

Kriteriebaseret scoring i de nationale test - den statistiske model

I de nationale test beregnes elevernes dygtighed i et givent profilområde ved anvendelse af en adaptiv algoritme. Elevernes dygtighed er bestemt på baggrund af besvarelsenerne af et varierende antal opgaver.

Den bagved liggende statistiske model¹ til bestemmelse af elevdygtigheden er den såkaldte Rasch model. Her er sandsynligheden for rigtigt svar på en dikotom opgave k (item Z_k) givet ved:

$$P(Z_k = 1) = \frac{\exp(\theta - \delta)}{1 + \exp(\theta - \delta)}$$

hvor θ = elevens dygtighed (personparameter)

δ = opgavens sværhedsgrad (opgaveparameter)

Elevens beregnede dygtighed er siden opstarten af de nationale test i 2010 efterfølgende omsat til en percentilværdi (1-100) med henblik på formidling af elevens resultat. Percentilværdien er et udtryk for elevens præstation i forhold til andre elever på landsplan.

Ved den kriteriebaserede scoring omregnes elevens dygtighed til en forventet score på et antal udvalgte opgaver – de såkaldte kriteriebaserede opgaver.

Lad Z_1, Z_2, \dots, Z_n være n kriteriebaserede opgaver i et profilområde, hvor den maksimale score er lig med M . Hvis de kriteriebaserede opgaver alle er dikotome, da vil M være lig med antallet af opgaver, n .

Hvis en elev med dygtigheden θ svarede på disse opgaver, ville fordelingen af den samlede score $T = \sum_j Z_j$ være givet ved sandsynlighederne²:

$$P(T = t|\theta) = \frac{\exp(t\theta)\gamma_t}{G(\theta, \gamma_0, \dots, \gamma_n)}$$

hvor γ_t er symmetriske polynomier i item parametrene for kriterieopgaverne og

$$G(\theta, \gamma_0, \dots, \gamma_n) = \sum_{x=0}^M \exp(x\theta)\gamma_x$$

¹ Carsten Bendixen & Svend Kreiner: Test i folkeskolen. Hans Reitzels Forlag 2009

² Karl Bang Christensen, Svend Kreiner & Mounir Mesbah: Rasch Models in Health. Wiley 2013

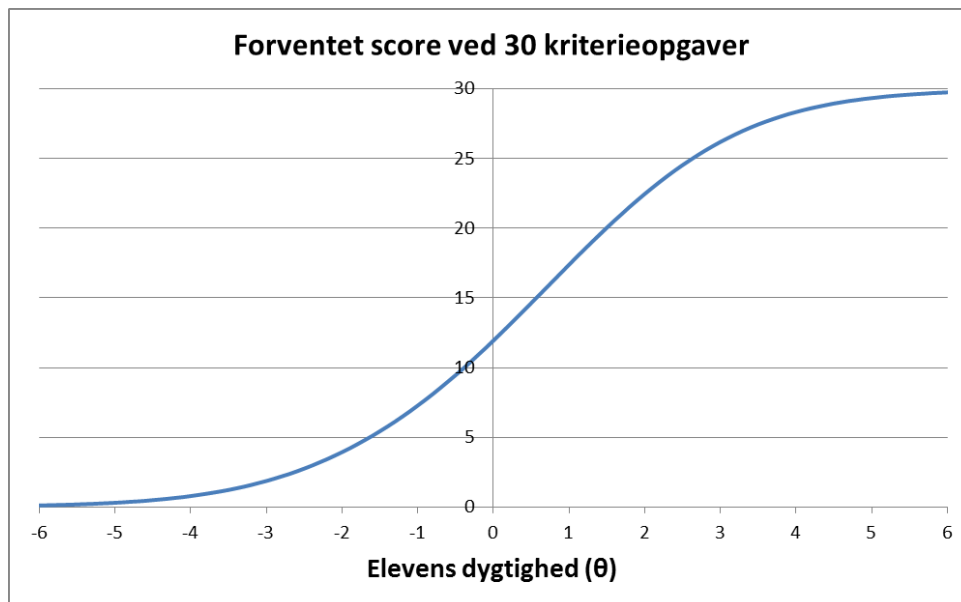
Elevens forventede score, $E(T|\theta)$, på de n kriteriebaserede opgaver kan nu beregnes ved³:

$$f(\theta) = E(T|\theta) = \sum_{t=0}^M t \frac{\exp(t\theta)\gamma_t}{G(\theta, \gamma_0, \dots, \gamma_n)}$$

Da $f(\theta)$ er en monoton voksende funktion af θ , kan $f(\theta)$ betragtes som en alternativ parametricering af personparameteren, θ , i Rasch modellen. Denne parametricering omtales som en kriteriebaseret parametricering.

I figur 1 er parametriceringen illustreret ved 30 udvalgte dikotome kriterieopgaver.

Figur 1 Sammenhæng mellem elevens dygtighed på Rasch skalaen og den forventede score i 30 kriterieopgaver (maxscore=30)



Beregningerne foretages herefter trinvis:

Første trin er at udvælge opgaver i hvert profilområde til den kriteriebaserede skala. Opgaverne skal være karakteristiske for det pågældende profilområde. Dette er foretaget af Undervisningsministeriets opgavekommissioner knyttet til de nationale test.

Andet trin er at beregne de symmetriske polynomier for de n kriteriebaserede opgaver på baggrund af de estimerede opgavesværheder. Dette er gjort i forbindelse med opgaveafprøvingerne og skal kun beregnes én gang.

³ Everett V. Smith, Jr. & Gregory E. Stone: Criterion Referenced Testing: Practice Analysis to Score Reporting Using Rasch Measurements Models. JAM Press 2009

Tredje trin er at estimere elevens dygtighed, θ , ud fra svarene på de opgaver, der rent faktisk er stillet. Denne beregnes ved anvendelse af den adaptive algoritme i forbindelse med afvikling af testene.

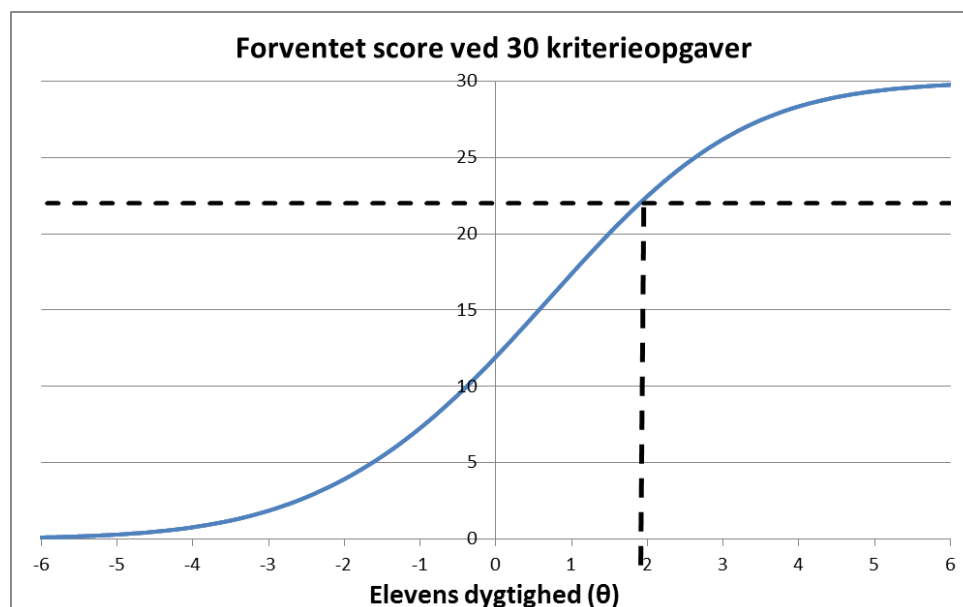
Fjerde trin er at indsætte estimatet for elevens dygtighed, $\hat{\theta}$, i formlen $f(\hat{\theta})$. Dette giver således den forventede score for eleven på de n kriteriebaserede opgaver. Det skal bemærkes, at det ikke er en forudsætning for beregningen, at eleven under afviklingen af testen svarer på én eller flere af de kriteriebaserede opgaver.

Ved den kriteriebaserede scoring omregnes elevens dygtighed på Rasch skalaen således til en forventet score på et antal udvalgte opgaver.

I figur 2 er vist et eksempel på sammenhængen mellem elevens estimerede dygtighed og den forventede score i 30 kriteriebaserede opgaver med en samlet maxscore på 30.

Her er angivet en elev, hvor dygtigheden ud fra den adaptive algoritme er estimeret til 1,9 på Rasch skalaen. 1,9 svarer til en forventet score på 22 på den kriteriebaserede skala i det pågældende eksempel. Hvis eleven fik stillet de 30 kriteriebaserede opgaver, da ville vi således forvente at eleven ville kunne svare korrekt på 22 af disse.

Figur 2 Eksempel på sammenhæng mellem elevens estimerede dygtighed (1,9 på Rasch skalaen) og den forventede score (22) i 30 kriterieopgaver (maxscore=30)



På baggrund af faglige vurderinger er der for hvert profilområde i dansk, læsning og matematik opstillet en tabel, der viser sammenhængen mellem faglige kriterier og den forventede score i de udvalgte kriterieopgaver. Scoren til hvert kriterie kan nu omregnes til en dygtighed på Rasch skalaen. Denne kaldes cutscoren. Tabel 1 viser

et eksempel på, hvordan sammenhængen mellem de faglige kriterier og elevens dygtighed kunne se ud.

Tabel 1 Eksempel på sammenhæng mellem faglige kriterier og den forventede score på de 30 kriterieopgaver

Kriterium	Antal korrekte (maxscore=30)	Elevdygtighed på Rasch skalaen
Fremragende præstation	27	3,3
Rigtig god præstation	22	1,9
God præstation	17	1,0
Jævn præstation	12	0,0
Mangelfuld præstation	7	-1,0
Ikke tilstrækkelig præstation	Under 7	<-1,0

Når elevens dygtighed er estimeret på baggrund af den adaptive algoritme, da kan det vurderes, hvilket fagligt kriterie eleven forventes at opfylde.

Det skal bemærkes, at antallet af kriterieopgaver, den tilhørende maxscore, fastlæggelsen af de faglige kriterier samt de beregnede cutscores afhænger af det enkelte profilområde.