer, att det som samtliga ramverks- och riktlinjeförfattare beskriver som ett huvudproblem tycks kvarstå, nämligen efterlevnad (International Test Commission, 2017, s. 30). De hänvisar till en systematisk översikt som visar att majoriteten av de testadaptionsprojekt som genomförs inte sker i linje med ITC:s riktlinjer (Rios och Sireci, 2014) och som jag förstår ej heller andra akademiskt publicerade riktlinjer. De uppmanar därför till att som läsare av deras dokument sprida information om riktlinjerna.

Detta (spridningen) avser hur första, andra och femte ögla är förbunden med tredje ögla i Latours modell: att nå ut till och skapa allierade. I detta fall tänker jag att allierade utgörs av åtminstone testutvecklare, testförmedlare och testanvändare. De (ITC men även andra riktlinjeförfattare) har uppenbarligen inte lyckats skapa intresse nog utanför sin egen krets eller så är kanske inte riktlinjerna tillräckligt hänsynstagande till praktisk verklighet. Eller så har olika intressenter andra preferenser än riktlinjeförfattarna, av olika skäl. Finns det några belägg för dessa antaganden?

Det framstår som att det finns inflytelserika och högst engagerade personer, som både tar fram riktlinjer och samtidigt ingår i intresseorganisationers inre sfär. Intresseorganisationer är de huvudsakliga kanalerna för att upplysa, sprida kunskap och informera om riktlinjeinriktad praxis (mig veterligen). Ändå lyckas riktlinjeförfattarna inte nå ut till allierade (i ögla nummer 3) i den

utsträckning som de önskar. I alla fall inte att få effekt gällande efterlevnad av standarder och riktlinjer. Det framstår som att det är en liten klick av personer som återkommer på olika roller inom testutvecklings- och testanpassningsbranschen. Detta förmedlas genom intervju med Sofia Sjöberg, leg. psykolog på det svenska testutvecklings- och testförmedlande företaget Assessio. Sjöberg beskriver att det är ett fåtal personer som engagerar sig i riktlinje- och testutvecklingsfrågor och att dessa samma personer ofta återfinns i flera olika bransch- och intresseorganisationer. I den cirkulation av referenser jag mött på under mitt uppsatsarbete så bekräftas den bilden i stort. Samma person kan således vara riktlinjeförfattare, testutvecklare och även arbeta för branschorganisationer.

Till viss del är det samma personer som både utvecklar riktlinjer och anpassar tester. Varför anpassas inte riktlinjerna efter användarna bättre? Är det för att ögla 5 och ögla 2 har sin egen dynamik och inte ”klarar av” att anpassa sig efter logik och behov i ögla 3? Jag kan inte konkret uttala att det är så men man bör kunna sluta sig till att riktlinjeförfattarna (i de fall de också är testutvecklare) inte *kan* gå ifrån sin vetenskapliga logik och det rådande fokuset på ekvivalens. I så fall tyder det på att öglornas logik (nummer 2 och 5) är otroligt starka. Med dessa resonemang kvarstår således frågan om varför riktlinjerna inte bättre når ut till, eller får effekt rent praktiskt för testutvecklare och testanvändare. Branschorganisationer (ögla nummer 2) utgör ju en väsentlig kanal mellan riktlinjeproduktion och åtminstone vissa grupper av testutvecklare och testanvändare. Om det blir någon skillnad i och med ITC:s senaste utgåva (2017) med att få effekt för praktisk tillämpning kvarstår förstås att se. Man kan fråga sig hur testadaptionens vetenskaplighet och dess representanters syn på giltig kunskap (i Latours ögla nummer 2) gör sig gällande för olika intressenters behov (allierade i ögla nummer 3). Dessa intressenters situationer och preferenser kan behöva synliggöras för att bättre förstå varför riktlinjer inte praktiseras. Vi rör oss därför över till att belysa några olika intressenter inom testadaptionens system och kommentera de kunskapsperspektiv och den verklighet de företräder och har att förhålla sig till.



## 7. Testutvecklare: riktlinjernas irrelevans

Så långt riktlinjeutvecklarnas kamp för ekvivalens i ögla nummer 2 och 5. Nu rör vi oss över till ögla nummer 3. Jag har under uppsatsarbetets gång kontaktat två informanter med gedigen erfarenhet av testanpassning. Ett sammandrag av dessa samtal följer nedan. Fokus är vad som är viktigt för användare av riktlinjer, alltså testutvecklare men indirekt även andra aktörer i testadaptationens system.

### 7.1. Testernas marknad

Om vi ser till testanvändares perspektiv – de som till exempel certifierar sig i, köper eller upphandlar tester – så beskriver informanten Sofia Sjöberg på Assessio, som representant för ett privat testutvecklande och testförmedlande företag, en situation som kännetecknas av att kunder idag efterfrågar allt snabbare testproduktion, detta ibland så snabbt som ett par till tre månader. Hon ser det som en konsekvens av digitaliseringens frammarsch. Det medför en förväntan på snabb produktion trots att ett väl anpassat test i vissa fall (kliniska eller arbetspsykologiska tester) kan ta två år eller mer att åstadkomma, enligt konstens alla regler. Sjöberg anger att Standards och ITC:s riktlinjer är de kvalitetsnormer som Assessio eftersträvar att följa men nämner även att dessa riktlinjer utgör en slags utopisk ”golden standard” i rådande läge. Assessio arbetar därför med iterativa processer i sin testanpassning och kommunicerar i detta förfaringssätt löpande till sina kunder.

Vidare beskriver Sjöberg en bild av kunder, som till exempel kan utgöras av HR-konsulter och policyskapare, som vanligtvis väljer (anpassade) tester som de antingen redan känner sig bekväma med utifrån kompetens och erfarenhet och/eller tester som de tidigare investerat certifieringsavgifter för, något som framstår som rimligt (trots eventuell bristande evidens för ekvivalens). Det kan även gälla organisationer som påbörjat och är trygga att göra analyser med ett visst test och ser det som problematiskt att byta till något som är mer ”up to date” och som är mer validerat. Få kunder har någon djupare kunskap om testadaptation som sådan och ser därför till andra, ofta mer pragmatiska, aspekter när de gör sina val av vilka tester de använder och satsar på menar Sjöberg.

Det framkommer även att flertalet kliniska tester generellt skulle ha kunnat utvecklas och förfinas om marknaden varit öppen för det men det görs en bedömning att det är för stora ekonomiska risker att göra genomlysning av och ändra testen. Sjöberg berättar att vissa av dessa produkter ej har utvecklats på femtio år. Med detta förhållningssätt så får vi inte någon utveckling och inte heller tester som är anpassade till dagens kunskapsläge säger Sjöberg. Att få ihop akademien (stringens) och business (pragmatism) är en ständig avvägning menar hon. Det vi blir varse här är att det finns etablerade testutvecklare som har grundläggande avsikt att genomföra testadaptation med stöd av angivna standarder men där detta ställs mot vad marknaden efterfrågar och kunders behov och preferenser. Det bör dock tilläggas att detta inte gäller samtliga kunder, då det även finns kunder som är måna om stringent anpassade test och därmed ”fogar” sig i den tidshorisont som en sådan process kräver.

Berthelsen – den andra informanten i denna uppsats – som lett anpassningsarbete av svensk version av Copsoq och som även arbetar aktivt internationellt med detta instrument (notera, jag väljer att ange Copsoq som *instrument* istället för *test* då denna benämning bättre matchar denna typ av formulär). Copsoq är ett instrument som ej har ett kommersiellt syfte och det finns fritt att nyttjas och ladda hem via internet på olika språk.

Berthelsen berättar att forskarna som var involverade i anpassningsarbetet av svenska Copsoq betonade att ta fram ett instrument som skulle vara användbart i praktiken (applicerbart på arbetsplatser). De avsåg i arbetet att följa riktlinjer utifrån bästa möjliga mån. Berthelsen menar dock att ökat fokus på stringens i anpassningsarbetet får påtagliga konsekvenser. Stringens betyder i sammanhanget mer rigorösa procedurer och avancerade statistiska, psykometriska testningar. Alla anpassnings- och valideringsprocesser kan inte se likadana ut. Informanten menar att stringens kräver resurser – personellt; kompetensmässigt; ekonomiskt och; tidsmässigt. Med krav på en allt mer

strukturerad och strikt anpassningsprocess följer det även med en del ideologiska och etiska följdfrågor, så som hon ser det.

Det är inte alltid enkelt att säkra resurser till valideringsstudier och det betyder bland annat att hittills genomförda nationella översättningar/anpassningar av Copsoq ligger i olika stadier av valideringsprocessen. Om man definierar en starkt standardiserad procedur som något som måste följas i varje liten detalj kan det innebära att man i praktiken riskerar utesluta länder från det internationella Copsoq-samarbetet säger Berthelsen. Istället arbetar det internationella nätverket med inkluderande och samarbeten mellan länder, för att på ett mer *pragmatiskt* vis styrka att etableringen av nationella versioner håller kvaliteten. I vissa länder används fortfarande den första versionen av Copsoq-instrumentet, ett instrument som motsvarar vår svenska arbetsmiljö för cirka femton år sedan. Frågan är hur strikta vi ska vara med vilka versioner som ska och får användas säger Berthelsen. Hon framhåller att för att jämföra data mellan länder behöver skalor och items konceptuellt och psykometrisk överensstämma. Samtidigt ser hon att lokala variationer som fungerar och är relevanta behövs. Om man framtvingar att använda senaste versionen tappar copsoq-nätverket och dess forskning vissa länder helt och hållet, främst de med sämre förutsättningar att vaska fram resurser säger Berthelsen. Detta innebär för vissa länder risker att uteslutas ur kollegiala nätverk som i sig skulle kunna fungera som auktoritets- och påtryckningsmedel för viktiga samhällsförändringar (författarens notering: i detta fall handlar det om organisatoriska och sociala arbetsmiljöfrågor). Om man omvänt inte uppdaterar instrumentet i till exempel de skandinaviska länderna så det anpassas till utvecklingen i samhället, då förlorar instrumentet sin relevans i dessa länder. Med stora internationella samarbeten är det en balansgång och efterhand som ett instrument sprids behövs även mer formalisering av minimikrav menar Berthelsen.

Berthelsen reflekterade kring att det förstas är viktigt att mäta rätt saker och att mäta rätt saker med en viss precision men: ”Ska vi endast satsa på ”metaanalyserbara” versioner av Copsoq eller tänka mer praktiskt?” Informanten ser hellre att nationella versioner av instrumentet får finnas kvar, versioner som uppenbart fungerar men där översättning inte sker eller skett standardiserat. Hon menar att utifrån ett samhällsuppdrag bör man fråga vad som behövs totalt sett. Berthelsen anser att hänsyn behöver tas till de behov som uppstår i olika sammanhang och utifrån olika syften (författarens notering: alltså kan inte endast riktlinjer för testanpassning vara det som vägleder eller avgör hur anpassningsarbetet ska gå till).

Efter denna genomgång av ett par testutvecklarens perspektiv följer relaterade reflektioner på resursfrågor som berör riktlinjernas utformning i relation till aktörers (testutvecklare och testanvändare) verklighet.

## 7.2. Resurskrav

Det är uppenbart att riktlinjeförfattarna ser ytterst seriöst på översättning och testanpassning som sådan. Samtidigt tycks åtminstone två övergripande, realistiska dilemman framträda (även uppmärksammat av Byrne, 2016, dock främst som ett konstaterande), där det första dilemma skulle kunna ha sina rötter i vad jag benämner som en alltför hård disciplinering. Detta första dilemma gäller svårigheten att förvärva välutbildade och kunniga experter i lingvistik, psykometri och relevanta fält inom psykologi, för att ingå i testutvecklande team under de former av adaptionsprocesser som anges i riktlinjerna. Med hård disciplinering menas att det ställs (för) höga krav på disciplinär kunskap och expertkunskap. Inom lärarkåren har en liknande process pågått, där man har försökt rensa upp bland icke-examinerade lärare genom legitimation. Detta är en beprövad väg där nog läkare och psykologer har gått först. Läkarna har en tydlig hierarki och psykologerna har en likande utveckling också, även om den är mindre uppstyrd. Detta skulle kunna förväxlas med mekaniska krav men detta (det mekaniska) handlar dock alltid om metoder eller procedurer. Disciplinära krav handlar om legitimation, hierarki, eminens (i motsats till evidens), experter, examina etcetera. Den hårda disciplineringen består i detta fall i att presentera en praktisk vägledande forskningsprocess som nära

nog är omöjlig att leva upp till som testanpassande praktiker. Denna process framstår således inte som särskilt verklighetsförankrad. Enligt min bedömning – men även adresserat av ITC (2017) och Byrne (2016) – framstår det som att dessa akademiker (som föreslås ingå i testutvecklande team) utgörs av en mindre klick experter från en rad olika discipliner. Hur möjliggör man rekrytering av dessa experter? Testadaption tenderar eventuellt att bli en elitistisk subdisciplin, som vissa kan dra fördel av förstås.

Det andra dilemmat, som också relaterar till rekrytering, är de uppenbara kostnaderna som det innebär att värva sådana experter (om de finns tillgängliga vill säga) som kan genomföra anpassningsförfaranden enligt givna standarder – detta såvida man inte har dem som kollegor. En möjlig politisk koppling är därför denna: det framstår som att främst välfinansierade testförmedlingsföretag eller framstående forskare som lyckas ansöka om finansiella medel mäktar med en riktlinjebaserad adaptationsprocess (om kunderna/testanvändarna tillåter den tid som krävs). I strävan att strama upp och i riktlinjerna allt mer precisera rigorösa modeller skärps fältet, med påföljande uteslutning av vissa oseriösa aktörer men sannolikt även seriösa sådana som är sämre bemedlade (finansiellt, kontakter och inflytande). Riktlinjeförfattarna är som sagt inte ovetandes om dessa resursfrågor men frågorna problematiseras ej i den utsträckning som skulle kunna behövas, i alla fall inte enligt informanternas utsagor.

Riktlinjeförfattare, testutvecklare och branschorganisationer utgörs dessutom som nämnts ovan ej ovanligt av samma människor. Vilkas intressen företräder de i de olika rollerna? Plake och Wise (2014) rapporterar att inför den senaste utgåvan av Standards har man backat på kravet att ej inkludera experter som kan ha kommersiellt intresse av hur standarder utformas. Dessa experter förväntas ha testkvalitet som sådan som det främsta att eftersträva i framarbetandet av standarder. Kruxet är att utesluta sådana personer är nära nog lika med att utesluta de främsta forskarna, eftersom de vanligtvis är engagerade så väl inom akademien som inom egna företag och näringsliv. En liknande situation återfinns till exempel inom European Federation of Psychologists' Association (EFPA). Det behöver förstås inte betyda något problematiskt i sig men likväl viktigt att denna box packas upp och synliggörs. En sådan genomlysning är dock utanför omfattningen på denna studie.

### 7.3. Sammanfattande diskussion

Med utgångspunkt i reflektioner kring resurskrav och de utsagor som informanterna Sjöberg och Berthelsen lägger fram ovan så kan vi notera att som testutvecklare har man inte endast riktlinjer att förhålla sig till utan också till exempel vad som efterfrågas på marknaden, kunders behov och preferenser, samt återigen resursfrågor. Ingen av dessa informanter ifrågasätter behovet av riktlinjer, däremot har de en klämd position mellan stringens och pragmatism. Med detta menas att de å ena sidan har formaliserade riktlinjer och standardiserade procedurer att följa, som i sig anger en uppsättning resurskrav, och å andra sidan en verklighet som inte alltid matchar det som dessa formaliserade riktlinjer anger eller kräver. De har även angränsande etiska frågor att hantera (så som vad är en miniminivå i anpassningsprocessen, för att med gott samvete släppa ett test på marknaden?).

Därtill, finns en ytterst politisk aspekt i det Berthelsen problematiserar kring skarpare krav på stringens (att följa standarder) i förhållande till möjlighet till allokering av resurser. Om detta dras till sin spets riskerar vissa länder i det nätverk hon representerar att falla ur ett givande internationellt samarbete. Däremot om ett mer flexibelt förhållningssätt till standardisering – till exempel acceptans för att olika versioner av instrumentet används, även om det hindrar att ingå i metaanalytiska studier för vissa parter – får råda finns möjlighet att bidra till att samhällsutvecklingsfrågor fortsätter att föras upp på agendan. Jag tycker att resursfrågan framstår som politisk till sin natur.

Utifrån ovan reflektioner ser det ut som att riktlinjeförfattarna – utan att nödvändigtvis ge avkall på sin grundläggande strävan efter rättvisa för testrespondenterna och (funktionell) ekvivalens – behöver förstå fler aktörers behov och situation, i alla fall om de är mån om att riktlinjer praktiseras i högre utsträckning. Det framstår som att ett förfiningsarbete pågår men inom samma (snäva) idésfär.

Riktlinjerna behöver därför sannolikt visa på en större grad av kontextuell flexibilitet för att matcha praktiska villkor.

Vi ska nu övergå till att, med stöd i Porters (1996) objektivitetskategorier och de kritiska perspektiven på standardisering som lyftes fram ovan via Timmermans och Berg (2003), göra en analys av valda teman och grundelement som ingår i testadaptionstödande ramverk och riktlinjer. Vi rör oss därför över till frågeställning nummer 3 – Vilka eventuella spänningar finns gällande synsätt på giltig kunskap och vetenskaplighet, inom och sinsemellan ramverk och riktlinjer? – och frågeställning nummer 4 - Finns det kunskapsperspektiv som riktlinjeförfattarna skulle behöva beakta och inkludera för att främja ökad tillämpbarhet av standarderna och riktlinjerna? Vilka i så fall?

## 8. Analys av ramverk och riktlinjer – Objektivitetssyn och standardisering

### 8.1. Introduktion

Här inleds en analys av ramverk och riktlinjer med inriktning på frågeställningarna nummer 3 och 4 - som berör spänningar gällande synsätt på giltig kunskap och vetenskaplighet, och om det finns kunskapsperspektiv som riktlinjeförfattarna behöver ta hänsyn till för att främja tillämpning av riktlinjer. Analysen riktar ett särskilt fokus på objektivitetssyn, standarder/standardisering och hur detta tenderar att samspela med varandra. Med stöd i dels ANT och Latours modell (se s. 13 ovan) som har ett systemperspektiv på vetenskaplig produktion och dels Porters (1996) objektivitetsbegrepp, tillsammans med Timmermans och Bergs (2003) utvärderande av standarder tar vi oss an att problematisera ramverk och riktlinjer inom testadaptation. I denna introduktion pekas några gemensamma drag ut och därpå behandlas i tur och ordning några centrala teman i mer detalj.

Gemensamt kommunicerar riktlinjeförfattarna nödvändigheten av standarder och riktlinjer för testanpassning, de lyfter genomgående fram de fördelar som de ser med standarder. De verkar, vad jag kan se, ha en genuin omtanke om testmottagarna och strävar efter bättre tvärkulturell forskning. Som presenterats ovan så vittnar flera författare – både riktlinjeförfattare och de som kritiserar nuvarande utformning av riktlinjer - om problematiken med undermåliga anpassningsprocesser med påföljande negativt kännbara konsekvenser. En diskussion om vilka (nya) problem som standarder kan skapa uteblir dock samtidigt. Detta kan delvis förstås komma sig av det som Porter synliggör; unga (under)discipliners behov av att komma bort från spretighet i förfaringssätt, en möjlighet att positionera sig samt att bygga förtroende för det egna kunskapsfältet via standarder (I Latours modell att stärka sin disciplin inom ögla nummer 2 men även att framstå som attraktiv gentemot allierade i ögla nummer 3). Det framstår som att det som förklaras som en rörelse mot en mer exakt vetenskap har en social och politisk dimension.

Standardiseringsrörelsen och dess bakomliggande idéperspektiv som granskats av Porter (1996) men även Timmermans och Berg (2003) innefattar vissa kunskapsideal, som gynnar viss objektivitet medan andra former får stå tillbaka (då de ej anses lika tillförlitliga i sammanhanget). Vid en första anblick kan det te sig som att riktlinjeproducenterna står för en rörelse bort från mekanisk objektivitet, som man kan säga representerats av den tidigare rådande logiken att endast tillbaka-översättning behövs för att säkra ekvivalens mellan kulturer. I stället har man ökat betoningen på kulturell sensitivitet och mer intrikata, expertbaserade anpassningsprocedurer. Samtidigt består paradoxalt nog både tron på allt mer strikt formaliserade standarder och tron på att det går att uppnå ekvivalens mellan kulturer: Med rätt riktlinjebaserade metoder går det att ta fram (standardiserade) tester som är likvärdiga över flera kulturer.

Riktlinjeförfattarna är dock inte okritiska till sina egna idéer, då det i litteraturen diskuteras till exempel att *alla tester* inte är anpassningsbara. I Standards (2014, s. 58) anges att principerna för universell design – som syftar till ekvivalens mellan testpopulationer – förvisso ger värdefull guidning i anpassningsprocesser, men utan evidensbaserade belägg för de principer som ingår. I övrig text genom handboken, där universell design nämns, behandlas dock denna design som något som bör eftersträvas och därmed att ekvivalens kan uppnås. Med andra ord, skulle någon undgå att i Standards läsa fåtalet rader där förbehållet om bristen på evidens nämns så framkommer inte denna information.

Så som riktlinjernas procedurer är utformade ges i vissa delar en stor tillit till professionellt bedömande och överväganden, som görs experter emellan. Som belysts ovan är de dock uppstyrda med diverse krav som är ytterst resurskrävande gällande till exempel olika expertkompetenser och tid som behöver tas i anspråk. Riktlinjerna öppnar ändå upp för expertmedverkan, eller disciplinär objektivitet, i motsats till huvudsaklig riktning inom standardiseringsrörelsen (kokboks-principer) och

de resonemang som förs av både Timmermans och Berg (2003) och Porter (1996). På så vis går vissa delar i utformningen av riktlinjerna på sätt och vis emot den generella evidensrörelsen, åtminstone som den inledningsvis uttrycktes inom evidensbaserad medicin. Genom att i större utsträckning anamma behovet av subjektiv expertkunskap framstår det som att de företräder en vidare kunskapssyn än den som ofta ifrågasätts inom den medicinska evidensrörelsen och dess riktlinjeproducering (Timmermans och Berg, 2003, s. 19 – 20). Trots standardiseringsivern och den odelat positiva synen på standarder så driver inte alla på en renodlat procedurell objektivitet. Möjligtvis kan man i dessa öppningar för expertomdöme se en parallell till de anpassningar som har gjorts inom evidensrörelsen efter den inledande kritiken mot densamma. Efter den initiala lanseringen av begreppet evidensbaserad medicin har vissa centrala företrädare framhållit att evidens alltid måste samordnas med kliniskt omdöme (Howick 2011). Det disciplinära verkar dock bli allt mer uppstyrt så som riktlinjerna utformas.

Min slutsats är att det råder en preferens mot mekanisk objektivitet som även åtföljts av en rörelse mot hårdare disciplinering (vilket också framkom i avsnitt 7.2). I detta läge verkar det som att riktlinjeförfattarna söker att strama upp det disciplinära, kanske för att riktlinjerna ska framstå som mer rigoröst vetenskapliga, för att skapa förtroende både internt men också externt.

Det finns kritiker som vill se en mer samlad riktlinjeproduktion, där det framkommer att Epstein et al. (2015, varav en författare själv är riktlinjeförfattare) efterfrågar noggranna utvärderingar av de riktlinjemodeller som publicerats. Dessa författare ifrågasätter utformning, omfattning och bristen på utvärdering. Med en snabb blick skulle man kunna tänka att de analyserar riktlinjeproduktion med en mer symmetrisk vetenskapsteoretisk ansats. Jag vill hävda att de analyserar och granskar med ett inifrånperspektiv. Detta perspektiv ger ingen kritisk blick på kunskapsperspektiv (så som den mekaniska objektiviteten) eller på standarder som sådana. Tvärtom tycks en procedurell objektivitet vara eftersträvansvärd i efterfrågan på en tydligare gyllene standard – helt utan problematisering. Troligen återspeglar detta den tillitsstruktur som Porter (1996) men också Timmermans och Berg (2003) beskriver.

Efter denna inledande analys följer ett avsnitt som berör den centrala idén om kulturell ekvivalens inom fältet för teststandarder.

## 8.2. Epistemologisk förförståelse: (o)giltigheten i kulturell ekvivalens

Det framstår som att samtliga ramverks- och riktlinjeförfattare övergripande representerar en gemensam epistemologisk förförståelse och förenas i synsätt kring kunskapens möjligheter, ursprung och giltighet. Låt oss belysa synsätt på kultur och ekvivalens.

För det första, det finns gemensamma kulturella drag mellan olika kulturer (som är översättningsbara) och för det andra, det finns även icke-gemensam kultur. En slutsats att dra av detta är att allt är därför inte översättbart mellan kulturer. För det tredje – där det låter sig göras och stämmer överens med koncept som avses mätas – så är det eftersträvansvärt att ”vaska fram” och försöka få tag på det gemensamma i dessa kulturer. Eftersom tester bör spegla rättvisa (i fråga om likvärdiga möjligheter att demonstrera ”var man befinner sig” i förhållande till det som mäts) för samtliga testmottagare, så riktas ambition mot att skapa funktionellt ekvivalenta tester. Detta utgör tester som på respektive målspråk fångar kulturunika sätt att operationalisera, uttrycka och mäta de gemensamma konceptuella konstrukten (så som till exempel depression, lycka, arbetsrelaterad stress, olika personlighetsdrag) via reliabilitets- och validitetstestade items och skalor som ingår i frågeformulären. Detta låter sig förverkligas med rätt metodologier som utgörs av en kombination av subjektiv (uppstramad) expertkompetens och (opersonliga) statistiska granskningar. På så vis hävdas det möjliga i att ta fram tillförlitliga tester som möjliggör tvärkulturella jämförelser. För det fjärde ska icke-gemensam kultur ”tvättas bort” eller exkluderas ur tester som ska användas i tvärkulturella

sammanhang, då det riskerar bidra med bias och icke-rättvisande resultat för (vissa grupper av) testmottagarna, detta enligt riktlinjeförfattarna. De icke gemensamma kulturella dragen behöver således exkluderas i strävan mot konceptuell ekvivalens.

För det femte tycks det som att själva inkluderandet av psykometrisk logik med tillhörande statistiska procedurer i testadaptionen utgör en för givet antagen modell för slutlig kvalitetssäkring – detta utan någon problematisering eller granskning av vad de psykometriska kriterierna egentligen grundar sig på, i alla fall så vitt jag kunnat läsa mig till. Kanske sker dessa resonemang i fördjupande litteratur som återfinns i hänvisningar (dock utanför det jag läst mig till), eller så är det bara så att frågor som ”hur kom psykometrin att få en central plats inom psykologisk och pedagogisk forskning” och ”vad kännetecknar synsätt på objektivitet och kunskap för dessa statistiska bedömningsgrunder” helt rör sig utanför den sfär av frågeställningar som litteratur inom testadaption består av (se problematiserande resonemang i Porter, 1996, till exempel sidorna 199-200, och 209-213). Vi återkommer till statistisk granskning genom psykometri i ett eget avsnitt nedan.

Kanhända utgör ett ifrågasättande om i vilken mån (funktionell) ekvivalens mellan kulturer överhuvudtaget är möjligt – och i förlängningen hur vägledande riktlinjer bidrar till detta – en så kallad epistemologisk svart box. Även om riktlinjeförfattare problematiserar svårigheter med sådant som konceptuell och semantisk ekvivalens så präglas ändå kommunikationen av att möjligheten till funktionell ekvivalens (Herdman et al., 1998) är uppnåbart. (Adresserandet av denna spänning sker via förslag på kulturella förstudier men den grundläggande frågan består).

Utifrån det tredje och fjärde antagandet som lyfts fram ovan i detta avsnitt (som berör ekvivalens): översättningar innefattar exkluderingar och reduceringar av olika slag, i allra första skedet när ett test skapas i originalversion, från (psykologisk) värld till ord (operationalisering av begrepp/konstrukt), och därefter till nya ord anpassade för en ”annan värld” (kontextuell och kulturell anpassning av test) än originalversionens. Det är så att säga mycket som händer i den transformationen, i översättningens olika ingående steg. I riktlinjeförfattarnas resonemang framträder en tonvikt mot mekanisk syn på *livsvärlden* även om det betonas kulturell sensitivitet och hänsyn till disciplinär objektivitet i själva testadaptionsarbetet. Det är möjligt att operationalisering av funktionell ekvivalens tenderar ge sken av överförbarhet och ordning och reda i en utsträckning som kanske inte helt är möjlig i en ”rörig” värld (eng. *a messy world*), för att använda Latours begrepp (2005, s. 136). Det betyder inte per automatik att tvärkulturell forskning inte skulle gå att bedriva men att mer ingående verka för att osäkerhetsfaktorer accepteras och kommuniceras mer explicit kollegor emellan samt även gentemot testanvändare och testmottagare – utan att för den skull avsäga sig att vetenskapligt arbete pågår – framstår dock inte i nuläget som lika ”säljbart”. En vetenskap som bygger på en retorik som kännetecknas av universell giltighet, med dess tydlighet och säkerhet, inger en större trovärdighet och blir på det sättet ett starkare retoriskt verktyg att använda vid olika beslut, internt men också ut mot samhället (Porter, 1996, s. 230).

Inom ögla nummer 2 i Latours modell kan sägas att vad som räknas som vetenskap är i stor utsträckning vad forskare bestämmer att vetenskap ska vara. Strukturerna för hur denna vetenskap skapas består av normer, idéer, värderingar och metoder för vad som är ren vetenskaplig praktik. Det utgörs av kognitiva och sociala normer som behöver följas i produktionen, legitimeringen och spridningen av kunskap för att den ska klassas som vetenskaplig. Kvalitetskontroll för vetenskaplig kunskap sker således utifrån vad den vetenskapliga disciplinen anser vara god vetenskap följt av de normer som styr inom det specifika forskningsfältet (jfr. Gibbons et al., 1994). Att synliggöra de osäkerheter som finns i forskningspraktiken kan innebära en otydlighet och rörighet som inte får samma genomslagskraft inom forskargemenskapen (ögla nummer 2) men också i förhållande till allierade (ögla nummer 3) och samhället i stort (ögla nummer 4).

Ekvivalens är en ofrånkomlig del av det epistemologiska grundantagandet bakom testadaption som sådant. Det framstår som rimligt att om man tror på testöversättning är det klart att man någonstans har ekvivalens som ett mål. Om aktörer inom testadaption inte har främsta fokus på ekvivalens, så kan

man fråga sig vad som istället står i fokus? För testutvecklare i ögla nummer 3 (Latour) verkar snarare praktisk användbarhet vara det primära och inte ekvivalens, även om det också förstås är av signifikant intresse för vissa aktörer. Vissa av dessa aktörer (testutvecklare) tänker först och främst mer pragmatiskt på ekonomi och tid, inte nödvändigtvis ekvivalens. Spänningsfältet mellan riktlinjeförfattarna och delar av det skrå som utgörs av testutvecklare visar sig i detta. Testadaption kanske inte alltid handlar om att uppnå ”likhet mellan kulturer” utan att skapa något som är *användbart* i en annan situation än den ursprungliga. Det som söks är att använda något som fungerar ”där” men utan föreställning om en ”ren” och fullständig överförbarhet. Dessa resonemang går i linje med Cartwright och Hardie (2012) som i boken *Evidence-based policy* inriktar sig på att stödja policyproducenter hur de ska utgå ifrån och använda evidens på effektiva sätt. Snarare än direkt överföring av en hel intervention tänker sig Cartwright och Hardie att huvudpoängen i en intervention ska tas med men utan hänvisning till ”likhet” eller ekvivalens. Utifrån ett synsätt inspirerat av ANT finns det liknande förslag. Karin Fernler ger i *Perspektiv på implementering* (2012, s. 26 – 28) några bilder av vad implementering utifrån ANT skulle kunna innebära. Översättningar är ju alltid partiella och förändrande enligt ANT, det vill säga bokstavig och exakt översättning är inte möjlig. Enligt Fernler är det viktigt att inte försöka överföra ett sammanhållet paket utan genom personlig coaching, goda exempel och enklare verktyg sträva efter att i ett nytt sammanhang dra nytta av det som gjorts i ett annat. Viktigt blir att inte förlita sig på en engångslösning utan en kontinuerlig process med återkommande besök av och dialog med personer med kunskap från ett redan genomfört arbete. Inom testadaption skulle detta innebära nätverksaktiga samarbeten såsom Copsoq beskrevs ovan av Berthelsen. Copsoq:s diskussioner om pragmatiska anpassningar av krav på testadaption (exempelvis för olika länder beroende på ekonomiska förutsättningar) påminner om både Cartwright och Hardie och ANT:s översättningsbegrepp i en nedprioritering av strikt ekvivalens.

Vad gäller kollegial kritik, kritiserar Epstein et al. (2015) fältet och menar att riktlinjemodellerna överlag både spretar och saknar utvärdering. De är kritiska mot att riktlinjerna endast vilar på ett antagande om att vara funktionella (för att uppnå ekvivalens). Även Byrne (2016) sällar sig till liknande resonemang och föreslår mer rigorösa psykometriska modeller. Dessa olika författare är inte tillfreds med kvaliteten i utformning av riktlinjer. I denna kritik tycker jag att man kan spåra spänningen mellan mekanisk och disciplinär logik med ett tydligt driv mot det mekaniska. Epstein et al. och Byrne vill se en utveckling mot mer likriktade standarder, där en ”gyllemodell” eftersträvas. Detta framstår som en motsatt logik och riktning till det som Fernler lägger fram.

Utvärderingar är säkerligen nödvändiga i strävan att utveckla testadptionsfältet och dess riktlinjer. Det kvarstår ändå en fråga om hur de än mer enhetliga procedurer som Epstein et al. och Byrne förespråkar – men som jag tolkar det också är drivkrafter och riktning bakom riktlinjeförfattarnas arbete – ska kunna kombineras med de intrikata kulturella jämförelser som behövs. Den springande frågan synes vara i vilken utsträckning sådana jämförelser kan mekaniseras. I vilken utsträckning kan kulturell kontext fångas in av en mekanisk objektivitet? Jag frågar mig hur mycket disciplinär objektivitet kan omvandlas till (mekaniska) procedurer utan att kulturellt känsliga element reduceras bort för mycket. Möjligheten bör sannolikt avgöras av hur de professionella utifrån lokala förutsättningar, verktyg och metoder *anpassat* kan implementera det som standarderna fordrar (jmf. Timmermans och Berg, 2003, s. 202 och s. 211). Underförstått finns det i så fall någon slags gräns för procedurerna och vad en mekanisk objektivitet kan leverera.

Efter att ha analytiskt belyst synen på kulturell ekvivalens inom det akademiska testadptionsfältet övergår uppsatsens analys till några relaterade spänningar. Centrala begrepp, eller processkomponenter, som delvis redan berörts ovan undersöks: universell design och rättvisa (eng. *fairness*; dokumentation; och slutligen psykometrisk utvärdering. Dessa komponenter är, som vi sett, nödvändigtvis inte (lika) prioriterade för andra aktörer (så som allierade i ögla nummer 3 i Latours modell) som för riktlinjeförfattarna.



### 8.3. Universell design och rättvisa

I riktlinjelitteratur återfinns återkommande resonemang som kan sammanfattas i strävan mot ekvivalens som antas kunna uppnås genom *universell design* och med fokus på *rättvisa*, om än med olika begreppsliga benämningar mellan författarna<sup>6</sup>.

Den eller de kulturer vi lever i formar till exempel det sätt vi uppfattar hälsa eller ohälsa på och influerar det relativa värdet eller betydelsen av vissa symptom menar riktlinjeförfattarna. Dessutom menar de är det vanligt förekommande att de beteendemässiga manifestationerna av vissa koncept variera mellan kulturer. Om testets items avser att mäta till exempel depression eller kognitiv förmåga, behöver det mätas på det sätt som depression eller kognitiva förmågor manifesterar sig i just detta land/kultur (för fördjupade resonemang se till exempel Behling och Law, 2000 i Kristjansson et al., 2003 och Hambleton et al., 2005).

Vilka perspektiv och antaganden ingår i koncepten för universell design och rättvisa utifrån riktlinjeförfattarnas perspektiv? Universell design adresserar behovet av att anpassa test som är likvärdigt tillgängliga för alla testmottagare inom den avsedda test-populationen, oavsett språk och kultur, allt för att minimera risken för systematiska fel i testresultaten (så kallad konstrukt-irrelevant varians). Med universell design som förhållningssätt i testadaptionsprocessen så har testutvecklare redan inledningsvis i testanpassningsprocessen ett öga mot att maximera rättvisa för testrespondenterna (American Educational Research Association et al., 2014, s. 50). Oavsett vad syftet med testning är så är det rättvisande siktet inställt på att maximera, till möjlig grad, möjligheten för de som testas att kunna prestera/svara så korrekt som möjligt utifrån var de ”befinner sig” i förhållande till testets konstrukt, alltså det eller de begrepp som mäts i testet.

Med tanke på att de exempel som angavs i början av detta avsnitt – depression och kognitiv förmåga – involverar en myriad av språkliga och kulturella aspekter att beakta blir behovet av enhetliga procedurer ett sätt att ta sig an den vetenskapliga praktiken. Standarder blir förståeligt attraktiva verktyg då de ger stabilitet, auktoritet och funktionalitet (Timmermans, 2015, s. 81). De förmedlar bilden av att kunna undvika fallgropar, spara tid och resurser och skapa tester som är giltiga, både där och här. Den gemensamma betoningen på universell design och rättvisa lämnar ändå öppet för frågan om hur detta ska gå till rent praktiskt – trots ingående procedurellt stöd i riktlinjerna. I bakgrunden ligger den mer grundläggande frågan om huruvida, eller i vilken mån sådan ekvivalens alls är möjligt att mekaniseras eller uppnå.

Strävan efter ekvivalens via universell design länkar väl samman med standardiseringsrörelsen i stort, de följer likande logik i grund och botten. Med en global marknad eller ett globalt forskningsfält som präglas av mångfald, är enhetlighet och gemensamma nämnare en förutsättning. Frågan om huruvida enhetliga standarder som strävar efter (funktionell) ekvivalens har bärighet att förflytta och överföra relevansen i ett psykologiskt test från ett språk och kultur till ett annat språk och kultur, via universell design och fokus på rättvisa är återkommande och uppenbart intrikat.

Härnäst belyses några aspekter av transparens och dokumentationens roll inom testadaptationens fält.

---

<sup>6</sup> Det bör noteras att American Psychological Association et al. (2014) viger ett helt kapitel till *fairness* (kap. 3) och problematiserar och beaktar begreppet i fråga (som jag som författare översatt till *rättvisa*) som ett mångfasetterat begrepp som inte har någon enskild teknisk definition eller betydelse och att begreppet används på ett flertal sätt inom ramen för offentliga diskurser (s. 49). De belyser även att det råder motsättningar inom det egna fältet i hur begreppet ska förstås. Något som författarna lyfter fram som centralt inom ramen för Standards är att förstå och beakta *fairness* (*rättvisa*) som strävan mot mesta möjliga *rättvisa villkor* för samtliga testmottagare för de begrepp (så kallade konstrukt) eller skalor som avses att mätas i (de psykologiska) testen. Så som anges i texten ovan, det som främst avses är att sträva mot att samma villkor ska gälla för samtliga individer i olika subgrupper som ingår i testning – utifrån att motverka skevhet i mätningar eller olika villkor för tillgänglighet – så att testmottagarna erhålls möjlighet att demonstrera var de ”befinner sig” i förhållande till det begrepp eller den skala som de mäts på (s. 50 - 51). Se kap. 3 för utförligare presentation.

I samtliga övriga riktlinjesammanställningar som ingår i denna uppsats går detta tema igen i texterna. Dessa författare problematiserar rättvisa och argument läggs fram för vikten av att sträva efter just detta i testadaptionsprocessen, som ett led i den övergripande strävan efter ekvivalens. Se till exempel International Test Commission (ITC) (2017) s. 9 och s. 14 – 15 men även rekommenderad litteratur från ITC så som kapitel 3 (till exempel s. 73) i Hambleton et al. (2005); Beaton et al. (2000) s. 3186; och Kristjansson et al. (2003) s. 128 – 129.

## 8.4. Dokumentation

Som jag kommer att belysa präglas resonemanget om dokumentation av spänningar mellan 1) riktlinjeförfattarnas strävan efter transparens i kombination med 2) omöjligheten att helt tydliggöra den tysta expertiskunskapen och 3) risken att avskräcka ”konsumenter” genom att i manualer synliggöra testers begränsningar (etiskt patos).

I studier som undersökt redovisning av testdokumentation framkommer det att detta vanligtvis brister (Elousa och Iliescu, 2012). De flesta riktlinjeförfattare framhåller samtidigt dokumentation som särdeles viktigt för att kollegor ska kunna förstå, följa och eventuellt replikera sättet att arbeta. Dock påpekar Porter (1996) att manualer och artiklar publicerade i vetenskapliga tidskrifter utifrån rådande format ej låter den tysta, ibland kanske viktigaste kunskapen och informationen, nedtecknas. Där lyser den mekaniska objektiviteten igenom, med en tilltro till manualer som förmedlare av det väsentliga och nödvändiga i den vetenskapliga processen.

Utförlig testdokumentation och därmed hög grad av *transparens* av hela adaptationsprocessen står centralt på agendan, detta tydligast hos tre av riktlinje författarna. Min bedömning är dock att, utifrån hur jag läst Kristjansson et al. (som inte betonar dokumentation på samma sätt) så sällar de sig sannolikt snarare till än att de går emot de övriga författarnas resonemang. Det framkommer att transparens kan beröra olika aktörer i anpassningsprocessen. I Standards vigs ett helt kapitel till olika aspekter av testdokumentation, där normer för transparens presenteras och diskuteras. Standards, ITC och Beaton et al. betonar dokumentation via testmanualer som tydligt beskriver testadaptionsprocedurerna samt eventuella (negativa) konsekvenser för tillförlitlighet. De menar att testanvändare men även testrespondenter ska kunna läsa sig till vilka förändringar som gjorts och vad det betyder i förhållande till testets nuvarande giltighet och tillförlitlighet.

Dokumentation och transparens framstår som en nyckelfaktor för testanvändare (så som policyproducenter och HR-konsulter) för att kunna bedöma tester enligt gängse normer då det bidrar med ökad möjlighet att värdera ett tests kvalitet, genom dokumentation i testmanualer menar riktlinjeförfattarna. Samtidigt ställer det ett antal förväntningar och krav på dessa testanvändare. Först och främst behöver dessa känna till att testadaptation är en intrikat process som kan förändra tillförlitligheten hos ett test och dels behövs kompetens att kunna tolka och bedöma rimligheten i redovisningen (samt att det är praxis att det redovisas i manual). Det behövs förstås även ett intresse för dessa frågor inom denna målgrupp. Frågan är i vilken grad testanvändare efterfrågar och läser dessa manualer. I samtal med informanten Sjöberg informerade hon om att det är vanligt förekommande att testanvändare utgår från att ett anpassat test är lika tillförlitligt som på originalspråk. Det visar på att det kräver i sig en egen expertkunskap att förstå och kritiskt analysera det bakomliggande anpassningsarbetet men det krävs också att detta är prioriterat för kunden (andra preferenser kan föreligga). Utifrån Laoturs modell (s.13 ovan) och vetenskaplighetens cirkulation så ser vi att olika prioriteringar uppdragas i de olika öglorna, i detta fall främst spänningar mellan riktlinjeförfattarna (ögla 2) och testutvecklare eller testanvändare (ögla 3). Enligt ovan resonemang ligger ansvaret således i hög grad hos testutvecklarna att redovisa anpassningsprocessen så som riktlinjeförfattarna föreslår – detta kan ses som en förväntan på striktare följsamhet från riktlinjeförfattarna. Detta är inte oproblematiskt utifrån Sjöbergs erfarenheter som testutvecklare på Assesio. Sjöberg menar att Assesio är ytterst noggranna med att beskriva testadaptionsprocessens alla steg inklusive tillhörande styrkor och svagheter. Däremot är det långt ifrån alla kunder som läser eller sätter sig in i sådana detaljer i medföljande manualer. Ur ett konkurrensperspektiv kan det till och med vara negativt att redovisa öppet och transparant menar Sjöberg, då andra kommersiella testutvecklare (medvetet) undviker att redovisa sådant, med påföljande effekt att deras tester framstår som mer relevanta än vad de är. Att öppet redovisa brister i en adaptationsprocess och därmed påvisa försämrad tillförlitlighet av ett test i dess nya version, det kräver ett slags professionellt ”etiskt patos” hos testutvecklare, kanske särskilt när tester utvecklas i kommersiellt syfte. Inom akademiska kretsar kan detta förstås även beröra nederlaget av att behöva redovisa ”akademiskt misslyckande” och därmed

avstå från status och en fjäder i hatten gentemot (forskar)kollegor.

Dokumentation framstår alltså som både en resurs- och kanske rent av en prestigefråga, utifrån tankegångarna ovan. Sannolikt exkluderas delar av dokumentering då det tar tid i anspråk, och tid kostar. Sjöberg menar i fortsatt resonemang att det ofta handlar om att som testutvecklare behöva göra val gällande allokering av de begränsade resurser som finns tillgängliga.

Som nämndes inledningsvis i detta avsnitt råder det tydliga spänningar vad gäller dokumentationens roll och dess möjligheter. Vi kan se riktlinjeförfattarnas strävan efter transparens, för att följa vetenskaplig praxis, delvis på kollisionskurs med det som kanske sällan synliggörs i diskussioner, nämligen omöjligheten att helt tydliggöra den tysta expertiskunskapen som ligger bakom en forskningsprocess. Därtill visar den praktiska verkligheten på risken att avskräcka ”konsumenter” genom att i manualer synliggöra testers begränsningar, även där dessa tester kanske anpassats mer i linje med standarder än andra tester, där sådant inte dokumenteras och därför får fördel på marknaden.

Efter att ha belyst ett urval av spänningar inom dokumentation låt oss gå vidare till en framträdande statistisk processingrediens: psykometrisk utvärdering.

## 8.5. Psykometrisk utvärdering

Resonemang som förs fram av riktlinjeförfattarna är att det psykometriska testningsförfarandet har fått en allt större plats i rekommendationer för testanpassning, detta i takt med ett mer globaliserat forskarfält med ökat intresse för tillförlitliga(re) tvärkulturella jämförelser och även ökad efterfrågan på standardisering av test som förväntas fungera med god precision (i testpoäng) för flera målgrupper. Psykometri avser metoder som kan granska och utvärdera testers kvalitet gällande att mäta de psykologiska egenskaper eller attityder som de utger sig för att mäta. Psykometriska resonemang och procedurer berörs inom samtliga ramverk och riktlinjer<sup>7</sup>.

Något som berör praktikerperspektiv utifrån epistemologiska - men också metodologiska - resonemang är den genomgående avsaknaden av kritisk reflektion över vad statistiska beräkningar (som psykometri utgörs av) kan – och inte kan – bistå med, något som även nämnts tidigare. Jag uppfattar riktlinjeförfattarna – men även de som kritiserar nuvarande modeller av riktlinjer, så som Epstein et al. (2015) och Byrne (2016) – som att de förbehållslöst ger stor tilltro till en rad olika psykometriska metoder och de värden som produceras därigenom.

Inom riktlinjelitteraturen ligger Standards och ITC i framkant gällande produktion och publicering av vägledande manualer, kapitel i böcker och artiklar. Beaton et al. och Kristjansson et al. hänvisar istället till externa källor för metodstöd. De statistiska modellerna framstår som att de blivit alltmer avancerade och sofistikerade med nyare utgåvor av riktlinjer.

Som belysts ovan är psykometri en statistisk disciplin med syfte att granska kvalitet för bland annat psykologiska tester. Med goda mått för precision och olika former av validitet anses generellt sett ett test vara mer tillförlitligt och därmed mer lämpligt, än tester med sämre sådana värden, det är författarna överens om. Till exempel kan i Standards läsas (American Educational Research Association et al., 2014, s. 85) att det är brukligt att testutvecklare presenterar och ger test en kvalitetsreferens genom tillhörande psykometriska analyser, som uttrycks framför allt i begreppen *validitet* som syftar på giltighet och relevans (att testet mäter vad det avser att mäta) och *reliabilitet* som berör tillförlitlighet (noggrannhet och precision för det som ska testas och mätas). Innebörden av detta är att resultatet bör vara detsamma vid upprepade mätningar (så kallad test-retest-reliabilitet), och oberoende av vem som utför testet (så kallad interbedömar-reliabilitet). Olika former av validitet som återkommande beskrivs som väsentliga och vitala att undersöka och rapportera om är främst: innehålls-; kriterie-; begrepps-; konvergent- och; diskriminerande validitet (utförlig beskrivning av de

<sup>7</sup> Se till exempel i Standards s. 79 – 80, s. 82 (American Educational Research Association et al., 2014); ITC s. 16 - 24 (International Test Commission, 2017); Beaton et al (2000) s. 3189; samt Kristjansson et al. (2003) s. 138 - 139.

olika validitetsbegreppen återfinns till exempel i Standards, se kapitel 1 i American Educational Research Association et al., 2014 som avhandlar validitet men även i hänvisningar i de övriga riktlinjerna som återges i uppsatsen).

Statistik används ofta som ett verktyg för att stärka påståenden och ge en objektiv bild. Porter (1996) menar att (statistisk) kvantifiering är att fatta beslut utan att fatta beslut (s. 8). Forskaren blir i detta sammanhang en budbärare och siffrorna ges legitimitet att tala för sig själva. Som belysts ovan beskriver Porter hur nya discipliner i större utsträckning *behöver* statistik för att hävda sin auktoritet. För testadaptation som disciplinärt fält tycks detta inte orimligt (de svaga gör detta för att ta sig in i vetenskapens finrum) där statistiken blir en slags grundkärna för evidensbaserad, enligt epistemologiska normer och mekanisk golden standard. Med en kritisk blick kan psykometriska utvärderingar antas anammas som giltig kunskap/resultat alltför oreflekterat. Vad som inkluderas respektive exkluderas i bakomliggande processer problematiseras ej – siffrorna ges tolkningsföreträde och något så komplext som tvärkulturella jämförelser blir via psykometriska procedurer tillsynes genomförbara (om siffrorna tillåter detta). Jag ser det som att det finns en risk att sökandet efter signifikansvärden grumlar analytisk reflektion kring vad beräkningarna egentligen kan uttala sig om (och vad som faktiskt osynliggörs). Ur ett disciplinärt perspektiv och med blick på Latours ögla nummer 5 och 2 (men även den tredje ögla) så ter det sig som att en mekanisk logik förväntas ge en förstärkt tilltro till kunskapsfältet i fråga och i förlängningen dess företrädare.

Om vi följer Porters tankegångar är psykometri en teknik som hanterar avstånd. Siffror, tabeller och grafer bidrar med att lägga fram världen som förståbar, på ett igenkänningsbart, överblickbart och standardiserat sätt som kan förstås och (mer eller mindre) enkelt förflyttas världen över. Det som är mest avgörande är att, genom att förlita sig på siffror och kvantifiering minimeras behovet av intim kunskap och personligt tillitsskapande. Forskning tenderar att ses som mer upphöjd eller betrodd, ju längre den är från mänsklig (subjektiv) inblandning (Bohlin och Sager, 2011). Ett sätt att uttrycka det är att vi alltså ska hålla fingrarna i styr. Detta ger en form av objektivitet som kämpar med och mot subjektivitet, en objektivitet som utesluter personligt omdöme. Som belysts tidigare, logiken bakom strategin säger att det går att lösa (eller vinna) interna konflikter och kontroverser via kvantifiering likväl som att skydda sig mot yttre hot. Därtill så definierar detta även synsätt på kunskap som i sig ger (vissa) företräden. Jag återkommer därför till det Porter säger om att kvantifiering *icke endast* är en politisk strategi för politiska problem, men det är likväl detta också och det bör förstås och synliggöras. Mekanisk objektivitet har under en längre tid setts som vetenskapens rågång. Det sociala i forskningsprocessen har dock alltid funnits med men ej synliggjorts eller framhållits i sfärer av mekaniska ideal.

Efter att analytiskt utforskat dessa ramverk och riktlinjer – i relation till praktisk tillämpning – genom några huvudsakliga genomlysningar gällande centrala idéer och begrepp, utformning och innehåll, samt att ha begrundat olika tillhörande kunskapsperspektiv är det dags att samla ihop trådarna och summera de lärdomar som framkommit.

## 9. Diskussion och slutsatser

Uppsatsen har syftat till en mer samlad kunskapsbild av testadaption med särskilt fokus på samband mellan riktlinjer och praktik (riktlinjers relevans och stöd för praktiker). Det analytiska syftet har bestått av att djupare förstå spänningsfälten, snarare än att presentera en lista med punkter på ”gör så här”. Kunskapsökandet har väglatts dels av det underliggande symmetriska synsättet som presenteras i ANT och Latours modell för vetenskaplighetens cirkulation (blomman med dess öglor, s. 13 ovan). Modellen – dock ej den enda existerande – företräder ett idéperspektiv på hur vetenskap blir vetenskap. Även om riktlinjeförfattarnas prioritet inte är detta lämpar sig modellen eftersom den för blicken till samspelet mellan olika delar av ett teknovetenskapligt system och hur beroende det vetenskapliga arbetet är också av ”icke-vetenskapliga” öglor.

Uppsatsens vetenskapsteoretiska analyser har vidare haft sin grund i Porters (1996) objektivitetskategorier (mekanisk/procedurell kontra disciplinär/expert) samt standardiseringens sociologi (Timmermans och Berg, 2003 och Timmermans, 2015) som ger kritiska perspektiv på standarder som sådana. Dock förespråkar de sistnämnda en kombination mellan mekanisk objektivitet (procedurell standardisering) och disciplinär objektivitet (de professionellas subjektiva bedömande, hur arbeta utifrån standarder inom ramen för deras rådande förutsättningar), i det som de kallar flexibel standardisering. De menar att det intressanta inte är att debattera standarders vara eller icke vara utan snarare hur kan dessa utformas och anammas med ett aktivt aktörskap från de professionella som ska arbeta utifrån dessa. Med hjälp av dessa analytiska begrepp och perspektiv: vilken kunskapsbild har framkommit och har frågeställningarna besvarats?

Utifrån *frågeställning 1 och 2* så har vi fått kunskap om olika ramverk och riktlinjer och deras utformning samt information om hur de främst (eller inte) sprids. De standarder och riktlinjer som belysts ovan är avsedda att fungera som stödjande och guidande ramverk för språkliga och kulturella testadaptionprocesser. Tvärkulturell adaption är en förberedande process för att anpassa och använda ett befintligt test i ny miljö, med strävan efter att uppnå likvärdighet mellan ursprunglig och ny målgrupp vad gäller sådant som mätbara begrepp och karaktäristika (konstrukt) och innehåll. Detta är vad som kan kallas kampen för ekvivalens.

Utifrån den *tredje frågeställningen* har vi sett att riktlinjeförfattarnas kamp för ekvivalens innehåller både disciplinär och mekanisk objektivitet, dock med betoning på den mekaniska. Det intressanta är spänningen mellan de synsätt som framkommer i riktlinjerna och även i adaptionfältet i stort. Trots den samsyn som finns kring svårigheterna med tvärkulturell översättning drar företrädare till viss del åt olika håll just vad gäller tilliten till det mekaniska och det disciplinära. Inom riktlinjeproduktionen ryms som jag ser det en slags (dold) dragkamp mellan det mekaniska och det disciplinära som innefattar inte bara en proceduralisering, men också en rörelse mot hårdare disciplinering. Detta är ett oväntat resultat som går mot Porters teser och även, till viss del, resultaten inom standardiseringssociologin. Vad skulle hittas om studier av standarder i högre utsträckning parallellt fokuserade på förekomsten av en disciplinär åtstramning? Timmermans och Berg (2003) visar i kapitel 4 hur mekanisk objektivitet kan *öppna för* disciplinär objektivitet i försäkringsmedicinen. Det som antyds i riktlinjerna för testadaption är att betoning av mekanisk objektivitet följs åt av ökade krav på den disciplinära objektiviteten, en åtstramning av vad som ska räknas som expertkunskap, vilket är något annat. Timmermans och Epstein (2010) visar med sitt begrepp *flexibel standardisering* att omdöme och standarder förutsätter varandra, men de ser inte en åtstramning av kraven på vems omdöme som räknas, vilket är fallet här.

Kritiker (Epstein, 2015 och Byrne, 2016) driver frågan mer åt mekanisk objektivitet i sina krav på

en gyllene standard då de menar att detta ej finns på plats idag (i alla fall inte när artiklarna skrevs). I vilken utsträckning riktlinjeförfattarna låter sig influeras av kritik kan jag inte uttala mig om. Det råder dock överlag bland riktlinjeförfattarna en strävan efter att renodla procedurerna, inte mist inom de psykometriska utvärderingarna. Det förekommer en uppsjö av idéer gällande att säkerställa ekvivalens via psykometriska modeller och metoder, detta samtidigt som anpassningsprocessen av tester är så intrikat förknippat med de professionellas egna bedömningar och reflektioner. Den disciplinära expertisen framhålls i ramverk och riktlinjer men samtidigt hänvisas det till psykometriska metoder för att undkomma just bedömningsmomentet. Sådant kan förklarligt vara ytterst avgörande gentemot kollegor i forskningssamhället (Latours ögla nummer 2). Detta kan tolkas som att bedömningar och reflektioner inte betraktas som tillräckliga eller likvärdiga i riktlinjernas giltighetshierarki. Det är i slutändan statistiska/psykometriska procedurer och kvantitativa mått som avgör ett tests kvalitet. Funktionell ekvivalens via psykometri tenderar ge sken av överförbarhet i en utsträckning som kanske inte alltid stämmer eller är möjlig, åtminstone inte om Latour (2005) har rätt i att vi lever ”in a messy world” – men det ger uppenbarligen vetenskaplig status. Riktlinjefältet återspeglar i denna mening den tillit till kvantifiering och mekaniserade procedurer som till stor del präglar såväl vetenskap och samhälle.

Utifrån den *fjärde frågeställningen* – som frågar om det finns kunskapsperspektiv som riktlinjeförfattarna skulle behöva beakta och inkludera för att främja ökad tillämpbarhet av standarderna och riktlinjerna – så har det framkommit ett antal svar som relaterar till Latours tredje ögla (allierade): Vad gäller riktlinjernas utformning behöver testutvecklarna – så som Berthelsen och Sjöberg beskriver - ett ännu mer anpassat arbetssätt utifrån syfte och mål med testanpassningen utifrån egna tillgängliga resurser, men också yttre krav och förväntningar (från kunder till exempel). Detta gäller genom hela anpassningsprocessen. Som konstaterats ovan: en ”fullskalig” adaptationsprocess är svår att möjliggöra på grund av att den är resurskrävande – finansiellt, kompetensmässigt, tidsmässig och så vidare – men också för att testanvändare – det vill säga de som köper eller nyttjar tester för olika bruk, så som HR-konsulter eller policyproducenter - har andra preferenser än evidensbaserad ekvivalens. För testanvändarna är exempelvis logistisk smidighet och redan gjorda investeringar ofta mer styrande än vetenskaplig stringens. Denna typ av olika prioriteringar och perspektiv som kan upplevas som ”störande” ur riktlinjeförfattarnas horisont är en del av den större verkligheten. Detta är en del i kunskapsbilden som vuxit fram genom att inkludera Latours olika öglor i analysen.

För testutvecklarna – i alla fall de två som intervjuats i uppsatsen – råder ständig balansgång och avvägningar som utgör en naturlig del i att försöka översätta kultur och samtidigt hantera resursfrågor och olika intressenters behov och preferenser. Där är det inte i första hand en dragkamp mellan olika typer av objektivitet. Snarare än olika objektiviteter handlar det här om en dragkamp mellan riktlinjeförfattarnas objektivitet och testutvecklarnas tillvaro och situation som inte alls alltid bryr sig om objektivitet. Så som Latours modell hjälpt oss att få syn på så råder olika behov och perspektiv i de olika öglorna. Det är inte alltid traditionell, vetenskapligt definierad objektivitet som är det primära. Öglorna med dess intressenter har olika perspektiv. Detta hjälper oss blommans symmetri att se: det som vissa skulle säga är ogiltigt, ovetenskapligt, ”smutsigt”, ”biased” och så vidare men som faktiskt finns där och spelar roll. Det betyder att fler logiker behöver få rymmas.

Sett till resultaten av denna studie – Vad kan vi ta med förutom förståelse av att fler logiker behöver utrymme och länkas samman i det vetenskapliga systemet?

Inför *den femte frågeställningen* – Hur kan teoretiska instruktioner behöva förändras för att bli mer användbara? – ger Timmermans och Bergs (2003) idéer om flexibla standarder – ej mer detaljerade än behövt, ej mer stringenta än nödvändigt, ej mer tvingande än användbart – delvis vägledande

perspektiv. Utifrån olika syften och behov skulle det kunna finnas flera alternativa *omfattningar* av testadaptionprocessens ingående steg. Olika grader av resurskrävande alternativ för de obligatoriska stegen skulle kunna arbetas fram. Inom evidensrörelsen är sådana nivåer av omfattning ett växande fenomen. Det har blivit tydligt att en fullödlig genomgång av all tänkbar litteratur på ett område ibland tar för lång tid för att möta behoven av vägledning. Tydligaste exemplet är framväxten av så kallade *rapid reviews* (Hailey et al., 2000; Tricco et al., 2015). En liknande situation tycks råda för riktlinjeförfattarna och testutvecklarna, de är således inte ensamma i sin problematik. De tidigare idealen visar på hur öglorna i Latours modell inte hakar i varandra eller inte hänger samman, de blir utan gemensam knutpunkt. Det framstår som lite märkligt att vissa riktlinjeförfattare som själva är testutvecklare inte anammat detta med anpassade utformningar av standarder. Dessa samma personer med olika ”hattar” på verkar inte kunna kompromissa vid ”hattbyten”. Slutsatsen som jag dragit är att den epistemologiska logiken mellan kollegor är väldigt stark, eller riskfylld att ifrågasätta.

Ett annat sätt är att ta till vara på och stärka utbyte av olika kunskaper och erfarenheter genom att bjuda in till interdisciplinära forum för att involvera fler intressenter, så som testanpassade praktiker och testanvändare i utformning av riktlinjer. Detta pågår delvis via ITC:s arbete med konferenser men även Copsoq-nätverket är ett sådant exempel. Enligt Berthelsen är deras nätverk nära nog unikt när det gäller internationell samverkan inom ett särskilt instrument. Denna form av kunskapsutbyte och kunskapsbyggande skulle behöva uppvärderas där testutvecklarnas pragmatiska vardag ges samma legitima utrymme som strävan efter de perfekta psykometriska modellerna för att värdera ekvivalens. Olika kunskapsperspektiv behöver således samsas och integreras. Det framstår som att kommunikation och utbyte av kompetens och kunskap mellan fler aktörer skulle vara en framgångsfaktor, för att forma riktlinjer som än bättre verkligen stödjer praktisk verklighet. Sådana forum och arbetssätt framhålls i anslutning till ökade krav på evidens inom praxisnära forskning och evidensinformerad praktik (Carlgren et al., 2015). Praxisnära kunskapsbyggande och forskning är att orientera sig mot problem som andra än forskarna har, alltså problem för deltagarna i den verksamhet man avser att stödja. När det gäller testadaption är det inte särskilt svårt att identifiera sådana problem, om vi betraktar det hela från praktikernas, det vill säga testutvecklarnas sida. Slutsatsen är att standarder skapas bäst när det finns en tydlig respekt för disciplinära aspekter av objektivitet men även pragmatiska faktorer, såsom tillgång på resurser. Denna objektivitet innefattar det som är socialt, personligt och omdömesgrundande utifrån kontextuella förutsättningar som legitim kunskapskälla.

Vi har tillika den intrikata epistemologiska frågan, som knyter an till den fjärde frågeställningen, om i vilken utsträckning kultur är översättningsbart – via psykologiska test. Finns det en realistisk potential i visionen om eller kampen för ekvivalens? Utan denna strävan är det enligt förespråkarna givet att testmottagare ej får tillräckligt rättvisa villkor i testsammanhang. Detta kan i så fall påverka deras chanser till både studier och arbete, eller stöd för psykologisk behandling och välbefinnande. Begreppet ekvivalens blir samtidigt mycket viktigt för riktlinjeförfattarnas strävan att stärka sin egen disciplin och den tvärkulturella forskningens vetenskaplighet (öglan nummer 2 men även öglan nummer 3 och 4 som berör inflytande i samhället). En fråga som dock uppstår är om (kulturell) ekvivalens är möjlig eller är det en för mekanisk syn på människor och kultur? Enligt den modell som denna uppsats utgår ifrån, nämligen ANT, är världen rörig (”messy”, Latour, 2005, s. 136). När ekvivalens fungerar är det för att system eller svarta lådor har skapats för friktionsfri överföring inom vissa specifika områden, såsom mellan laboratorier och hamburgerrestauranger, inom ett järnvägsnät eller ett valutasystem. Huruvida ekvivalens i testning av människors psykologiska egenskaper kan etableras är en öppen fråga för framtida studier med ANT (eller andra ansatser). Utifrån de riktlinjer som studerats här verkar dock hindren stora, kanske till och med oöverkomliga.

Ett sätt att se det är att strävan efter detta – kulturell ekvivalens - ingår i en större samhällsrörelse, där tvärkulturell forskning och standardisering behövs, för att stödja behov som anses prioriterade av olika samhällsinstanser. Då hamnar de epistemologiska problemen med ekvivalens i bakgrunden till förmån för ”praktisk hantering” genom en mekanisk objektivitetssyn. Här liknar adaptation av tester de samhällsekonomiska beräkningar som ofta är prioriterade, men omöjliga att göra. Porter (1996) menar att just ett sådant behov av att praktiskt hantera komplexa problem låg bakom framväxten av så kallade cost-benefit-kalkyler (kap. 7). Trots att komplexiteten är ohanterlig skapas en modell som ger ett sken av att vara exakt. Detta vet de flesta om, men modellerna hanterar problemen vilket gör att aktörer kan fortsätta med ”business as usual”.

Det kanske framstår som att enda alternativet är att bli paralyserad men det är inte bättre som jag ser det. Det finns fortfarande ett arbete att göra, då uppgifterna kvarstår trots epistemologiska problem. Det är svårt att argumentera emot strävan efter så rättvisa testvillkor som möjligt för alla de testmottagare som blir föremål för eller utsätts för dessa av olika anledningar. Detta parallellt med att se att standarder företräder makt och politik (intressekonflikter).

Den slutliga frågan om vilka lärdomar som kan dras kan sammanfattas i det som framkommit i vetenskaplighetens cirkulation – där koncept och kontext samspelar. Slutsatserna är:

- Riktlinjernas ”ankdamm”: Utifrån de standarder som granskats framstår det som att en mindre krets av forskare och associerade organisationer är tongivande för publicering, produktion, spridning och diskussion av riktlinjer och standarder. Vissa är även själva testutvecklare
- Visionen om ekvivalens: Överlag har den uppsättning standarder som presenteras övervägande gemensamma synsätt som kan sammanfattas i strävan efter funktionell ekvivalens, universell design och rättvisa.

Ett antal spänningar mellan olika sätt att argumentera för riktlinjers kvalitet blir tydliga, spänningar som riktlinjeproducenter och testutvecklare antagligen bör reflektera över:

- Mellan procedurer och människor: I visionen om eller kampen för ekvivalens förlitar sig aktörerna (riktlinjeförfattarna) mestadels på vad som kan beskrivas som en mekanisk objektivitet, alltså en tillit till procedurer, även om disciplinär objektivitet, alltså tillit till professionellt omdöme, ges visst utrymme.
- Hårdare disciplinering: Inslagen av disciplinär objektivitet skiljer sig på avgörande sätt från generella hänvisningar till standardiseringens saliggörande effekter inom delar av evidensrörelsen. Proceduralisering följs åt av nästan omöjliga krav på expertis vid testutveckling. Det handlar inte enbart om en öppning för bedömning, utan krav på vem som ska få genomföra bedömningarna.
- Begränsad transparens: Trots att många ställer krav på dokumentation i riktlinjebaserad testadaptation verkar inte tyst expertkunskap inkluderas i sådana krav på transparens.
- Hopplös kamp för ekvivalens? En springande fråga synes vara i vilken utsträckning kulturella jämförelser kan mekaniseras: Kan disciplinär objektivitet omvandlas till (mekaniska) procedurer utan att kulturellt känsliga element reduceras bort för mycket? I förhållande till detta presenteras idéer om flexibel standardisering liksom pragmatisk evidensinformerad praktik.



- Riktlinjernas irrelevans: Förslaget för mer flexibel standardisering riskerar dock att bli irrelevant. Testutvecklarens faktiska arbete verkar i vissa eller ibland stora drag bortse från riktlinjeförfattarnas kamp för att uppnå ekvivalens i översättning av riktlinjer, främst utifrån resursfrågor men även av affärsrättsliga skäl.

Inom detta (testadaptionens) fält tycks det finnas en direkt parallell mellan universell design och idealet om mekanisk objektivitet och därmed också tilliten till standarder, i alla fall från riktlinjeförfattarna. Detta är något som skulle kunna kritiserars hårt. Samtidigt, kan jag som analytiker stå på ett visst avstånd från interna stridigheter eller andra uppenbara spänningar utan några egna uppenbara risker att förlora något och utan att min position på något sätt blir hotad. Denna position är analytikerens privilegium medan aktörerna däremot – såsom riktlinjeförfattarna – är fullt inblandade i sina öglor och system. Deras egna positioner står på spel tillsammans med de kunskapsanspråk och de föreställningar om forskning som de presenterar (jmf. Sager, 2015). Kunskap för aktörerna är inte enbart en akademisk fråga utan också en fråga om att själva överleva epistemiskt och socialt, det vill säga erhålla resurser för att fortsätta sin forskning. Som analytiker – eller läsare av analyser – behöver vi med viss ödmjukhet väga in denna skillnad i position.

## 10. Referenslista

- Alonso-Arbiol, I., & van de Vijver, F. J. R. (2010). A historical analysis of the European Journal of Psychological Assessment. *European Journal of Psychological Assessment*, 26, 238–247.
- American Educational Research Association, American Psychological Association, National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- American Educational Research Association, American Psychological Association, National Council on Measurement i Education. (2014). *Standards for Educational and Psychological Testing*. Publied by American Educational Research Association, US, Washington, DC.
- Banville, D., Desrosiers, P., & Genet-Volet, Y. (2000). Translating questionnaires and Inventories Using a Cross-Cultural Translation technique. *Journal of Teaching in Physical Education*, 19, 374-387.
- Beaton, D., Bombarider, C. & Guillemin, F. (2000). Guidelines for the Process of Cross-Cultural Adaption of Self-Report Measures. *SPINE*, 25 (24), 3186-3191.
- Behling, O., & Law, K. (2000). *Translating questionnaires and other research instruments: Problems and solutions*. Sage University Papers Series on Quantitative Applications in the Social Sciences, 07-131. Thousand Oaks, CA: Sage.
- Berg, M., Horstman, K., Plass, S., & van Heusden, M. (2000). Guidelines, professionals and the production of objectivity: standardisation and the professionalism of insurance medicine. *Sociology of Health & Illness* no. 22 (6):765 – 791.
- Bernal, G., & Domenech Rodriguez, M. M. (2012). *Cultural Adaptations: Tools for Evidence-Based Practice with Diverse Populations*. Footprint Books, Australia.
- Berthelsen, H., Westerlund, H., & Söndergård, T. (2014). COPSOQ II - en uppdatering och språklig validering av den svenska versionen av en enkät för kartläggning av den psykosociala arbetsmiljön på arbetsplatser. *Stressforskningsrapport* nr: 326.
- Bloor, D. (1976). *Knowledge and social imagery* (2nd ed.). Chicago, Illinois: University of Chicago Press.

Bohlin, I., & Sager, M. (2011). *Evidensen många ansikten*. Lund: Arkiv förlag.

Busch, L. (2011). *Standards: Recipes for reality*. Cambridge, MA: The MIT press.

Byrne, B. (2016). Adaption of Assessment scales in Cross-National research- Issues, Guidelines, and Caveats. *International Perspectives in Psychology: research, Practise, Consultation*, Vol. 5, No 1, 51 – 65.

Candell, G. L., & Hulin, C. L. (1987). Cross-lanugauges and cross-cultural comparisons in scale translations. *Journal of Cross-Cultural psychology*, 17, 417 – 440.

Carlgren, I., Josefson, I., Liberg, C., Anward, J., Evaldsson, Ann-Carita., Marton, Ference., Nordenstam, T., Orre, I., & Weiner, G. (2005). *Forskning av denna värden II – om teorins roll i praxisnära forskning*. Vetenskapsrådet. ISBN 91-7307-065-3.

Caro, I., & Stiles, W.B. (1997). Vamos a traducir los MRV (Let's translate the VMR): Linguistic and cultural inferences drawn from translating a verbal coding system from English to Spanish. *Psychiatry*, 60, 233 – 247.

Cartwright, N., & Hardie, J. (2012). *Evidence-based policy*. Oxford University press.

Elousa, P., & Iliescu, D. (2012). Tests in Europé: Where we are and where we should go. *International Journal of Testing*, 12: 157–175, 2012. DOI: 10.1080/15305058.2012.657316.

Epstein, J., Miyuki Santo, R., & Guillemin, F. (2015). A review of guidelines for cross-cultural adaption of questionnaires could not bring out a consensus. *Journal of Clinical Epidemiology*, 68, 435 – 441.

Erickan, K. (2002). Disentangling sources of differential item functioning in multilanguage assessments. *International Journal of testing*, 2 (3), 199 – 215.

Feinstein A. R. (1987) The Theory and Evaluation of Sensibility. In: Feinstein, A. R., Ed. *Clinimetrics*. New Haven: Yale University Press, 141–166.

Fernler, K. (2012). Perspektiv på implementering. Vad är god implementering och kan det stödjas? *Leading Health Care Report*, nr: 8.

Gibbons, M., Liboges, N., Nowotny, H., Schwartzman, S., Scott, P. & Trow, M. (1994) *The new production of knowledge: The dynamics of science and research in contemporary societies*. London: SAGE.

Geisinger, K. F. (1994). Cross-Cultural Normative Assessment: Translation and Adaptation Issues in Influencing the Normative Interpretation of Assessment Instruments. *Psychological Assessment*, 6, 304 – 312.

Gregoire, J., & Hambleton, R.K. (2009). Advances in Test Adaptation Research: A Special Issue. *International Journal of Testing*, 9: 75–77, 2009. DOI: 10.1080/15305050902880678.

Gudmundsson, E. (2009). Guidelines for adapting psychological instruments. *Nordic Psychology*, 61(2), pp. 29 – 45.

Guillemin, F., Bombardier, C., & Beaton, D. (1993). Cross-Cultural Adaption of Health-Related Quality of Life Measures: Litterature Review and proposed Guidelines. *Journal of Clinical Epidemiology*, 46 (12), 1417-1432.

Hailey, D., Corabian, P., Harstall, C., & Schneider, W. 2000. The use and impact of rapid health technology assessment. *International Journal of Technology Assessment in Health Care*, 16, 651-656.

Hambleton, R. K. (2001). The next generation of the ITC test translation and adaptation guidelines. *European Journal of Psychological Assessment*, 17, 164 -172.

Hambleton, R. K. (2002). Adapting achievements tests into multiple languages for international assessments. In A. C Porter & A. Gamoran (Eds.), *Methodological advances in cross-national surveys of educational achievement* (pp. 58-76). Washington: National Academy of Sciences.

Hambleton, R. K. (2005). Issues, Designs and Technical Guidelines for Adapting Tests Into Multiple Languages and Cultures. In R. K. Hambleton, P. F. Merenda and C. D. Spielberger (Eds.). *Adapting Psychological and Educational Tests for Cross-Cultural Assessment*. NJ: Lawrence S. Erlbaum Publishers.

Hambleton, R. K., & de Jong, J. (Eds.). (2003). Advances in translating and adapting educational and psychological tests. *Language Testing*, 20(2), 127-240.

Hambleton, R. K., Merenda, P., & Spielberger, C. (Eds.). (2005). *Adapting educational and psychological tests for cross-cultural assessment*. Hillsdale, NJ: Lawrence S. Erlbaum Publishers.

Hambleton, R. K., & Patsula, L. (1998). Adapting tests for use in multiple languages and cultures (special issue on Validity Theory and the Methods Used in Validation, edited by B.D. Zumbo). *Social Indicators Research*, 45, 153 –171.

Herdman, M., Fox-Rushby, J., & Badia, X. (1998). A model of equivalence in the cultural adaptation of HRQoL instruments: the universalist approach. *Qual Life Res*; 7:323-335.

Howick, J. (2011). *The Philosophy of Evidence-Based Medicine*. Wiley-Blackwell.

Hulin, C.L. (1987). A psychometric theory of evaluations of item and scale translations: Fidelity across languages. *Journal of Cross-Cultural Psychology*, 18, 115–142.

International Test Commission. (2005). *International Guidelines on Test Adaptation*. Hämtad 15 juni 2016 från International test Commission: <https://www.intestcom.org>.

International Test Commission. (2017). *The ITC Guidelines for Translating and Adapting Tests (Second edition)*. [www.InTestCom.org]

Kristjansson, E. A., Desrochers, A., & Zumbo, B. (2003). Translating and Adapting Measurement Instruments for Cross-Linguistic and Cross-Cultural Research: A Guide for Practitioners. *Canadian Journal of Nursing Research*, 35 (2), 127-142.

Latour, B. (1999). *Pandora's Hope. Essays on the Reality of Science Studies*. Cambridge, MA; London, UK: Harvard University Press.

Latour, B. (2005). *Reassembling the Social - An Introduction to Actor-Network-Theory*. USA: Oxford University Press.

Law, J. (2007). *Actor-Network Theory & Material Semiotics*. (2007). Kan nedladdas via: <http://www.heterogeneities.net/publications/Law2007ANTandMaterialSemiotics.pdf>.

Levinsson, M. (2013). *Evidens och existens: evidensbaserad undervisning i ljuset av lärares erfarenheter*. Göteborg: Acta Universitatis Gothoburgensis.

Nationalencyklopedin, psykometri. <http://www.ne.se/uppslagsverk/encyklopedi/lång/psykometri> (hämtad 2016-09-15).

Nilsen, P. (2014). *Implementering av evidensbaserad praktik*. Gleerups, Malmö.

Nilsson Kajermo, K., Böe, H., Johansson, E., Henriksen, E., McCormack, B., Gustavsson, J. P., & Wallin, L. (2013). Swedish Translation, Adaptation and Psychometric Evaluation of the Context Assessment Index (CAI). *WVN* 2013; 10:41–50.

Nordin, S., Elf, M., McKee, K., & Wijk, H. (2012). Att översätta och anpassa ett engelskt instrument för bedömning av fysisk miljö på särskilt boende till svenska förhållanden – utmaningar och erfarenheter. Presenterad vid International Conference ARCH12 (Architecture, Research, Care and Health), Göteborg, 12-14 November, 2012.

Oakland, T. (2005). Selected Ethical Issues Relevant to Test Adaptions, i: Hambleton, R. K., Merenda, P., & Spielberger, C. (Eds.). (2005). *Adapting educational and psychological tests for cross-cultural assessment*. Hillsdale, NJ: Lawrence S. Erlbaum Publishers.

Oakland, T., Poortinga, Y. H., Schlegel, J., & Hambleton, R. K. (2001). International Test Commission: Its History, Current Status, and Future Directions. *International Journal of Testing*, 1(1), 3-32.

Plake, B. S., & Wise, L. L. (2014). What Is the Role and Importance of the Revised AERA, APA, NCME Standards for Educational and Psychological Testing? *Educational Measurement: Issues and Practice*, Winter 2014, 33(4), pp. 4–12.

Porter, T. (1996). *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton, NJ: Princeton University Press.

Rios, J., & Sireci, S. (2014). Guidelines versus practices in cross-lingual assessment: A disconcerting disconnect. *International Journal of Testing*, 14(4), 289-312.

- Sager, M. (2015). Etiska aspekter av osäker vetenskap och vaccinkontroverser: konstruktivistiska bidrag. *Socialmedicinsk tidskrift*, 6.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2003). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Houghton Mifflin Company, USA, Boston, NY.
- Sireci, S., & Faulkner-Bond, M. (2014). Validity evidence based on test content. *Psicothema*, Vol 27(6), pp. 465 - 476.
- Sireci, S. G., Yang, Y., Harter, J., & Ehrlich, E. J. (2006). Evaluating Guidelines for Test Adaptations - A Methodological Analysis of Translation Quality. *Journal of Cross-Cultural Psychology*, Vol. 37, 5: pp. 557-567.
- Slaney, K. L., Tkatchouk, M., Gabriel, S. M., & Maraun, M. D. (2009). Psychometric assessment and reporting practices: Incongruence between theory and practice. *Journal of Psychoeducational Assessment*, Vol. 27(6), pp. 465-476.
- Timmermans, S. (2015). Trust in Standards – Transitioning clinical exome sequencing from bench to bedside. *Social Studies of Science*, Vol. 45 (1), pp. 77-99.
- Timmermans, S., & Berg, M. (2003). *The Gold Standard – The Challenge of Evidence-Based Medicine and Standardization in Health Care*. Philadelphia; Temple University Press.
- Timmermans, S., & Epstein, S. (2010). A World of Standards but not a Standard World: Toward a Sociology of Standards and Standardization. *The Annual Review of Sociology*, Vol. 36, pp. 69–89.
- Triandis, H. C. (1994). *Culture and social behaviors*. New York; McGraw-Hill.
- Tricco, A. C., Antony, J., Zarin, W., Striffler, L., Ghassemi, M., Ivory, J., Perrier, L., Hutton, B., Moher, D., & Straus, S. E. (2015). A scoping review of rapid review methods. *BMC Med*, 13, 224.
- van de Vijver, F. J. R., & Hambleton, R. K. (1996). Translating tests: Some practical guidelines. *European Psychologist*, 1, 89.
- van de Vijver, F. J. R., & Leung, K. (1997). *Methods and data analysis for cross-cultural research*. Thousand Oaks, CA: Sage.
- van de Vijver, F. J. R., & Tanzer, N. K. (1997). Bias and equivalence in cross-cultural assessment: An overview. *European Review of Applied Psychology*, 47, 263 – 279.

## 11. Bilagor

### 11.1. Bilaga 1: Litteratursökning.

Avsikten med litteratursökningen har varit att hitta, jämföra och sammanställa riktlinjer för testadaptation samt relevant angränsande kunskapsbärande litteratur. Jag sökte efter publicerade riktlinjer och standarder samt vägledning i systematiska tillvägagångssätt för testadaptation.

Litteratursökning genomfördes via databaser tillgängliga via universitetsbiblioteken vid Umeå Universitet och Göteborgs Universitet. Initialt behövde jag sätta mig in i fältet genom att förutom att söka riktlinjer för anpassning av psykologiska test även söka angränsande litteratur som problematiserar och perspektiviserar inom kunskapsområdet. Litteratursökningens karaktär skiftade under tidens gång. Den rörde sig från att innefatta ett vidare sökljus och fält till att allt mer avsmalna utifrån bedömd relevans. Samtidigt har en slags snöbollsmetod nyttjats då vissa sökvägar lett till ökad tillgång till relevant litteratur.

Inledningsvis genomfördes sökning i databasen PsycInfo, eftersom studien är inriktad på tester av psykologisk karaktär. Processen upprepades parallellt i PsycArtic för att se om denna databas innehöll kompletterande material. När så varit fallet redovisas det nedan. Detsamma gäller sökning i databaserna Scopus och PubMed, för att se om det kunde finnas kompletterade litteratur i någon av dessa databaser (PubMed har dock en mer medicinsk inriktning). Google Scholar nyttjades vid tillfällen när artiklar ej var tillgängliga för nedladdning via universitetens databaser (Detta gav tillgång till en artikel som framstod som angelägen i sammanhanget). Sökvägar och så kallade söksträngar redovisas nedan. Sökning har genomförts vid olika tillfällen från augusti 2015 till oktober 2016. Angivna inklusionskriterier har varit desamma vid samtliga sökningar.

#### 11.1.1. Avsikt med litteratursökning

Avsikt med litteratursökningen har varit att hitta, jämföra och sammanställa riktlinjer för adaptation av tester samt relevant angränsande kunskapsbärande litteratur.

#### 11.1.2. Inklusionskriterier

Detta är inklusionskriterierna för litteratursökningen:

- Artiklar som granskats och betecknas som peer reviewed
- Böcker med bedömt relevant innehåll
- Årtal för publikation: 2000 och framåt. Där avsteg skett finns detta redovisat nedan.
- Språk: engelska eller svenska
- Doktorsavhandlingar som är nedladdningsbara
-



### 11.1.3. Exklusionskriterier

Artiklar som syftar till att validera specifika test på ett specifikt språk där tillvägagångssätt för testadaption med stöd i riktlinjer ej explicit adresseras.

### 11.1.4. Litteratursökningsperiod augusti 2015 - november 2016

Sökning genomfördes i databasen PsycInfo. Den första sökning präglades av att novist undersöka olika sökord. Sökorden förändrades allt eftersom jag genom ”trial and error” provade mig fram och via nyckelord som framkom i artiklar som gav träffar i sökningen.

Sökning utan någon begränsning i sökfältet:

*test adaption + psychometrics + guidelines*

Tolv (12) träffar. Samtliga artiklar icke relevanta. Innehöll artiklar inriktade på till exempel validering av enskilda tester. Jag skummade dock igenom enstaka nyckelord.

Jag valde därefter att söka på nyckelord (eng. *key words*):

*Psychometrics + test validity* = Två (2) icke relevanta träffar (Spanska artiklar).

Ny sökning på key words:

*psychometric assessment + theory and practise* = Sextiosju (67) träffar. En av dessa artiklar var intressant i sammanhanget:

Slaney, K. L., Tkatchouk, M., Gabriel, S. M., & Maraun, M. D. (2009). Psychometric assessment and reporting practices: Incongruence between theory and practice. *Journal of Psychoeducational Assessment*, Vol 27(6), pp. 465 – 476.

Slaney et als artikel och de riktlinjer de utvärderar avhandlar ej explicit översättning och anpassning av instrument och tester. Jag läste den dock och undersökte vilka artiklar de hänvisade till i referenslistan. Där hänvisade de till:

American Educational Research Association, American Psychological Association, National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Denna handboks utgivningsår var 1999. Jag undersökte om det utgivits någon reviderad version därefter. På så vis fann jag:

American Educational Research Association, American Psychological Association, National Council on Measurement in Education. (2014). *Standards for Educational and Psychological Testing*. Published by American Educational Research Association, US, Washington, DC.

Boken fanns ej i denna reviderade upplaga och köptes därför in till universitetsbiblioteket på Umeå Universitet. Under denna ”väntetid” fortsatte jag sökning via databaserna. Jag valde bort att undersöka tidigare utgåvor av handboken.

Fortsatt sökning skedde i fritt fält på:

*”translation of psychometric assessments + guidelines”*

Etthundraåttiosex (186) träffar. Närmast uteslutande icke relevanta artiklar. Flertalet artiklar utgjordes av fokus på validering av enstaka instrument på ett antal olika språk. En artikel framstod dock som intressant:

Nilsson Kajermo, K., Böe, H., Johansson, E., Henriksen, E., McCormack, B., Gustavsson, J. P., & Wallin, L. (2013). Swedish Translation, Adaptation and Psychometric Evaluation of the Context Assessment Index (CAI). *WVN* 2013; 10:41 – 50.

Artikeln var intressant eftersom den dels var skriven av svenska författare med relevant inriktning för denna uppsats och dels för att jag såg möjligheten att kontakta artikelns huvudförfattare för en intervju gällande deras anpassningsarbete.

Utifrån (nyckel)ord som jag fann i artikeltexter eller i abstract och ”keywords”, så genomfördes ny sökning i fritt fält:

*translation + psychological tests + guidelines*

En (1) träff som utgjordes av denna bok:

Hambleton, R. K., Merenda, P., & Spielberger, C. (Eds.). (2005). *Adapting educational and psychological tests for cross-cultural assessment*. Hillsdale, NJ: Lawrence S. Erlbaum Publishers.

Denna bok beställdes och har ingått i uppsatsens litteraturläsning.

Andra ord som jag mötte på i texterna och valde att fånga upp var ”test construction”. Via sökning i fritt sökfält på:

*”test constructions + recommendations + translation”*

Tvåhundraåttioen (281) st träffar. Efter en genomgång av de femtio (50) första artiklarna, där ingen av dessa framstod som relevant bytte jag till att söka på samma ord men i abstract (alla tre orden i sökfält abstract). Detta gav tre (3) träffar varav en mycket givande:

Kristjansson, E. A., Desrochers, A. & Zumbo, B. (2003). Translating and Adapting Measurement Instruments for Cross-Linguistic and Cross-Cultural Research: A Guide for Practitioners. *Canadian Journal of Nursing Research*, 35 (2), 127 – 142.

Artikeln var ej tillgänglig via PsycInfo eller PsycArtic men via Google Scholar.

Via träffen på *The Standards for Educational and Psychological Testing* (American Educational Research Association et al., 2014, se ovan) genomfördes sökning i sökfält för abstract på denna handboks titel. Det gav åtta (8) träffar, varav en intressant (Sireci och Faulkner-Bond, 2014) men denna artikel var ej tillgänglig för att laddas hem (ej heller via Google Scholar). Artikelns huvudförfattare, Stephen Sireci återkom dock i andra sökvägar och söksträngar (se längre ned).

Parallellt med sökningarna ovan fann jag hänvisningar till tidskriften *International Journal of testing* via artiklar i referenslistan i Kristjansson et al (2003, se ovan):

Hambleton, R. K. (2002). Adapting achievements tests into multiple languages for international assessments. In A. C Porter & A. Gamoran (Eds.), *Methodological advances in cross-national*

*surveys of educational achievement* (pp. 58 – 76). Washington: National Academy of Sciences.

Hambleton, R. K., & Patsula, L. (1998). Adapting tests for use in multiple languages and cultures (special issue on Validity Theory and the Methods Used in Validation, edited by B.D. Zumbo). *Social Indicators Research*, 45, 153 – 171.

Obs: Äldre publiceringsår utifrån inklusionskriterierna men artikelns titel och abstract verkade kunna bidra signifikant och leda mig vidare i mitt sökande.

Dessa källor vägledde mig till International test Commission (ITC). Genom sökning via Google på International Test Commission fann jag ITC:s hemsida samt länkar till deras tidskrift *International Journal of Testing*. Via hemsidan fann jag ITC:s riktlinjer för testadaptation:

International Test Commission. (2005). *International Guidelines on Test Adaptation*.

Via *International Journal of testing* fann jag flertalet artiklar som var relevanta för denna uppsats (se uppsatsens referenslista). Jag valde bort att undersöka den första utgåvan av riktlinjer som presenterades under 1999 (Hambleton, 2001).

Ytterligare en sökning innehöll sökning för *guidelines* i abstract och i fritt sökfält *psychological test + translation*. Detta resulterade i etthundrafemtiosex (156) träffar. Detta gav träffar på till exempel äldre riktlinjer från ITC och redan för mig kända artiklar skrivna av Roland K. Hambleton och kollegor till honom. En träff avsåg riktlinjer för psykologiska test:

Gudmundsson, E. (2009). Guidelines for adapting psychological instruments. *Nordic Psychology*, 61(2), pp. 29 – 45.

Dessa riktlinjer valdes bort då de inte hade ett tvärkulturellt perspektiv.

En (1) träff enligt ovan angiven sökning avsåg enskild studie av ett test och bedömning av översättningskvalitet:

Sireci, S. G., Yang, Y., Harter, J., Ehrlich, E. J. (2006). Evaluating Guidelines for Test Adaptions: A methodological Analysis of translation Quality. *Journal of Cross-Cultural Psychology* 37.5 (Sep 2006): 557 – 67.

Denna artikel analyserar ett specifikt test men den ledde mig till International Association for Cross-Cultural Psychology (IACCP) som ger ut *Journal of Cross-Cultural Psychology* som publicerar artiklar inom testanpassning.

Upprepad sökning skedde under senare delen av höstterminen 2016 för att eventuellt finna färskt publicerat material men det gav inga nya relevanta träffar.

I intervju med Hanne Berthelsen kom jag i kontakt med Beaton et als riktlinjer genom att Berthelsens team arbetat med en äldre utgåva av dessa riktlinjer (jag kände vagt till dessa innan):

Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *Journal of clinical epidemiology*. 1993;46(12):1417 – 32.

Dessa riktlinjer återfanns dock ej i databasen PsycInof men de fanns registrerade i databasen PubMed. Där fann jag även en nyare utgåva av riktlinjerna:

Beaton, D., Bombarider, C. &, Guillemin, F. (2000). Guidelines for the Process of Cross-Cultural Adaption of Self-Report Measures. *SPINE*, 25 (24), 3186 – 3191.

Efter ovan sökningar bedömde jag litteratursökningen som tillräckligt mättad för denna uppsats omfattning. Dock besökte jag ITC's hemsida under juli 2017, där det kom till min kännedom att en uppdaterad version av riktlinjerna publicerats. Jag valde att inkludera dessa i uppsatsen.

Nedan presenteras de ramverk och riktlinjer som ingår i uppsatsen.

### **11.1.5. Ramverk och riktlinjer**

De ramverk och riktlinjer för testadaption som slutligen kom att ingå i uppsatsen är:

American Educational Research Association, American Psychological Association, National Council on Measurement i Education. (2014). *Standards for Educational and Psychological Testing*. Publied by American Educational Research Association, US, Washington, DC.

Tjärnberg, Nalika (2017). Masteruppsats. *Den utopiska visionen: Etablering av ekvivalens i anpassning av psykologiska test*. Institutionen för filosofi, lingvistik och vetenskapsteori, Göteborgs Universitet.

International Test Commission. (2005 och 2017, inledningsvis 2005 och senare 2017). International Guidelines on Test Adaptation. Hämtad 15 juni 2016 respektive 14 juli 2017 från International test Commission: <https://www.intestcom.org>.

Kristjansson, E. A., Desrochers, A. & Zumbo, B. (2003). Translating and Adapting Measurement Instruments for Cross-Linguistic and Cross-Cultural Research: A Guide for Practitioners. *Canadian Journal of Nursing Research*, 35 (2), 127 – 142.

Beaton, D., Bombarider, C. & Guillemin, F. (2000). Guidelines for the Process of Cross-Cultural Adaption of Self-Report Measures. *SPINE*, 25 (24), 3186 – 3191.