

Måling i kvantitativ utdanningsforskning: Et instrumentelt mistak?

Sigve Høgheim

Høgskolen på Vestlandet, Norge

SAMMENDRAG

Til tross for at måling er en hyppig anvendt metode i kvantitativ utdanningsforskning, blir lite oppmerksomhet viet til logikken som ligger bak tallene. I denne artikkelen argumenterer jeg for at utdanningsforskning har adoptert psykologiens syn på måling sammen med en rekke vitenskapelige tankefeil. Artikkelen gir en gjennomgang av innholdet i de to overlappende tilnærmingene til måling som preger dagens metodologi: måling som tallfesting og måling som validering. Basert på en sammenligning med en klassisk forståelse av måling argumenteres det for at kvantitativ utdanningsforskning ikke måler, men kvantifiserer teoretiske begreper. Denne forståelsen av måling sår tvil om andre rådende oppfatninger i kvantitativ forskning, spesielt ontologiske og vitenskapelige antakelser. Basert på gjennomgangen av psykologisk måling argumenteres det for at utdanningsforskning ikke har tilknytning til ontologisk realisme, som ofte antas i metodelitteraturen. Jeg viser også til den utestede hypotesen om den iboende kvantitative strukturen til teoretiske begrep, som danner fundamentet for at noe i det hele tatt er målbart. Hovedtema i artikkelen er at «måling» i utdanningsforskning blir gjort utelukkende til en instrumentell oppgave forstått som instrumentutvikling, mens vitenskapelige oppgaver blir tatt for gitt.

Nøkkelord: måling; kvantitet; konstruktvaliditet; validering; metode

ABSTRACT

Measurement in Quantitative Educational Research: An Instrumental Mistake?

Although measurement is a frequently used method of data collection in educational research, little attention is devoted to the logic behind the obtained numbers. In this article, I argue that educational research has adopted its view of measurement from the field of psychology alongside several scientific fallacies. This article gives an account of two complementary approaches to measurement that characterize contemporary methodology: measurement as the assignment of numbers, and measurement as validation. Based on a comparison with a classic view on measurement, I argue that quantitative educational research does not measure, but *quantifies* concepts. Furthermore, I argue that this method of measurement influences other prevailing conceptions in quantitative educational research, particularly ontological and scientific assumptions. Based on the review of psychological measurement I argue that educational research has no connection to ontological realism, which is often assumed in methodological literature. Furthermore, I also highlight the

Korrespondanse: Sigve Høgheim, e-post: sigve.hogheim@hvl.no

© 2023 Sigve Høgheim. This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>), allowing third parties to copy and redistribute the material in any medium or format and to remix, transform, and build upon the material for any purpose, even commercially, provided the original work is properly cited and states its license.

Citation: Høgheim, S. (2023). Måling i kvantitativ utdanningsforskning: Et instrumentelt mistak?. Nordisk tidsskrift for pedagogikk og kritikk, 9, 144–159. <http://doi.org/10.23865/ntp.v9.5658>

untested hypothesis concerning the quantitative structure and measurability of theoretical concepts. The main theme of this article is that “measurement” in educational research is solely based on instrumental tasks, understood as instrument development, while scientific tasks are taken for granted.

Keywords: *measurement; quantity; construct validity; validation; method*

Mottatt: Mai, 2023; Antatt: Mai, 2023; Publisert: August, 2023

Introduksjon

Hva vil det si å måle noe? Mange vil trolig svare «å finne ut hvor mye det er av noe», tilsvarende det vi gjør når vi måler fysiske objekter. Når vi måler et objekt tar vi utgangspunkt i et attributt, slik som lengde eller masse, og finner ut mengden av attributtet i henholdsvis meter eller kilogram. Dette omtales som et klassisk perspektiv på måling (Michell, 1990). Dersom en forsker innen utdanningsfeltet svarer på spørsmålet om måling kan vi få et svar som verken handler om attributt eller mengde. Det er ikke fordi forskeren måler fysiske objekt på en annen måte, men i studier av menneskers psykologiske prosesser brukes en fundamentalt annerledes forståelse av måling. Så ulik er metoden at det åpner for spørsmålet: *Måler egentlig kvantitativ utdanningsforskning noe som helst?*

Måling har en sentral plass i utdanningsfeltet. Målinger gjennomsyrrer skolediskursen (se Meld. St. 22 (2010–2011); Meld. St. 28 (2015–2016)), hvor internasjonale undersøkelser (Gabrielsen et al., 2017; Jensen et al., 2019; Nilsen & Kaarstein, 2021), nasjonale tilstandsrapporter (Wendelborg et al., 2020) og metaanalyser (Hattie, 2009) retter fokus mot utfordringer i skolen og mulige løsninger. Ut over slike undersøkelser finnes det en rekke kvantitative studier rettet mot utdanningssektoren med hensikt å måle psykologiske egenskaper. Kritikken av de internasjonale studiene har ofte rettet søkelyset mot metodiske svakheter knyttet til årsakssammenhenger (f.eks. Melby-Lervåg & Lervåg, 2013; Sjøberg, 2014) og de praktiske følgene av slike undersøkelser. Det er derfor ikke overraskende at bruken av *randomiserte kontrollerte forsøk* (RCT), som anses som mest egnet for studiet av kausale relasjoner (Maxwell, 2004; Shadish et al., 2002), har hatt stor vekst de siste tiårene (Pontoppidan et al., 2018). Det blir som regel tatt for gitt at måling, som ligger til grunn for all kvantitativ forskning, er en gyldig metode.

Hensikten med denne artikkelen er å løfte fram metodologien bak måling i kvantitativ utdanningsforskning for å synliggjøre logikken bak tallene som forskning bringer fram. I denne artikkelen argumenterer jeg for at pedagogikkfeltet har adoptert psykologiens perspektiv på måling, men også dette perspektivets tankefeil – som i ytterste konsekvens kan så tvil om vitenskapeligheten til kvantitativ utdanningsforskning. Ifølge Michell (1997) har kvantitativ forskning to oppgaver knyttet til målinger: en *vitenskapelig* og en *instrumentell*. Den vitenskapelige oppgaven handler om å teste de underliggende antakelsene ved måling, slik som at det som måles faktisk er målbart.

Den instrumentelle oppgaven viser til det å gjøre noe målbart med instrumenter, for eksempel med observasjons- eller spørreskjema. I denne artikkelen argumenterer jeg for at den vitenskapelige oppgaven blir tatt for gitt i pedagogisk forskning som sammen med en underspesifisert instrumentell oppgave utgjør et mistak i måling, særlig ved operasjonaliserte psykologiske begrep. Jeg begynner framstillingen av måling med først å vise til en klassisk forståelse av måling, som en kontrast til det psykologiske og pedagogiske synet på metoden. I resten av artikkelen bruker jeg betegnelsen «psykologisk måling» også for måling i pedagogisk forskning. Til slutt vil jeg kontrastere perspektivet på måling mot vanlige antakelser i kvantitativ forskningsmetodologi som preger pedagogikk, særlig det som gjelder ontologi og epistemologi.

Det klassiske perspektivet

Det klassiske perspektivet på måling kan knyttes til måten metoden blir forstått i naturvitenskapene, med røtter i antikken og Euklids verker (se Michell, 1999, for detaljert historikk) og ytterligere presisert i blant annet Otto Hölders aksiomer for kontinuitet fra 1901 (engelsk oversettelse av Michell & Ernst, 1997). Som en vitenskapelig metode blir måling forstått som å oppdage verdien av en mengde som et objekt innehar (Michell, 2001), hvor mengde blir forstått som et gitt tilfelle av et attributt (for eksempel lengden av person x). Måling defineres som en numerisk relasjon (ratio) mellom en mengde og et gitt nivå fra samme attributt (Michell, 1990), uttrykt som $X = rY$. X er mengden som skal oppdages, Y er et gitt nivå av attributtet som skal oppdages og r er forholdstallet. Når vi måler en persons lengde (X) kan vi bruke en standard meter som et gitt nivå av attributtet, som da blir en *enhet*, for å oppdage forholdet mellom lengde ved personen og r antall meter.

En forutsetning for måling i et klassisk perspektiv er *kvantiteter*, som er kontinuerlige variabler som tillater oppdagelsen av numeriske forholdstall. En variabel viser til gruppen av mulige verdier et attributt kan inneha, og «kontinuerlig» referer til at verdiene teoretisk sett er uendelige og udelelige ved at det alltid vil finnes en mengde c mellom to andre, a og b . For et spesifikt objekt er variabel noe som det innehar singulært, altså for eksempel én lengde eller masse av gangen. Dersom vi har verdien fra flere objekter på samme attributt, kan relasjonen mellom disse brukes for å demonstrere om et attributt er kvantitativt etter Hölders aksiomer for kontinuitet (se Michell, 1990). De definerende egenskapene til en kvantitet er *orden* og *additiv* struktur. Orden viser til at mengdene kan rangeres etter størrelse (f.eks. $a > b > c$), og additiv referer til en struktur hvor mengder står i en relasjon som kan uttrykkes med addisjon, uttrykt som $a + b = c$, som gir en kontinuerlig sammensetning. Relasjonen er ikke adderende i en ren matematisk forstand, da $a + b = c$ viser at c er en mengde som er sammensatt av mengdene a og b . Se appendiks 1 for aksiomene for orden og additiv samt en ikke-teknisk beskrivelse av disse.

Ratio danner utgangspunktet for måling, som også er tilfelle der hvor det brukes intervallmålinger. Et eksempel er den *intensive* kvantiteten temperatur, som i

motsetning til *ekstensiv* kvantiteter som lengde, ikke kan direkte sammenkobles. En blanding av to væsker med temperatur x og y gir ikke $x + y = z$ på samme måte som lengder. Intervall viser til måling gjort med utgangspunkt i forskjeller mellom objekter på et attributt, som krever både orden og addisjon mellom forskjellene. For temperatur viser dette til forskjeller i endring av volum, for eksempel ved gamle kvikksølvtermometer, eller spenning, som i nyere elektriske termometer. Siden både volum og spenning er ratio blir intervallmål også et forholdsmål, men basert på forskjeller.

Det klassiske synet på måling forutsetter ontologisk realisme (Wolff, 2020), forstått som antakelsen om en objektiv virkelighet som er uavhengig av den som observerer (Borsboom, 2005). Siden måling omhandler relasjoner mellom mengder antas det en eksistens av kontinuerlige strukturer og reelle tall i form av relasjonen mellom dem. Wolff (2020) summerer de ontologiske og epistemologiske antakelsene som *restriktiv realisme*: kvantitative attributter har en uavhengig eksistens (realisme), men ikke alle attributter er kvantitative (restriktiv). En forlengelse av det restriktive prinsippet er at det ikke finnes en logisk nødvendighet for at et attributt er en kvantitet. Antakelsen om kvantitet er en testbar hypotese som uten empirisk støtte forblir spekulasjon. Det som også verdt å merke seg, er at realisme i denne sammenheng viser til eksistensen av kvantitative *attributter*, ikke «kvantitative» objekter. Hvis det er en person som måles, antas det ikke en objektiv eksistens av «person». Det er attributter ved personen som antas å være objektivt eksisterende og universelle (Michell, 1999), slik som lengde, masse og temperatur.

Psykologisk måling

Måling i pedagogisk og psykologisk metodologi representerer et brudd med det klassiske perspektivet, uten at det endrer bruken av betegnelsen måling. Psykologisk måling tar ofte utgangspunkt i abstrakte og ikke-observerbare egenskaper (Kleven, 2002), slik som motivasjon og evner, som antas å gi egne utfordringer ved måling. Som Kleven (2008) beskriver: «Anyways, the heart of the measurement problem is not a problem of numbers [...] A more important problem is: How well is the concept represented by the indicators?» (s. 224). Sitatet fra Kleven (2008) er en standard posisjon i pedagogisk forskning: utfordringen ved måling handler å fange det som undersøkes i instrumentet (m.a.o. *indicators*). Men hvordan kan en metode for å verdien av mengder ikke ha utfordringer knyttet til tall? Trolig ligger svaret i framveksten av to komplementære syn på måling som har preget psykologisk metodologi som sammen har gjort tall sekundære for metoden.

Måling som tallfesting

Den mest sentrale bidragsyteren for psykologisk måling er Stanley Smith Stevens (1946), som utformet en målingsteori i kjølvannet av sterk kritikk mot psykologiens metoder (Ferguson, 1932). Løsningen for Stevens' del var en målingsteori

sammensatt av to perspektiv modifisert som et forsvar av psykologisk måling, hvor to bidrag fremdeles preger forskningsmetodologien.

Det første bidraget er Stevens' (1946) definisjon av måling som *regelstyrt tildeling av tall til objekter eller hendelser* (se s. 677). Dette har blitt den dominerende definisjonen på måling i psykologi (se Borsboom, 2005; Michell, 1999) og er et etablert perspektiv i pedagogisk forskningslitteratur (f.eks. Høgheim, 2020; Kleven et al., 2011; Lund, 2001; Ringdal, 2018). Definisjonen henter Stevens eksplisitt fra *det representasjonalistiske perspektivet* (DPR) på måling (Campbell, 1920), hvor tall anses som en symbolsk logikk som kan brukes for å representere relasjoner mellom objekter etter definert premiss, framfor noe som er iboende i realiteten som kan oppdages (Suppes & Zinnes, 1963). Framfor å definere premisser for psykologisk måling bruker Stevens heller det mer diffuse prinsippet om *regelstyring*. Reglene blir aldri tydelig definert, men kommer til uttrykk i Stevens' (1946, 1958) andre sentrale bidrag til målingslitteraturen: *skalaer* (eller *målenivåer*). Stevens definerte den kjente, hierarkiske fire-delingen av nominal-, ordinal-, intervall- og rationivå (se også Stevens, 1951), som blir «reglene» i tildelingen av tall. Uten å gå i detalj er det verdt å merke seg at Stevens (1946) her utvider synet på måling ved å inkludere nominal- (kategorisering) og ordinalnivå (rangering) i tillegg til ratio og intervall. Reglene for de ulike nivåene finner Stevens i *operasjonalismen*, hvor grunnprinsippet er at et begrep er synonymt med måten det blir identifisert (Bridgman, 1927). Hos Stevens (1946) kommer operasjonalismen til uttrykk i forskerens valg om hvordan objekter skal tallfestes: «the type of scale achieved depends upon the character of the basic empirical operations performed» (s. 677).

Når et begrep blir forstått som prosedyrene for å identifisere det (operasjonalisme) og det er forskerens valg som avgjør tallrepresentasjon (skalaer), blir DPR modifisert i Stevens' teori ved at det er forskeren som også definerer relasjonene mellom begreper (se også Michell, 1997). Premissløs DPR sammen med operasjonalisme gjør den vitenskapelige oppgaven ved måling overflødig, da det er forsker som definerer hva som måles og hvordan det skal måles.

Måling som validering

Forskning på psykologiske begreper krever at disse operasjonaliseres til indikatorer i et instrument. Stevens' målingsteori, annet enn med operasjonalisme, tar lite hensyn til spørsmål om hva som måles (Zeller, 1997), som er sentralt i studier av abstrakte, uobserverbare egenskaper. Spørsmålet om hva som måles behandles i litteraturen om *konstruktvaliditet*.

Konstruktvaliditet har noe ulike betydninger (se Borsboom, 2005): spørsmålet om en test måler det den har til hensikt å måle (Kelley, 1927), om testskårer korresponderer til teoretiske relasjoner i nomologiske nettverk (et sett lover som knytter teoretiske begreper sammen i et nettverk: Cronbach & Meehl, 1955), om tolkninger og bruk av testskårer er berettiget (Messick, 1995). Parallelt blir validitet definert som «the approximate truth of an inference» (Shadish et al., 2002, s. 34), hvor

konstruktvaliditet knyttes til spørsmålet om det er (grad av) sannhet i slutninger om et konstrukt basert på antakelsen om at «instrument X måler konstrukt A» (se Kleven, 2008). Innen valideringslitteraturen blir en ny definisjon på måling foreslått som: «the process of linking concepts to indicants» (Zeller, 1997, s. 823), samtidig som det antas at Stevens' målingsteori ligger til grunn for de empiriske operasjonene ved måling. Måling som validering kan derfor ikke vurderes løsrevet fra Stevens' målingsteori, men som en nødvendig forlengelse for å kunne si noe om *hva* som måles utover operasjonalisme.

Psykologiske egenskaper er komplekse, og løsningen i testing blir ofte å operasjonalisere én egenskap med flere indikatorer, for så analysere indikatorene for å si noe om mulige underliggende konstrukt. En sentral tanke i validering ser ut til å henge sammen med *klassisk test-teori* (CCT: Lord & Novick, 1968), hvor hver person antas å ha en sann skår (T) som bare kan bli gjort kjent gjennom konkrete observasjoner (O) som alltid vil være preget av feil (E). T er en teoretisk skår definert som forventet O over replikasjoner, men som i enkelttilfeller er ukjent. Tanken som tilsynelatende preger validitetstenkning er at vi kan komme nærmere konstruktet T representerer ved å eliminere feil i observasjon, altså å gjøre avstanden fra T til O så liten som mulig ved å minske E. Feil handler om både tilfeldige feil, flyktige variasjoner (f.eks. dagsform, tidspunkt) som antas å elimineres med økende antall mål basert på sannsynlighet, men også systematiske feil som inntreffer ved hver testing, slik som i operasjonaliseringer (Kleven, 2008). Målet er å redusere systematiske feil, slik som korrespondanse mellom operasjonalisering og teoretisk begrep (innholdsvaliditet: se f.eks. Cronbach & Meehl, 1955), og demonstrere dette i nomologiske nettverksanalyser om faktorstruktur, relasjoner innad og til andre instrumenter samt begrepers «atferd» (kriteri-ektivitet, konvergent og diskriminant validitet: f.eks. Campbell & Fiske, 1959; Zeller, 1997), og vurdering av studieeffekter (se også Shadish et al., 2002). Validitet blir koblet til arbeidet med validering (Messick, 1995), hvor data må berettige antakelsen om underliggende konstrukt basert en teoretisk nettverksmodell.

Innenfor måling som validering spiller antakelsen om *latente variabler* en stor rolle som analytisk redskap, enten det er gjennom eksplorerende og konfirmerende faktoranalyser eller strukturelle ligningsmodelleringer. Selv om det finnes mange forståelser av hva en latent variabel egentlig benevner, som «hypotetiske konstrukter» eller «umålbare, uobserverbare variabler» (se Bollen, 2002), er et generelt felles kjennetegn antakelsen om at noe uobserverbart kan estimeres basert på variasjoner i testskårer.

Drøfting

Jeg begynner med å svare på det innledende spørsmålet: *Måler* kvantitativ utdanningsforskning noe? Psykologisk måling representerer et brudd med klassiske perspektiver og tar ikke sikte på å finne ut noe om et studieobjekt etter definerte premisser. Psykologisk måling studerer relasjoner mellom operasjonaliseringer basert på forskerens valg om tallfesting og fortolkning av data. Som en konsekvens kan

ikke mål fra kvantitativ utdanningsforskning fortelle oss noe ut over relasjonen mellom instrument, som med Stevens' åpne definisjon på måling ofte ikke overgår den enkelte studiens rammer. Til tross for forskningslitteraturens gjentakende bruk av Stevens' (1946) rammeverk og valideringslitteraturens fokus på konstrukt, er ikke praksisen som beskrives måling. Som Barrett (2003) beskriver: «it is *applied numerics*, not quantitative measurement» (s. 12). Framfor å bruke begrepet «måling» om metoden foreslår jeg at psykologisk måling kan omtales som *relasjonell koding*, for å fange DPRs fokus på relasjoner, men med en avstand til det klassiske perspektivet på måling (koding). Men har mangelen på måling og kvantitet egentlig noe å si for forskningsfeltet? Og er kvantiteter oppnåelig? Jeg vil videre rette oppmerksomheten mot noen konsekvenser av mangelen på måling, særlig mot ontologiske og epistemologiske antakelser i utdanningsforskning.

Implisitt realisme

Måling forutsetter realisme; at det finnes noe som kan måles. Dette antas kjent i utdanningsforskning, da mye av litteraturen om konstruktvaliditet baseres på realisme (Shadish et al., 2002), særlig *kritisk realisme* (f.eks. Kleven, 2008; Lund, 2005; Lund et al., 2006). Kritisk realisme kan oppsummeres som ontologisk realisme, epistemologisk relativisme og forkasting av *bedømmende relativisme* (Bhaskar, 1979). De epistemologiske aspektene viser til at kunnskap vurderes som foranderlig selv om alle oppfatninger ikke er tilsvarende valide – det finnes grunnlag for å foretrekke en oppfatning over en annen. Til tross for posisjoneringen, er det lite ved den vitenskapelige oppgaven i psykologisk måling som har realistiske tilknytninger. Operasjonalisme er en måte å komme unna metafysiske spørsmål ved å definere konsepter som handling, mens betingelsesløs representasjonalisme løsrives fra virkeligheten ved å fjerne forbehold i tildeling av tall.

Konstruktvaliditet bidrar tilsvarende lite til realisme i den instrumentelle oppgaven ved måling. Validitet antas å være en funksjon av sannhet (Shadish et al., 2002), samtidig som validitet brukes synonymt med validering (Borsboom et al., 2004), som igjen er en funksjon av testskårer (Kane, 2009; Messick, 1995). Disse to perspektivene sammenfaller ikke: Sannhet handler om ontologiske kriterier (Bridges, 1999), mens evidens handler om det observerte. Det er tilsynelatende testskår-tolkningen som er dominerende i validitetslitteraturen (se Borsboom et al., 2009), motstridende kritisk realisme, hvor: «To be is *not* to be the value of a variable; though it is plausible (if, I would argue incorrect) to suppose that things can only be *known* as such» (Bhaskar, 2008, s. 29). Testskår-tolkning ser vekk fra spørsmålet om hva som eksisterer til fordel for hva som kan tolkes ut fra det foreliggende.

Det mest problematiske begrepet innenfor konstruktvalidering er «konstrukt», som forstås både som (1) et teoretisk begrep, et språklig symbol for noe (Kane, 2012), og (2) en egenskap som antas å spille en rolle i en psykologisk realitet (Shadish et al., 2002). (1) kan ikke måles eller oppdages i data fordi det er en språklig representasjon (Maraun & Halpin, 2008). (2) kan måles, men kan ikke defineres teoretisk da det er

et reelt fenomen (Borsboom et al., 2009). Det er (1) forståelsen som tilsynelatende er dominant i validitetslitteraturen til tross for antakelsen om ontologisk realisme (Cronbach & Meehl, 1955; Messick, 1995). Hos Cronbach og Meehl (1955), et av de mest innflytelsesrike perspektivene, blir betydningen av et konstrukt «set forth by stating the laws in which they occur» (s. 294; i nomologiske nettverk), samtidig som at det erkjennes at disse lovene er ukjente, eller i beste fall på vei til å bli oppdaget. Dette åpner for vage nettverk om språklige representasjoner av studieobjektene (konstrukt) som gir en begrensning hvor enhver tolkning av testskårer kan være valid gitt teori-koherens framfor ontologisk realisme. Som Bhaskar (2008) poengterer: «For if to be were just to be the value of a variable we could never make sense of the complex process of identification and measurement» (s. 29).

Borsboom et al. (2004) foreslår et alternativt perspektiv hvor konstruktvaliditet knyttes til den kausale sammenhengen mellom et konstrukt (2) og et instrument. Konstruktvaliditet som en kausal fortolkning fra konstrukt til instrument krever en teoretisk forutsetning som sier hvorfor og hvordan et gitt konstrukt skal påvirke de enkelte indikatorene i et instrument. Dette vil jeg argumentere for at er i tråd med kritisk realisme og Bhaskars (2008) antakelse om stratifisert virkelighet hvor reelle strukturer og mekanismer (det reelle) påvirker hendelser (det faktiske) og erfaringer (det som observeres). Bhaskars (2008) prinsipp om retroduktive slutninger, «inferences from present effects to prior (perhaps hidden, perhaps just unrecorded) causes» (s. 125), og transcendentale spørsmål kan åpne for antakelsen om reelle egenskaper som påvirker instrument. Transcendentale spørsmål, som Bhaskar (1979) ikke knyttet direkte til måling, kan i denne sammenheng være: Hva må vi anta eksisterer for at instrumentet vårt skal fange opp noe? Forutsetningen må være at det finnes en psykologisk egenskap som «listens to its name [...] and that actually *does* steer the measurement outcome in one or the other direction» (Borsboom et al., 2009, s. 150), som innenfor kritisk realisme kan baseres på retroduktiv resonnering om skjulte og antatt reelle mekanismer og strukturer. Og her kan Zellers (1997) definisjon på måling, som lenken mellom indikator og konstrukt, være passende for validitet.

Psykologisk måling har unngått tilknytning til (kritisk) realisme, både med Stevens' målingsteori og det generelle rammeverket for konstruktvaliditet. Jeg vil argumentere for at psykologisk måling baseres i stor grad på en logisk positivistisk metafysikk med utgangspunkt i relativisme (Michell, 2021) med et konsensusbasert sannhetskriterium for validering (Borsboom, 2005). Dersom vi ser på validitet som sannhet, er hovedproblemet at psykologisk måling antar måling. Det som kan og bør valideres, er om et instrument *måler* noe samtidig som det drøftes hva dette *noe* kan være. Konstruktvaliditet er derfor ikke uten betydning i psykologisk måling, men da bør en kausal forståelse av validering ligge til grunn for at *psykologisk* realisme skal kunne antas og for å spesifisere kriteriene i prosessen. En kausal forståelse av konstruktvaliditet kan alene gi tilknytning til kritisk realisme, men spørsmålet om måling står fremdeles åpent.

Psykologiske kvantiteter

«Whatever exists at all exists in some amount. To know it thoroughly involves knowing its quantity as well as its quality» (Thorndike, 1918, s. 16). Selv om sitatet fra Thorndike nærmest har blitt kanon, er det ingen logisk selvfølge at noe som tallfestes er en kvantitet. Det er en hypotese som sier noe om det vi antar eksisterer som kan være falsk (Michell, 1990). Måling i pedagogisk forskning er med dagens praksis en metode med hundre prosent suksessrate – måling finner sted selv om man kan under- eller overrepresentere konstrukt med instrumenter (Messick, 1995; Shadish et al., 2002).

Michell (1990, 2008) argumenterer for bruken av *additive conjoint measurement* (ACM) for å teste «den kvantitative hypotesen», basert på arbeidet til Luce og Tukey (1964). I korte trekk er ACM en teori hvor Y relateres til A og B og beskriver betingelsene som er nødvendige for at (1) variablene er kvantiteter og (2) Y er en funksjon av A og B ($Y = f(A, B)$) (Michell, 1990). Ingen av variablene trenger å være kjente kvantiteter på forhånd og kan være på ulike målenivåer etter Stevens' hierarki. ACM er basert på betingelsene (dobbel) kansellering, Arkimedes' betingelse og løsbarehet. For en teknisk beskrivelse av ACM, se Michell (1990, s. 70–77) eller Boorsboom (2005, s. 93–95). Sett for eksempel at deltakere må bedømme et sett matematikkoppgaver, hvor oppgavene er sammensatt av en kombinasjon (a, b) av (a) kontekstuell vanskelighetsgrad og (b) regnetekniske krav til løsning, hvorvidt forventning om suksess (Y) er høyere (\geq) enn en oppgave satt sammen av en annen kombinasjon. Det betyr at deltakerne tar stilling til flere A (a_1, a_2, a_3) og B (b_1, b_2, b_3) opp mot Y . Parene av A og B kan stilles opp i en matrise utledet fra Y , slik som i tabell 1.

Tabell 1 Eksempel på matrise for uavhengige variabler

| | | Uavhengig A | | |
|-------------|-------|--------------|--------------|--------------|
| | | a_1 | a_2 | a_3 |
| Uavhengig B | b_1 | (a_1, b_1) | (a_2, b_1) | (a_3, b_1) |
| | b_2 | (a_1, b_2) | (a_2, b_2) | (a_3, b_2) |
| | b_3 | (a_1, b_3) | (a_2, b_3) | (a_3, b_3) |

Alle tre betingelsene må være innfridd for å demonstrere kvantitet, men grunnet plassbegrensning illustreres ACM her med betingelsen dobbel kansellering, som er en testbar konsekvens av additiv struktur. Dobbelt kansellering forutsetter at hver (a, b) kan representeres med den additive funksjonen: $f(a) + g(b)$. I ethvert verdipar (a, b) må det være en \geq relasjon til at annet verdipar basert på Y . Eksempelvis, hvis $(a_2, b_1) \geq (a_1, b_2)$, så $a_2 + b_1 \geq a_1 + b_2$. Hvis dette stemmer, og det samme er gjeldende for, for eksempel, $(a_3, b_2) \geq (a_2, b_3)$, så følger det at dersom:

- (1) $a_2 + b_1 \geq a_1 + b_2$, og
- (2) $a_3 + b_2 \geq a_2 + b_3$, så
- (3) $a_2 + b_1 + a_3 + b_2 \geq a_1 + b_2 + a_2 + b_3$, som ved eliminering av fellesledd blir
- (4) $b_1 + a_3 \geq a_1 + b_3$, som blir $(a_3, b_1) \geq (a_1, b_3)$

Antakelsen som testes er at dersom $(a_2, b_1) \geq (a_1, b_2)$ og $(a_3, b_2) \geq (a_2, b_3)$ på Y, så følger det at $(a_3, b_1) \geq (a_1, b_3)$. Dette kalles dobbel kansellering siden to ulikheter kanselleres med et tredje verdipar. Dersom de resterende betingelsene er møtt – kansellering, Arkimedes' betingelse og løsbarehet – er det nødvendig med én test av dobbel kansellering (Michell, 1988). Tanken med ACM er å oppnå en enhet basert på A og B, hvor en gitt endring i A (f.eks. $a_1 - a_2$) kan brukes som en enhet for B: enhver endring i B som kan kansellere endringen i A vurderes som tilsvarende stor. Og tilsvarende for B som enhet for A. Dette gir en uendelig sekvens av par (a_p, b_p) , hvor p er \pm et naturlig tall, der hvert neste par går opp/ned en enhet i A og en enhet ned/opp i B. Psykologiske kvantiteter blir her å forstå som intensive kvantiteter basert på forskjeller i de uavhengige variablene.

Til tross for oppslutning om bruk av ACM for å teste den kvantitative hypotesen (f.eks. Barrett, 2003; Cooper, 2019; Kyngdon, 2008), er forskning på området sparsom. ACM er heller ikke uten kritikere, hvor blant annet Sijtsma (2012) argumenterer for at ACM ikke kan ta høyde for kompleksiteten i målefeil i psykologisk metodologi. Målefeil er et sentralt moment, men det er viktig å ta høyde for at ACM ikke er en teori om testskårer, slik som CCT, men en teori om strukturer til attributt (Michell, 2014). Dette betyr ikke at måling antas feilfri, da all observasjon har feilkilder, men det er ikke inkorporert som et sannsynlighetsestimat i teorien. Av denne grunn vil konstruktvaliditet fremdeles være relevant, men da i en kausal forstand etter Borsboom et al. (2004). Samtidig løfter Sijtsma (2012) fram et viktig poeng om at betingelsene i ACM er såpass stringente at det er vanskelig, om i det hele tatt mulig, å innfri alle. Gitt få alternativ til ACM for den vitenskapelige oppgaven ved måling kan det tenkes en realitet hvor metoden ikke er mulig i psykologisk forskning.

Trenger utdanningsforskning kvantiteter?

Utdanningsforskning trenger kvantiteter *dersom* det antas at måling finner sted, men ikke ellers. Begrepsbruk om metode er ikke et uvesentlig poeng siden, som Bhaskar (1979) påpeker, samfunnsvitenskapene ikke bare handler om fagstoff, men også er for et publikum som blir en mulig kilde til (mis)forståelse og (des)informasjon. Ved å omtale koding som måling, på lik linje med et klassisk perspektiv, åpner utdanningsforskning opp for misoppfatninger om anvendt metode og mulighetene i utfallene. Dette blir et aspekt ved den *vitenskapelige adekvatheten* i beskrivninger som brukes i utdanningsforskning, i Bhaskars termer.

Et annet element i denne diskusjonen er dataanalyse. Parametriske tester med bruk av gjennomsnitt og standardavvik forutsetter kontinuerlige skalaer (intervall og ratio), men det meste av tallfesting i utdanningsforskning baseres på kategoriske skalaer (nominal og ordinal), slik som antall oppgaver løst eller Likert-skalaer (Jebb et al., 2021). Kategoriske skalaer har verken gjennomsnitt eller standardavvik, noe som løses ved å lage summerte skårer som *simulerte* intervallskalaer. Dette er en kjent utfordring, men Stevens (1946) forsvarer praksisen med: «In numerous instances

it leads to fruitful results» (s. 679). Det er likevel ingen logisk argumentasjon for at kategoriske data kan konverteres til en kontinuerlig skala. I utgangspunktet betyr dette at mye av kvantitativ analyse med gjennomsnitt, standardavvik og påfølgende p-verdier er basert på spekulative mål fra parametere som i utgangspunktet er nonsens, altså ikke-meningsbærende. Og her er det to alternativer: (1) gjøre det vitenskapelige arbeidet tilknyttet måling, eller (2) basere seg på egnede alternativer til parametrisk analyse, slik som *observation oriented modeling* (Grice, 2011), som er en ikke-metrisk analyse av talldata.

Så lenge det antas måling og kontinuerlige variabler, trenger utdanningsforskning kvantiteter. Situasjonen endrer seg dersom man slutter å anta måling, men heller bruker mer passende termer (f.eks. koding) med adekvate antakelser og praksis. Det betyr ikke at utdanningsforskning ikke hadde vært tjent med kvantiteter, både for teoriutvikling og sammenligning mellom studier. I utvikling av teorier er måling en fordel, i det minste en tilknytning til realisme. Slik rammeverkene står er forskningsfeltet sårbar for teoretisk underutvikling (Maraun, 1998). Dersom et instrument X lages ut fra konstrukt A og en forventet sammenheng mellom X og Y ikke observeres, kan man konkludere med: (1) X ikke representerer A, eller (2): teorien om A må revideres. Siden valideringslitteraturen forutsetter tolkning av data mot teoretiske nettverk kan det argumenteres med (1), selv om (2) kanskje er gjeldende. På denne måten kan teorier forbli statiske siden data må stemme overens med det teoretiske rammeverket, ikke bare instrumentet og konstrukt. I beste fall kan teorier utvides med valideringsperspektivet. Med en kausal fortolkning av validering og validitet i form av kvantiteter, løsrives forsker fra teoretiske nettverk for å kunne trekke konklusjoner om konstrukt. En utfordring med de simulerte kvantitetene er at forskjeller mellom tall ikke er «oversettbar» til størrelser. For hva betyr, for eksempel, forskjellen mellom 3,1 og 3,7 på en «intervallskala» fra 1 til 5? For vurdering av størrelser og sammenligning mellom studier er utdanningsforskning avhengig av metaanalytiske tilnærminger med effektstørrelser for å korrigere data for å avgjøre om tall er «store» eller «små», som igjen må baseres på subjektive overveielser.

Avslutning

Argumentasjonen i denne artikkelen baserer seg ikke på at pedagogisk forskning må følge naturvitenskapene for å være vitenskapelig. Bruk av ordet 'måling' ser likevel ut til å emulere naturvitenskapene og dens forutsetninger, men tydeliggjør at utdanningsforskning ikke driver med måling – det er teoretisk mulig, men det trengs grunnforskning. Et forsvar for dagens metode er at den er praktisk anvendelig og kan gi innsikt i relasjon mellom konstrukt, men da må det blant annet tas høyde for at

- (1) tall ofte behandles som kvantiteter, uten at de er det
- (2) konstrukt hevdes å være basert på realisme, uten at de behandles slik
- (3) relasjoner som undersøkes er basert på teoretiske lover som ikke er oppdaget

Jeg vil hevde at psykologisk måling representerer et instrumentelt mistak ved å emulere naturvitenskapenes metode uten de metodologiske prinsippene, som verken er nyttig eller vitenskapelig. Måling blir gjort til en instrumentell oppgave, og mistaket viser til en praksis som går imot egne underliggende antakelser med redusert mulighet til oppdagelse. Dersom kvantiteter ikke er mulig, er et alternativ å definere en egen praksis for psykologisk «måling» både i begrepsbruk, tallfesting, kausal validering og behandling av data for å skape egenart i metodologien.

Om forfatteren

Sigve Høgheim er førsteamanuensis i pedagogikk ved Institutt for pedagogikk, religion og samfunnsfag, Høgskolen på Vestlandet, campus Sogndal. Forskningsinteressene hans er knyttet til motivasjon i grunnskolen og grunnskolelærerutdanningen, samt til vitenskapsfilosofi og psykometri.

Referanser

- Barrett, P. (2003). Beyond psychometrics: Measurement, non-quantitative structure, and applied numerics. *Journal of Managerial Psychology*, 18(5), 421–439. <https://doi.org/10.1108/02683940310484026>
- Bhaskar, R. (1979). *The possibility of naturalism. A philosophical critique of the contemporary human sciences* (3. utg.). Routledge.
- Bhaskar, R. (2008). *A realist theory of science*. Routledge.
- Bollen, K. A. (2002). Latent variables in psychology and the social sciences. *Annual Review of Psychology*, 53(1), 605–634. <https://doi.org/10.1146/annurev.psych.53.100901.135239>
- Borsboom, D. (2005). *Measuring the mind*. Cambridge University Press.
- Borsboom, D., Cramer, A. O. J., Kievit, R. A., Scholten, A. Z. & Franic, S. (2009). The end of construct validity. I R. W. Lissitz (Red.), *The concept of validity: Revisions, new directions, and applications* (s. 135–170). Information Age Publishing.
- Borsboom, D. & Mellenbergh, G. J. (2004). Why psychometrics is not pathological: A comment on Michell. *Theory & Psychology*, 14(1), 105–120. <https://doi.org/10.1177/0959354304040200>
- Bridges, D. (1999). Educational research: Pursuit of truth or flight into fancy? *British Educational Research Journal*, 25(5), 597–616. <https://doi.org/10.1080/0141192990250503>
- Bridgman, P. W. (1927). *The logic of modern physics*. MacMillan.
- Campbell, D. T. & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56(2), 81–105. <https://doi.org/10.1037/h0046016>
- Campbell, N. R. (1920). *Physics: The elements*. Cambridge University Press.
- Cooper, C. (2019). *Psychological testing. Theory and practice*. Routledge.
- Cronbach, L. J. & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4), 281–302. <https://doi.org/10.1037/h0040957>
- Ferguson, A. (1932). Quantitative estimates of sensory events. *Nature*, 130(3291), 810–810. <https://doi.org/10.1038/130810a0>
- Gabrielsen, E., Hovig, J., Rongved, E., Strand, O., Støle, H. & Toft, T. E. (2017). *Godt nytt! Norske resultater fra PIRLS 2016*. https://www.udir.no/globalassets/filer/tall-og-forskning/forskningsrapporter/pirls2016_hovedrapport.pdf
- Grice, J. W. (2011). *Observation oriented modeling. Analysis of cause in the behavioral sciences*. Elsevier.
- Hattie, J. (2009). *Visible learning: A synthesis of meta-analyses relating to achiev*. Routledge.
- Høgheim, S. (2020). *Masteroppgaven i GLU*. Fagbokforlaget.
- Jebb, A. T., Ng, V. & Tay, L. (2021). A review of key Likert scale development advances: 1995–2019. *Frontiers in Psychology*, 12, 637547. <https://doi.org/10.3389/fpsyg.2021.637547>

- Jensen, F., Pettersen, A., Frønes, T. S., Kjærnsli, M., Rohatgi, A., Eriksen, A. & Narvhus, E. K. (2019). *PISA 2018. Norske elevers kompetanse i lesing, matematikk og naturfag*. Universitetsforlaget.
- Kane, M. (2009). Validating the interpretations and uses to test scores. I R. W. Lissitz (Red.), *The concept of validity: Revisions, new directions, and applications* (s. 39–64). Information Age Publishing.
- Kane, M. (2012). All validity is construct validity. Or is it? *Measurement: Interdisciplinary Research and Perspectives*, 10, 66–70. <https://doi.org/10.1080/15366367.2012.681977>
- Kelley, T. L. (1927). *Interpretation of educational measurements*. World Book Company.
- Kleven, T. A. (2002). Begrepsoperasjonalisering. I T. Lund (Red.), *Innføring i forskningsmetodologi* (s. 141–184). Unipub.
- Kleven, T. A. (2008). Validity and validation in qualitative and quantitative research. *Nordic Studies in Education*, 28(3). <https://doi.org/10.18261/ISSN1891-5949-2008-03-05>
- Kleven, T. A., Hjørdemaal, F. & Tveit, K. (2011). *Innføring i pedagogisk forskningsmetode. En hjelp til kritisk tolkning og vurdering* (2. utg.). Unipub.
- Kyngdon, A. (2008). The Rasch model from the perspective of the representational theory of measurement. *Theory & Psychology*, 18(1), 89–109. <https://doi.org/10.1177/0959354307086924>
- Lord, F. M. & Novick, M. R. (1968). *Statistical theories of mental test scores*. Addison-Wesley.
- Luce, R. D. & Tukey, J. W. (1964). Simultaneous conjoint measurement: A new type of fundamental measurement. *Journal of Mathematical Psychology*, 1(1), 1–27. [https://doi.org/10.1016/0022-2496\(64\)90015-X](https://doi.org/10.1016/0022-2496(64)90015-X)
- Lund, T. (2001). *Måling av forandring. En innføring*. Uniped.
- Lund, T. (2005). The qualitative–quantitative distinction: Some comments. *Scandinavian Journal of Educational Research*, 49(2), 115–132. <http://dx.doi.org/10.1080/00313830500048790>
- Lund, T., Fønnebo, B. & Haugen, R. (2006). *Forskningsprosessen*. Unipub.
- Maraun, M. D. & Halpin, P. F. (2008). Manifest and latent variates. *Measurement: Interdisciplinary Research and Perspectives*, 6(1–2), 113–117. <https://doi.org/10.1080/15366360802035596>
- Maxwell, J. A. (2004). Causal explanation, qualitative research, and scientific inquiry in education. *Educational Researcher*, 33(2), 3–11. <http://dx.doi.org/10.3102/0013189x033002003>
- Melby-Lervåg, M. & Lervåg, A. (2013). En revurdering av evidensbasert praksis. *Bedre skole*, 4. <https://utdanningsforskning.no/artikler/2013/en-revurdering-av-evidensbasert-praksis/>
- Meld. St. 22 (2010–2011). *Motivasjon – mestring – muligheter*. Kunnskapsdepartementet. <https://www.regjeringen.no/no/dokumenter/meld-st-22-2010--2011/id641251/>
- Meld. St. 28 (2015–2016). *Fag – fordypning – forståelse – en fornyelse av Kunnskapsløftet*. Kunnskapsdepartementet. <https://www.regjeringen.no/no/dokumenter/meld.-st.-28-20152016/id2483955/>
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50, 741–749. <https://doi.org/10.1037/0003-066X.50.9.741>
- Michell, J. (1988). Some problems in testing the double cancellation condition in conjoint measurement. *Journal of Mathematical Psychology*, 32(4), 466–473. [https://doi.org/10.1016/0022-2496\(88\)90024-7](https://doi.org/10.1016/0022-2496(88)90024-7)
- Michell, J. (1990). *An introduction to the logic of psychological measurement*. Lawrence Erlbaum Associates.
- Michell, J. (1997). Quantitative science and the definition of measurement in psychology. *British Journal of Psychology*, 88(3), 355–383. <https://doi.org/10.1111/j.2044-8295.1997.tb02641.x>
- Michell, J. (1999). *Measurement in psychology: A critical history of a methodological concept*. Cambridge University Press.
- Michell, J. (2001). Teaching and misteaching measurement in psychology. *Australian Psychologist*, 36(3), 211–217. <https://www.doi.org/10.1080/00050060108259657>
- Michell, J. (2008). Is psychometrics pathological science? *Measurement: Interdisciplinary Research and Perspectives*, 6(1–2), 7–24. <https://doi.org/10.1080/15366360802035489>
- Michell, J. (2014). The Rasch paradox, conjoint measurement, and psychometrics: Response to Humphry and Sijtsma. *Theory & Psychology*, 24(1), 111–123. <https://doi.org/10.1177/0959354313517524>
- Michell, J. (2021). Denying Descartes and wary of Wittgenstein: Response to Franz. *Theory & Psychology*, 32(1), 151–157. <https://doi.org/10.1177/09593543211046204>
- Michell, J. & Ernst, C. (1997). The axioms of quantity and the theory of measurement. *Journal of Mathematical Psychology*, 41(4), 345–356. <https://doi.org/10.1006/jmps.1997.1178>
- Nilsen, T. & Kaarstein, H. (Red.). (2021). *Med blikket mot naturfag. Nye analyser av TIMSS 2019-data og trender 2015–2019*. Universitetsforlaget. <https://doi.org/10.18261/9788215045108-2021>

Måling i kvantitativ utdanningsforskning: Et instrumentelt mistak?

- Pontoppidan, M., Keilow, M., Dietrichson, J., Solheim, O. J., Opheim, V., Gustafson, S. & Andersen, S. C. (2018). Randomised controlled trials in Scandinavian educational research. *Educational Research*, 60(3), 311–335. <https://doi.org/10.1080/00131881.2018.1493351>
- Ringdal, K. (2018). *Enhet og mangfold* (4. utg.). Fagbokforlaget.
- Shadish, W. R., Cook, T. D. & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.
- Sijtsma, K. (2012). Psychological measurement between physics and statistics. *Theory & Psychology*, 22(6), 786–809. <https://doi.org/10.1177/0959354312454353>
- Sjøberg, S. (2014). PISA-syndromet – hvordan norsk skolepolitikk blir styrt av OECD. *Nytt Norsk Tidsskrift*, 31(1), 30–43. <https://doi.org/10.18261/ISSN1504-3053-2014-01-04>
- Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, 103(2684), 677–680. <https://doi.org/10.1126/science.103.2684.677>
- Stevens, S. S. (1951). Mathematics, measurement, and psychophysics. I S. S. Stevens (Red.), *Handbook of experimental psychology* (s. 1–49). Wiley.
- Stevens, S. S. (1958). Problems and methods of psychophysics. *Psychological Bulletin*, 55(4), 177–196. <https://doi.org/10.1037/h0044251>
- Suppes, P. & Zinnes, J. L. (1963). Basic measurement theory. I R. D. Luce, R. R. Bush & E. Galanter (Red.), *Handbook of mathematical psychology* (s. 1–76). John Wiley and Sons. <https://doi.org/10.2307/2270274>
- Thorndike, E. L. (1918). The nature, purposes, and general methods of measurements of educational products. I G. M. Whipple (Red.), *Seventeenth yearbook of the national society for the study of education* (Bd. 2, s. 16–24). Public School Publishing.
- Wendelborg, C., Dahl, T., Røe, M. & Buland, T. (2020). *Elevundersøkelsen 2019. Analyse av Utdanningsdirektoratets brukerundersøkelser* (Rapport). <https://hdl.handle.net/11250/2689870>
- Wolff, J. E. (2020). *The metaphysics of quantities*. Oxford University Press.
- Zeller, R. A. (1997). Validity. I J. P. Keeves (Red.), *Educational research, methodology, and measurement. An international handbook*. Pergamon.

Appendiks 1: Aksiomer for orden og additivitet

Orden defineres ut fra kriteriene for en enkel orden, gitt en variabel K med verdiene a , b og c :

1. hvis $a \geq b$ og $b \geq c$ så $a \geq c$ (transitivitet)
2. hvis $a \geq b$ og $b \geq a$ så $a = b$ (antisymmetri)
3. enten $a \geq b$ eller $b \geq a$ (total)

Orden er ikke i seg selv tilstrekkelig for å klassifiseres som en kvantitet. Relasjonen mellom verdiene må også være additiv, som her defineres ut fra Otto Hölders aksiomer fra 1901 (se Michell & Ernst, 1997, for engelsk oversettelse), som er ytterligere presisert i Michell (1997, 1999, 2001):

4. For hvert par mengder av K , a og b , er en og bare en av de følgende sann:
 - a. a er lik b ($a = b$)
 - b. det eksisterer en mengde, c , slik at $a = b + c$
 - c. det eksisterer en mengde, c , slik at $b = a + c$
5. For enhver mengde, a og b , i K , er $a + b > a$ og $a + b > b$
6. For enhver mengde, a og b , i K , er $a + b = b + a$
7. For enhver mengde av a , b og c i K , er $a + (b + c) = (a + b) + c$
8. For enhver mengde a finnes det en annen, b , slik at $b < a$
9. For enhver mengde a og b i K er det en annen, c , slik at $c = a + b$
10. Enhver ikke-tom delmengde av K som har en øvre skranke, har en minste øvre skranke.

I noe mindre tekniske termer, forteller aksiom 4 at to mengder er enten lik eller ulik, og om ulik vil en tredje mengde eksistere som differansen mellom dem. Aksiom 5 sier at enhver sum av to mengder er større enn de summerte, og aksiom 6 at rekkefølgen av sammensatt deler er irrelevant for helheten. Aksiom 7 forteller at summen av tre mengder er lik uavhengig om det tilføyelsen av den første mengden på summen av de siste to eller tilføyelsen av den tredje på summen av de to første. Aksiom 8 og 9 sier henholdsvis at det ikke finnes en minste eller største mengde av en kvantitet. Aksiom 10 ivaretar kontinuitet ved å sørge for at det ikke er noen hull i en rangert serie av tall. For eksempel, anta at vi har en delmengde A av en kvantitet, slik som lengde, som er mengder, a , begrenset opptil X (m.a.o. X er øvre skranke). Innad i delmengden er det ingen begrensninger slik at for hver lengde a som er mindre enn X , finnes det en annen lengde som er større enn a og mindre enn X . Anta videre at vi har en annen delmengde B som har X som nedre skranke, som medfører alle lengder i B , b , er en øvre skranke av a . Det er heller ingen indre begrensninger innad i B . Dersom vi bare har alle lengdene av A og B og rangerer disse vil det være et gap i tallrekken, nemlig verdien X . Siden aksiom 10 inkluderer en minste øvre grense, sørger den for

inkluderingen av X og dermed at det er kontinuitet i tallrekken (for mer detaljer om aksiomene anbefales Michell, 1999, s. 47–59).

Aksiom 4–10 viser ikke til matematiske operasjoner da de viser til *mengder*, ikke tall (m.a.o. verdier). De matematiske symbolene viser til *relasjoner* mellom mengder (Michell, 1997). Symbolet for addisjon (+) viser til relasjoner mellom mengder i den forstand at « $a + b = c$ » betyr at mengden c består av de separate delene a og b . En ren matematisk forståelse vil føre til at bare de kvantitetene som er direkte håndterlig vil oppfylle kravene for additivitet, slik som lengde og masse; eller ekstensive kvantiteter. De intensive kvantitetene, som ikke er direkte håndterbar, vil da ikke kunne oppfylle aksiomene, slik som temperatur og velositet.