

Appendiks 1: Om baggrund og teori bag valg af skala

De nationale test gav i 2010 for første gang danske lærere mulighed for at foretage en egentlig måling på en skala af deres elevers præstationer på grundlag af videnskabelige principper. Af de nationale test er testene i matematik blandt de vigtigste. Matematiske færdigheder er centrale for megen videre uddannelse og i nogen grad også i dagligdagen. Samtidig kan den matematiske forståelse være svær at følge udviklingen i. Der er brug for redskaber til at observere de latente egenskaber. Testen kan for læreren være, hvad termometeret og blodtryksmåleren er for lægen.

De nationale test skal passe til Rasch-modellen, som er den ultimative videnskabelige standard inden for test. Dette indebærer også at den bagved liggende skala til opgørelse af både opgavers sværhed og elevers dygtighed er Rasch-skalaen.

Grundlæggende Rasch-skalaer måles i enheden Logitter – Skalaen går principielt fra minus til plus uendelig (i praksis fra ca. -7 til +7), og nulpunktet fastlægges som regel ud fra den middelsvære opgave i opgavebanken. Når skalaen kan anvendes både til at fastlægge elevers dygtighed og opgavers sværhedsgrad kan man på et videnskabeligt grundlag afgøre hvilke opgaver der passer bedst til hvilke elever (baggrunden for den adaptive algoritme).

Et særligt kendetegn for Raschskalaen er at det er en intervallskala¹. Denne skalatype har en række belejlige egenskaber for dem, der anvender den til videnskabeligt og professionelt brug – først og fremmest at det er en lineær skala. Længden på ethvert interval på skalaen er sammenligneligt.

Det vil sige, at en forbedring i den matematiske færdighed fra -2 til 0 er lige så stor som en forbedring fra 1 til 3 og dobbelt så stor som forbedringen fra 1½ til 2½. I modsætning til andre skalaer (fx percentilskalaer) varierer forskelle målt på skalaen, ikke afhængigt af hvor på skalaen, man befinder sig. Man taler om, at målinger på denne skala er invariante. Betydningen og af og forklaringen herpå vender vi tilbage til nedenfor.

Men Raschskalaen er samtidigt meget abstrakt, og både lærere, elever og forældre har behov for at kunne sammenholde deres resultat med noget, som de kan forholde sig til. For at gøre testresultaterne forståelige er det valgt at rapportere resultater fra de nationale test i systemet på såkaldte percentil-skalaer, som går fra 1 til 100 point og opdeler landets elever i 100 grupper (procentdele) efter stigende præstation. Eleverne rangordnes efter deres raschscore og den dårligst præsterende procentdel af eleverne får tillagt værdien 1 – de midterste elever får 50 point, og den dygtigste procentdel får 100 point osv. Til brug for dialog med hjemmet og elever blev der på grundlag af percentilskalaerne konstrueret en fem-trinsskala, der minder om karakterer: 1-10=Klart under middel, 11-35=Under middel, 36-65=Middel, 66-90=Over middel og 91-100=Klart over middel. Denne type skala kaldes en ordinalskala², som ved rangordning af elever er meget velegnede til at give læreren et forståeligt mål for elevernes faglige niveau i forhold til det gennemsnitlige niveau (normen) i Danmark. Men denne skalatype har to mangler i forhold til brug i det pædagogiske arbejde, der kan illustreres på to måder:

1. Dels kan en progression på fx 10 percentiler repræsentere en meget forskellig fremgang forskellige steder på skalaen. For eksempel: En elev scorer "1" i 2010 i 3. klasse og "10" i 2013 i 6. klasse. En

¹ Intervallskala: En skala hvor alle afstande er lige store og der derfor kan regnes gennemsnit, summer og differencer. Temperatur-, længde- og vægtskalaer er gode eksempler på denne type. En oversigt over forskellige skalatyper kan findes her: S. S. Stevens (1946), "On the Theory of Scales of Measurement", Science, Vol. 103.

² Ordinalskala: En skala, hvor observationer (fx elevernes præstationer) kan rangordnes, men hvor værdierne på skalaen kun er knyttet til en rangorden og ikke kan bruges til at regne videre på en teoretisk holdbar måde, selvom man ofte gør det – fx ved beregning af eksamensgennemsnit.

Beregneren - progression i de nationale matematiktests
 - Vejledning til brug af beregner af progression i matematik

anden elev scorer "40" i 2010 og "50" i 2013. Hvad betyder en progression på ca. 10 point for de to elever?

2. Dels kan resultaterne fra de to matematiktest (til hhv. 3. og 6. klasse), der er opgjort på hver deres skalaer, ikke sammenlignes indbyrdes. Eksempel: En elev i 3. klasse scorer "60" i 2012 og 3 år senere i 2015 scorer eleven "45" i 6. klasse testen. Spørgsmålet man som lærer kan stille sig er, om "eleven fagligt er gået frem, tilbage eller har han stået stille?"

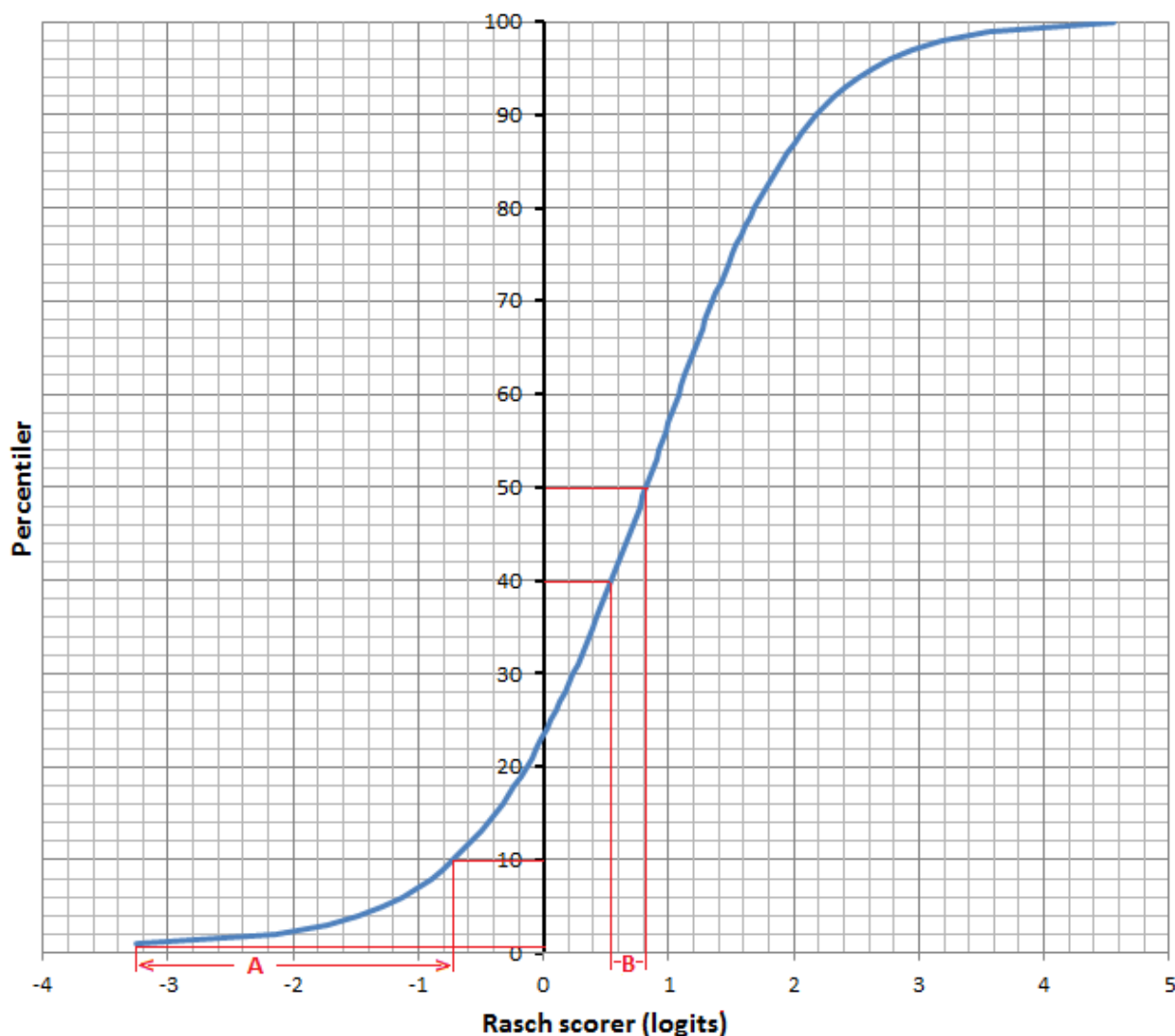
For at illustrere dette skal vi dels kunne placere resultaterne fra de fire forskellige test på den **samme sammenlignelige skala**, og dels skal denne skala være en **intervalskala**, ligesom det er tilfældet med den bagved liggende Rasch-skala.

Ad 1: Placering på skalaen

Det forekommer umiddelbart logisk at gå ud fra, at en fremgang på 10 point er nogenlunde det samme, uanset hvor man befinder sig på skalaen. Men sådan er det ikke nødvendigvis.

Figur 1

Sammenhængen mellem percentiler og Rasch-scorer (logits) for Matematik i Anvendelse, 3. kl.



Beregneren - progression i de nationale matematiktests
- Vejledning til brug af beregner af progression i matematik

Af Figur 1 fremgår forholdet mellem Percentil- og Rasch-skalaen for MiA i 3 kl.-testen (det ser nogen lunde tilsvarende ud for de andre tre profilområder på to klassetrin hvor der testes). En fremgang fra 1 til 10-percentilen svarer til ca. $2\frac{1}{2}$ enheder (logitter) på Raschskalaen (markeret med "A" i figuren), mens en fremgang fra 40 til 50-percentilen svarer til knap $\frac{1}{3}$ logit (markeret med "B"). En fremgang fra 1 til 10 point repræsenterer en forbedring af Matematik i anvendelse, der er ca. 7 gange så stor som fremskridtet fra 40 til 50-percentilen.

Uanset at de fleste lærere har en god fornemmelse af deres elever, så er det altså helt afgørende for vurderingen af progression, at man måler på en egnet skala, hvor man ved, hvad en enhed betyder. Især hvis man vil se på en elevgruppes (fx en classes) samlede fremskridt, vil valg af den rette skala betyde forskellen mellem en solid måling og et svært fortolkeligt, måske stærkt misvisende tal: Alene på grund af valg af skala kan der vise sig målefejl på gennemsnitsresultater svarende til næsten to års normal progression for en klasse.

Ad 2: Udviklingsarbejdet og valg af fælles skalaer

For hvert profilområde er der to percentilskalaer med bagved liggende Raschskalaer, der bør måle samme underliggende egenskab (fx Tal & Algebra 3. og 6. klasse).

Men hverken enheden (hvilken fremgang i Tal & Algebra betyder en tilvækst på en logit) eller 0-punktet (den gennemsnitlige opgaves sværhedsgrad i hver af testene) kan på forhånd antages at være den/det samme for de to Raschskalaer.

Relationerne mellem 3. klasses testen og 6. klasses testen blev i projektet undersøgt ved, at en række elever gennemførte begge test på samme vilkår inden for en uges mellemrum³.

Når den samme elev scorede forskelligt i de to test, kan det principielt skyldes to forhold:

- Forskelle i testene (fx opgavernes sværhedsgrad).
- Såkaldte "målefejl" (som følge af fx elevernes gode/dårlige dage, koncentration, kendskab til indhold af konkrete opgaver, teknik-bøvl, instruktion og andre omstændigheder ved testafviklingen).

Ca. 250 elever deltog i arbejdet, men ikke alle elever besvarede begge test på de skitserede vilkår. På grundlag af 156 elever (der hver besvarede to test inden for samme uge på et helt sammenligneligt grundlag) kunne de parvise relationer mellem matematiktestene til 3. og 6. klasse undersøges på grundlag af elevernes scorer. Skalaerne måtte naturligvis ækvivaleres profilområde for profilområde.

På det grundlag kan man placere besvarelsene på fælles skalaer med de rigtige egenskaber, givet at visse forudsætninger er opfyldt:

- At opgaverne i hver af de 6 opgavebanker hver især følger en raschmodel. Dette krav er afprøvet som et led i etableringen af opgavebankerne, hvor alle skalaerne ved idriftsættelsen levede op til Raschmodellen krav.
- At de to test måler præcis de samme underliggende egenskaber hos eleverne - Samme slags Tal & Algebra, Geometri og Matematik i Anvendelse⁴.

³ Dette er under normale omstændigheder ikke muligt, da man kun kan gennemføre de frivillige test på et klassetrin over og et under det klassetrin testen er beregnet til. Men det blev muliggjort i samarbejde med administratorer i forbindelse med Undersøgelsens gennemførelse ved at lade eleverne midlertidigt rykke klassetrin op eller ned.

⁴ Ved udviklingen af testene var det aftalt, at testene/testopgaverne skulle fremstilles/afprøves på den måde, men det er ikke så vidt vides undersøgt til bunds om dette er sket og/eller holder efter testene er sat i drift - og i givet fald for

Beregneren - progression i de nationale matematiktests
- Vejledning til brug af beregner af progression i matematik

Dannelse af mat-i alt-skala

I testsystemet fremgår en samlet vurdering af en elevs matematik færdigheder. Denne vurdering/score er et gennemsnit af percentilscore fra hver af de tre profilområder. Men som det er beskrevet ovenstående, er denne beregning rent matematisk problematisk (giver en matematisk betinget fejl alene på grund af skalavalg svarende til op til 2 års progression).

Derfor har vi dannet en samlet matematik i alt score ved at omsætte resultaterne fra profilområderne i 3. og 6. klasseset til en raschscore og derefter beregne et vægtet gennemsnit af de tre profilområders rasch-scorer. Principielt kunne man forudsætte at hvert af de tre profilområder skulle veje det samme i gennemsnittet (altså 1/3 hver). Men der kunne være belæg for en forud antagelse om, at visse færdigheder er vigtigere end andre for den samlede matematiske færdighed – ved at der er noget) der er mere grundlæggende og uden hvilket man ikke kan løse nogen opgaver overhovedet, fx kendskab til tallene og renearterne.

For at afprøve denne antagelse har vi set på sammenhængen mellem resultaterne fra profilområderne og den samlede score fra en generel matematiktest⁵, der er udviklet til at måle de samme underliggende egenskaber som de nationale test.

Gennem en regressionsanalyse viste det sig, at en meget betydelig del af variationen i det samlede matematikresultat kunne forklares ved scorerne fra de tre profilområde, og bidraget fra hver af de tre profilområder kunne derved estimeres⁶.

På dette grundlag er skalaen Mat-i alt-skalaen dannet.

Om valg af skalatype

Man kunne have valgt at arbejde videre med én af Raschskalaerne til testene (en logit-skala) eller en anden intervallskala (fx en TIMMS- eller en PISA-skala). Men samtidigt er det vigtigt, at valget af skala skal forbindes med en pædagogisk problemstilling og ikke rangordning, kontrol, politik eller administration. På grundlag af amerikanske forskningsbaserede erfaringer og tidligere arbejde med danske elever/lærere/skoler blev der valgt en skala (Lexile-skalaen⁷) som model for de fælles skalaer.

Konklusion

Begrundelsen for at afrapportere på enten percentilskalaer (1-100-skalaer) eller de karakterlignende 1-5 skalaer ift. middel (ordinalskalaer) holder stadigvæk. Denne type af skaler er egnede til formidling, fordi de afspejler måden, forældre, elever og skoler sædvanligvis tænker uddannelsesresultater på.

Men hvis kommuner og skoler skal kunne arbejde professionelt med testene som pædagogisk redskab og til monitorering af udvikling, så kræver det, at denne tilgang suppleres med udviklings-/progressionsskalaer, hvor elevernes faglige udvikling kan følges over tid. På den måde kan man meningsfuldt beregne den

alle klassetrin og for alle profilområder. Dette vil ikke kunne undersøges til bunds inden for dette projekts rammer. Det ville kræve en analyse af de enkelte opgaver og opgavebesvarelser (altså Rasch- og/eller Item Respons analyser).

⁵ NordicMetrics har udviklet en matematiktest (som bl.a. bruges i DR's dokumentarserie "Folkeskolen Forfra") med opgaver inspireret af de Amerikanske NAEP-test i matematik, denne test er tilpasset Raschmodellen og ækvivaleret med DNT. En beskrivelse af denne test er gengivet her: http://www.nordicmetrics.com/wp-content/uploads/2014/05/NordicMetrics-test_v27.pdf

⁶ Ved at tvinge regressionslinjen gennem origo, fås koefficienterne til de tre profilområders scorer (der er de forklarende variabler). Resultaterne viser at Tal og algebra vejer lidt tungere end de to andre profilområder i en samlet matematikscore.

⁷ Lexile-skalaen er udviklet i USA og bruges til både at beregne en teksts sværhed og elevens læsefærdigheder (Se <http://Lexile.com>). Når man kender elevens dygtighed og tekstens sværhed og har dem opgjort på samme skala, kan man let udvælge tekster, der lige præcis passer til eleven. I nærværende sammenhæng anvendes skalaen alene til beregning af elevernes dygtighed.

Appendiks 1

Beregneren - progression i de nationale matematiktests
- Vejledning til brug af beregner af progression i matematik



enkelte elevs udvikling og endvidere på et videnskabeligt korrekt grundlag beregne gennemsnit, så man kan vurdere klassens og skolens gennemsnitlige udvikling.

For hvert profilområde er der altså en række forskellige skalaer:

- Percentilskalaer (1-100-skalaerne), der måler op mod normen i 2010, da testene blev lanceret
- 5-skalaen, en gruppering af percentiler i forhold til middel (kaldet normskalaen)
- En kriteriebaseret udgave der måler i forhold til ministeriets fagfolks vurderinger (se også Appendix 2)
- De bagvedliggende logit-skalaer, der er specifikke for klassetrin
- Den fælles progressionskala for hvert profilområde og i alt

Du kan læse mere om, hvordan du kan bruge beregneren og testsystemet til at få indsigt i den enkelte elevs progression i Appendix 3.