



Notat

Vedrørende: Den adaptive algoritme i De Nationale Test

Olof Palmes Allé 38
8200 Aarhus N
Tlf.nr.: 35 87 88 89
E-mail: stil@stil.dk
www.stil.dk
CVR-nr.: 13223459

Den adaptive algoritme i De Nationale Test

januar 2015

Dette dokument beskriver i hovedtræk hvordan den adaptive algoritme fungerer i De Nationale Test og hvordan elevdygtigheden bliver beregnet ved anvendelse af Rasch modellen.

Opbygning af test og testforløb

Eleverne skal igennem deres skoletid gennemføre 10 obligatoriske nationale test inden for forskellige fag. Hver test består af tre såkaldte profilområder. Eksempelvis består dansk, læsning i 2. klasse af profilområderne: ”Sprogforståelse”, ”Afkodning” og ”Tekstforståelse”.

De nationale test er **selvscorende** og **adaptive**.

At testene er selvscorende betyder, at lærerne ikke selv skal rette testene, men får leveret resultaterne direkte. Når testen er afsluttet, kan der beregnes en elevdygtighed for hvert profilområde. Udregningen af elevdygtigheden sker ved anvendelse af Rasch modellen¹. Denne elevdygtighed er udtrykt på en såkaldt logit-skala, men kan efterfølgende transformeres til anden skala med henblik på formidling. Percentilskalaen (1-100) er et eksempel på dette.

At testene er adaptive betyder, at testene tilpasser sig den enkelte elevs dygtighed undervejs i testforløbet.

Eleven testes skiftevis i de tre profilområder, indtil usikkerheden i elevens estimerede dygtighed i et profilområde kommer under en fastsat grænse². Hvis usikkerheden i elevens dygtighed i ét profilområde kommer under den fastsatte grænse før de øvrige profilområder, koncentrerer testafviklingen omkring de øvrige områder. Testafviklingen fortsætter, indtil usikkerheden i elevdygtigheden i alle profilområder ligger under

¹ Se bilag 1 for forklaringer på statistiske og tekniske termer.

² I testsystemet anvendes en grænse på usikkerheden på elevdygtigheden (SEM) på 0,55.

den fastsatte grænse, og den afsatte tid³ til afvikling af testen er gået. Læreren kan forlænge testen, hvis der ikke er opnået tilstrækkelig præcision, inden tiden er gået.

I praksis består et testforløb således af tre individuelle test, som i forløbet blandes i mønsteret 1, 2, 3, 1, 2, 3, ... indtil usikkerheden er under den fastsatte grænse indenfor et profilområde. Derefter kunne mønsteret eksempelvis fortsætte med ..., 1, 3, 1, 3, 3, 3, så usikkerheden i elevdygtigheden i alle tre profilområder til sidst ligger under den fastsatte grænse.

Afviklingen af et testforløb indenfor hvert af de 3 profilområder forløber helt ens. Først beskrives afviklingen af et testforløb i ét enkelt profilområde.

Testforløbet består af to dele:

- **Opgavevalg**
- **Beregning af elevdygtighed**

Opgavevalg i forbindelse med afvikling af et testforløb

Første del af afviklingen af et testforløb afdækker, hvilken opgave eleven skal præsenteres for. Opgaverne stilles ikke tilfældigt, men udvælges på baggrund af den viden, som testen indtil da har givet om elevens dygtighed.

Udvælgelsesmetoden er forskellig for tre faser i testforløbet. De tre faser gentages for hvert profilområde. De første to faser kaldes samlet for **run-in perioden**.

1. Den første opgave er speciel, idet alle test påbegyndes ved at anvende en opgave med en sværhedsgrad⁴, der ligger på middelniveauet for det pågældende profilområde. Middelniveauet afhænger af profilområdet og er på enten 0,0 eller 0,5 med undtagelse af profilområde 2 i dansk, læsning, hvor middelniveauet er 1,5 eller 2,0 afhængig af klassetrin. Der anvendes kun den type opgaver, som kan besvares enten rigtig eller forkert. Dette kaldes opgaver med kompleksitet på 1 eller for dikotome opgaver.
2. Opgave 2 og 3 er ligeledes specielle, idet det også her sikres, at kompleksiteten i opgaven er 1. Til gengæld afhænger opgavens sværhedsgrad nu af elevens svar på den forrige opgave. Har eleven svaret rigtigt på forrige opgave, vælges en opgave med en sværhedsgrad som er omkring 0,5 højere. Ved forkert svar vælges en opgave med sværhedsgrad 0,5 lavere. For enkelte profilområder er ændringen dog kun på 0,25.
3. Fra og med opgave 4 kan eleven få stillet alle typer opgaver, også opgaver med kompleksitet over 1, altså opgaver, som har flere

³ Der afsættes 45 minutter

⁴ Værdierne for sværhedsgraden (også benævnt location værdier) befinder sig i intervallet -7 til 7 på logit-skalaen (Rasch skalaen).

delspørgsmål, og som derfor samlet set kan være delvist rigtigt besvaret.

Endvidere vil sværhedsgraden af opgave 4 og fremefter være baseret på en beregning af elevens dygtighed, som inkluderer alle indtil nu afgivne svar. Elevens dygtighed måles på samme skala som opgavernes sværhedsgrad, så eleven vil i næste opgave blive præsenteret for en opgave med en sværhedsgrad, der ligger tæt på den beregnede elev dygtighed.

Eleven kan på intet tidspunkt tildeles en opgaven, som eleven har haft før i testforløbet.

Hvis eleven undlader at besvare en opgave, bliver besvarelsen håndteret som en fejlbesvarelse i det videre forløb. Dermed trækker en blank besvarelse ned i beregning af elevdygtigheden.

Beregning af elevdygtighed i forbindelse med afvikling af et testforløb

Når eleven har besvaret den 3. opgave, sker der en statistisk behandling af elevens besvarelser. Elevens dygtighed estimeres her første gang. Herefter beregnes elevdygtigheden efter besvarelse af hver opgave.

Beregningen af elevdygtigheden sker ved anvendelse af Rasch modellen⁵.

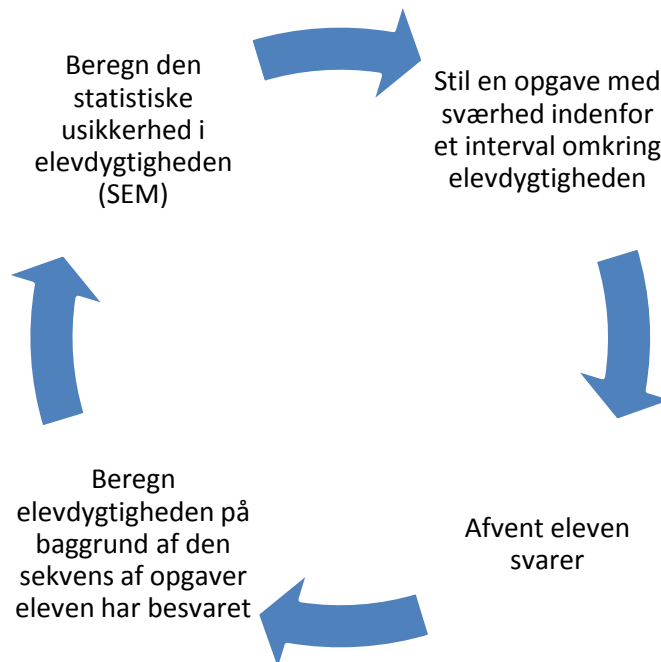
Den beregnede elevdygtighed danner udgangspunkt for udvælgelse af den efterfølgende opgave, der stilles til eleven (se afsnit vedr. opgavevalg).

Efter 3. opgave beregnes endvidere usikkerheden (SEM) på den estimerede elevdygtighed. Usikkerheden på den estimerede elevdygtighed beregnes efter elevens besvarelse af hver af de følgende opgaver. Usikkerheden bliver mindre, des flere opgaver eleven besvarer.

I figur 1 er testcyklus i form af det adaptive princip illustreret.

⁵ Se bilag 1 for forklaringer på statistiske og tekniske termer.

Figur 1: Testcyklus efter besvarelse af opgave 3 og fremefter

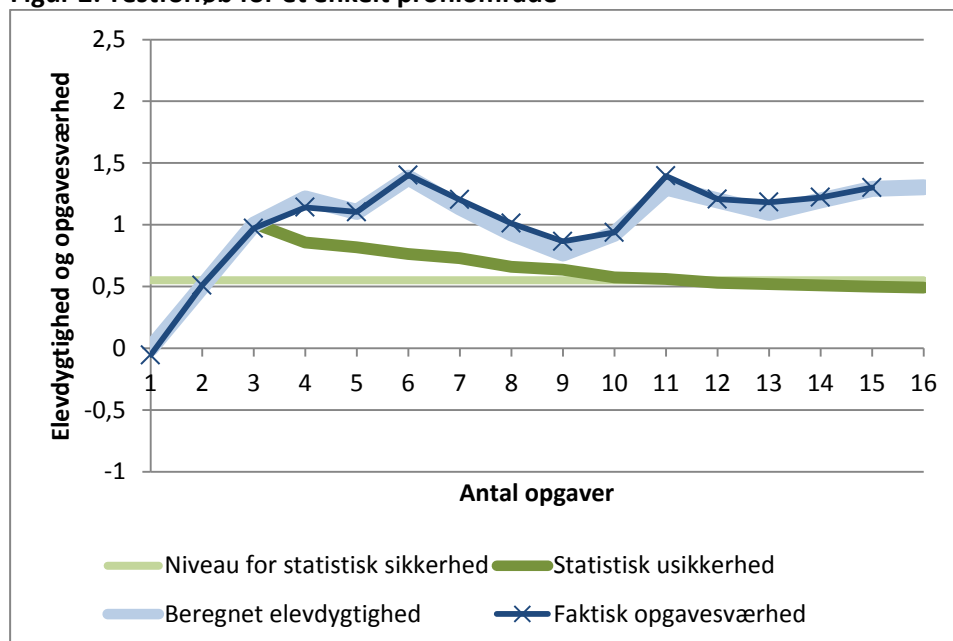


Er den beregnede usikkerhed over den fastsatte grænse på 0,55 fortsættes med flere opgaver i det pågældende profilområde.

Eksempel på et testforløb for et enkelt profilområde

Figur 2 viser udviklingen i henholdsvis opgavernes sværhedsgrad, elevens dygtighed og den beregnede usikkerhed på elevdygtigheden (SEM) under et testforløb.

Figur 2: Testforløb for et enkelt profilområde



Som skitseret i figuren er forløbet af de tre første opgaver specielt. Først efter den 3. opgave beregnes elevens dygtighed ved anvendelse af algoritmen, og de efterfølgende opgaver udvælges, så de har en sværhedsgrad, der ligger så tæt som muligt på den seneste estimerede elevdygtighed.

I det konkrete tilfælde svarer eleven rigtigt på de to første opgaver og afslutter således run-in perioden med en foreløbig dygtighed på +1,0. Den tredje opgave er også besvaret rigtigt, og ved 4. opgave er det algoritmen, der har bestemt elevdygtigheden (1,21) og dermed udgangspunktet for den næste opgaves sværhedsgrad.

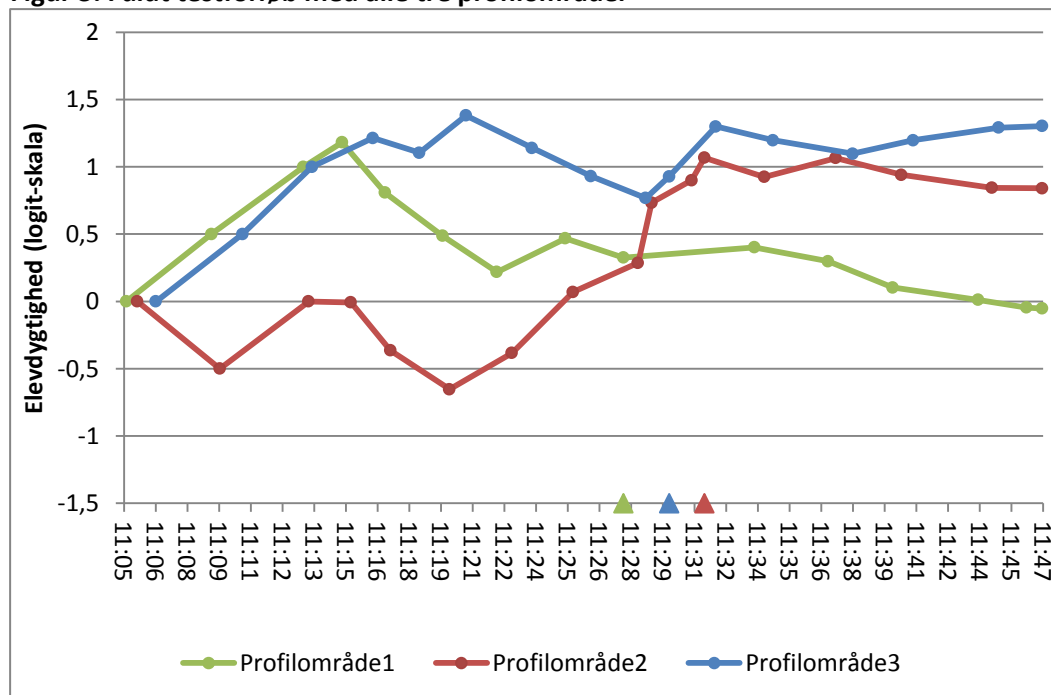
Opgaven, som eleven 4. gang præsenteres for, har dog en faktisk opgavesværthed på 1,14, idet der først findes et passende antal opgaver i et interval omkring den beregnede elevdygtighed på 1,21 og dernæst udvælges en opgave tilfældigt blandt disse mulige opgaver.

SEM-værdien beregnes første gang efter den 3. opgave, og som det fremgår af figuren er det, i dette konkrete tilfælde, først efter 11 opgaver, at eleven krydser grænsen på 0,55. Herefter kan elevens dygtighed i profilområdet estimeres med den fastsatte præcision.

Eksempel på et fuldt testforløb

Figur 3 viser et eksempel på et fuldt testforløb, som eleven oplever det.

Figur 3: Fuldt testforløb med alle tre profilområder



Hver opgave er markeret med en prik på graferne, og det ses, at eleven får tildelt opgaver skiftevis i de tre profilområder, indtil grænseværdien for SEM opnås (markeret med farvet pile på tidsaksen).

Omkring kl. 11:28 er usikkerheden på elevdygtigheden i profilområde 1 kommet under grænseværdien, og testafviklingen koncentrerer efterfølgende omkring profilområde 2 og 3.

Efter kl. 11:32 er usikkerheden på elevdygtigheden for alle profilområder kommet under grænseværdien, og i den resterende tid bliver der igen uddelt opgaver i alle profilområder.

Ved afslutning af testforløbet bliver der beregnet den endelige elevdygtighed på logit-skalaen i hvert af de 3 profilområder.

Bilag 1: Ordforklaringer og definitioner

Kompleksitet, dikotom og polytom

Alle opgaver har registreret et tal for opgavens kompleksitet (kaldes også Items). Hvis tallet er 1, så er opgaven stillet på en sådan måde, at svaret altid kan vurderes som enten rigtigt eller forkert. Hvis tallet er højere, eksempelvis 2, giver opgaven mulighed for, at svaret kan vurderes som fx værende 50 % korrekt. Algoritmen til beregning af elevdygtighed tager højde for dette.

En opgave betegnes som *dikotom*, hvis kompleksiteten er 1, og svaret herved kun kan være rigtigt eller forkert. Den maksimale score for en dikotom opgave er derfor altid 1. *Polytome* opgaver er opgaver, hvor der er flere delspørgsmål, der til sammen kan udtrykke om hele opgaven er svaret rigtigt, eller om kun nogle dele heraf er rigtige. Det kan fx være opgaver, hvor mere end et valg er rigtigt svar. Herved kan man eksempelvis for kompleksiteten 4 opnå at svare 2 rigtige ud af 4 mulige rigtige. Den maksimale score vil altid være større end 1 for polytome opgaver og i eksemplet med kompleksiteten 4 vil den maksimale score være 4.

Da en dikotom opgave blot er et særtilfælde af en polytome opgave, behandles de to opgavetyper ens af algoritmen.

I den nuværende opgavebank er ca. 2/3 af opgaverne dikotome, mens de øvrige 1/3 er polytome, og de har hyppigst en kompleksitet mellem 2 og 5, men der findes også en række opgaver med over 5 i kompleksitet.

Opgavesværhed

Hver opgave har en opgavesværhed (i nogle sammenhænge også benævnt location værdi), der udtrykker den enkelte opgaves relative sværhedsgrad, dvs. hvor den enkelte opgave er placeret, når man rangordner alle opgaver efter sværhedsgrad og indplacerer opgaverne på en kontinuert logit-skala fra -7 til 7.

Opgaverne indenfor hvert profilområde er fordelt i sværhedsgrad, og sværhedsgraden fastsættes ved en såkaldt opgaveafprøvning og en efterfølgende Rasch analyse af svarene herfra.

Elevdygtighed

Er en værdi i intervallet fra ca. -7 til 7, som er angivet på en logit-skala. Skalaen er den samme som for opgavesværheden. Elevdygtigheden er en beregnet værdi på baggrund af den sekvens af opgaver, som eleven allerede har løst for et givent profilområde. Elevdygtigheden er dermed et udtryk for, hvordan eleven klarer sig i det pågældende profilområde.

Elevdygtigheden for en elev estimeres ved anvendelse af formlen for den multiplikative Rasch model:

$$S = \sum_{x=1}^{\max score} \left(x \cdot \frac{\varphi^s \gamma_s}{\sum_{t=0}^{\max score} \varphi^t \gamma_t} \right)$$

hvor s er elevens samlede score og max score er summen af max scores for alle de opgaver, som eleven har fået.

γ_s er summen af alle produkter af item parametre, svarende til responsmønstret med samlet score lig med s

$$\gamma_s = \sum_{(y_1, \dots, y_k: \sum_i y_i = s)} \prod_i \gamma_{iy_i}$$

hvor y_1, \dots, y_k er k opgaver, som er besvaret med en total score på s .

Elevdygtigheden er udtrykt ved $\theta = \exp(\varphi)$ og bestemmes iterativt ved anvendelse af Newton-Raphson metoden.

Elevdygtigheden beregnes efter hvert afgivet svar, og bruges til at udvælge næste opgaves sværhedsgrad.

SEM – Standard Error of Measurement

SEM er et mål for hvor sikkert elevens beregnede elevdygtighed kan bestemmes ud fra den sekvens af opgaver eleven har besvaret. Med andre ord hvor stor variation der er i elevens evne til at besvare de stillede opgaver.

SEM er defineret som:

$$SEM = 1/\sqrt{s^2}$$

hvor s^2 er summen af varianserne af de opgaver, der er forsøgt besvaret, dvs.

$$s^2 = \sum_i (SS_i - M_i^2)$$

hvor i er den i 'te opgave der er forsøgt besvaret og,

$$SS_i = \sum_{x=1}^k \left(x^2 \cdot \frac{\varphi^x \gamma_{ix}}{\sum_{z=0}^k \varphi^z \gamma_{iz}} \right)$$

og

$$M_i = \sum_{x=1}^k \left(x \cdot \frac{\varphi^x \gamma_{ix}}{\sum_{z=0}^k \varphi^z \gamma_{iz}} \right)$$