

Appendiks 3: Elevens adaptive forløb, tidsforbruget og den statistiske usikkerhed

I de nationale test (DNT) i baseres resultaterne som regel på et begrænset antal opgaver (normalt 10-25 items pr. profilområde¹) – som regel noget flere i læsning end i matematik. Hensynet ved design af testene var, at de normalt skulle kunne gennemføres indenfor en lektion og at de skulle kunne give et mere nuanceret resultat end blot en enkelt testscore – en profil af eleven (derfor profilområder): Der skulle gives tre scorer for hver elev pr test. I læsning hedder profilområderne sprogforståelse, afkodning og tekstforståelse; i matematik Tal/Algebra, Geometri og Statistik/Sandsynlighed. De få opgaver per score betyder at de enkelte opgavers besvarelse i en national test får en relativ stor betydning, og at læreren derfor skal være meget opmærksom på, om elevens svar afspejler, hvad eleven faktisk kan.

Når man arbejder med måling af læring/progression som forskellen mellem to testresultater, skal man være ekstra opmærksom, idet hvert af de to testresultater kan være produkter af mere eller mindre problematiske forløb, i den forstand om de viser, hvad eleven kan på en god dag.

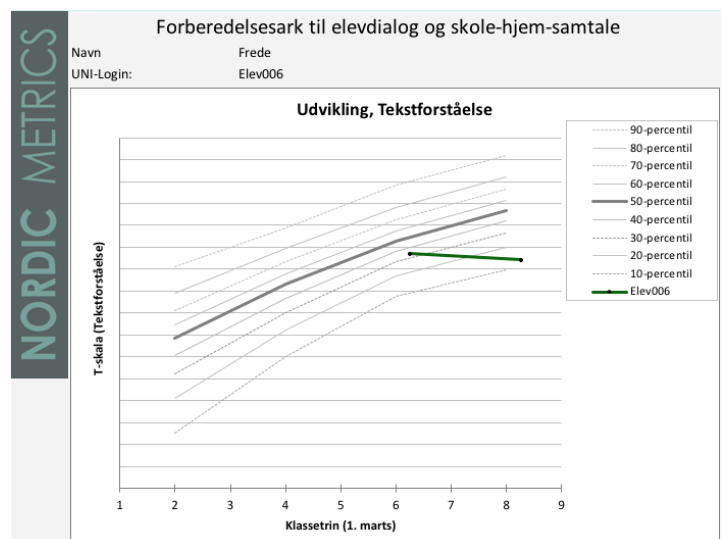
Af figur 1 fremgår en elevs udvikling fra 6. til 8. klasse i tekstforståelse, hvor eleven har præsteret svagere (målt absolut) i 8. end i 6. klasse: Dvs. at progressionen er negativ.

Når eleverne, som her, gennemfører de to test, er det en forudsætning for progressionsberegningen, at testene er gennemført på en ensartet måde (med samme instruktion, samme koncentrationsniveau, med samme hjælpemidler osv.). Således vil beregningen af progression som regel blive nogenlunde retvisende. Sker der imidlertid ændringer (hvis fx eleverne får ordbøger og masser af tid til den ene test, men ikke den anden, eller hvis instruktionen af eleven er forskellig ved de to test) vil progressionsberegningen kunne blive misvisende. Disse "ydre" forhold har læreren som regel mulighed for at kende til og kunne tage højde for ved vurdering af resultatet.

Men eleven kan over- eller underpræstere *uden*, at læreren registrerer det, fx hvis eleven har været ukoncentreret i hele eller dele af det ene eller begge testforløb, ikke har taget besvarelsen af testen alvorligt eller blot har haft en dårlig dag. Dette kan give misvisende resultater, som læreren ikke registrerer/observerer i testsituationen, hvilket, ved mange test er et kedeligt faktum, det er svært at gøre noget ved. Her adskiller DNT sig ved at give adgang til en betydelig mængde information om testforløbet, der i mange tilfælde kan afsløres en problematisk elevpræstation.

Dette fordrer dog, at læreren har en tilstrækkelig viden om testene og desuden færdighed i at fortolke testresultater og således kunne vurdere om de er retvisende for elevens dygtighed. Dette er der kun få, der

Figur 1 – elevforløb fra 6. til 8. klasse (jf. Appendiks 2)



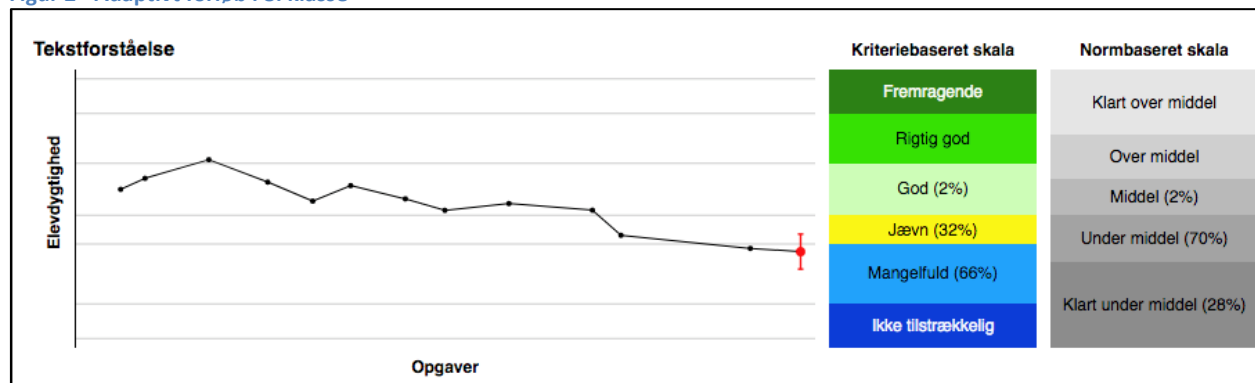
¹ Skal en test leve op til en solid, internationalt anerkendt standard for pålidelighed (realibilitet) skal usikkerheden på estimatet af dygtigheden ned på et lavt niveau. Ofte kræves at usikkerheden (målt ved Standard Error of Measurement eller SEM) reduceres til under 0,3 logits (enheder på Raschskalaen). At nå det niveau kræver besvarelse af mindst 40 optimalt udvalgte spørgsmål. Omsat til scorer på de tre skalaer i Beregneren svarer 0,3 Logits til godt 50 point.

er opmærksomme på, og det er heller ikke umiddelbart enkelt, men der er nogle generelle huskeregler og fif, som gør det mere tilgængeligt, og som illustreres i nedenstående eksempel.

Testforløbet i 8. klasse

I figur 2 er vist en grafisk illustration over elevens adaptive testforløb i 8. klasse. På den lodrette skala vises elevdygtighed målt på en Raschskala (målt i Logits jf. appendiks 1 om skalatyper, det skal bemærkes, at enheden ikke er angivet på akse i testsystemet). På den vandrette akse vises opgavenummeret. Intervallet mellem opgaverne svarer til besvarelsestiden. Grafen viser elevens testforløb, hvor hver sort prik indikerer elevens estimerede dygtighed, altså *hvor* eleven ville ligge, såfremt testen blev stoppet der. I Figur 3 viser en grafisk illustration over Fredes tidsforbrug i opgavebesvarelserne. Til højre i figur 2 vises de to skalaer: Norm- og kriteriebaseret skala. Det ses tydeligt, hvordan "Middel", der udgør 30 percentiler (området fra 35-65 percentilen), dækker over en væsentlig mindre spredning i faglighed end eksempelvis "Klart over middel", der kun udgør 10 percentiler (90-100 percentilen på den normbaserede skala). Læs mere om skalaerne og relationen mellem dem i appendiks 1 om skalatyper.

Figur 2 - Adaptivt forløb i 8. klasse



Figur 3 - Tidsforbrug i 8. klasse

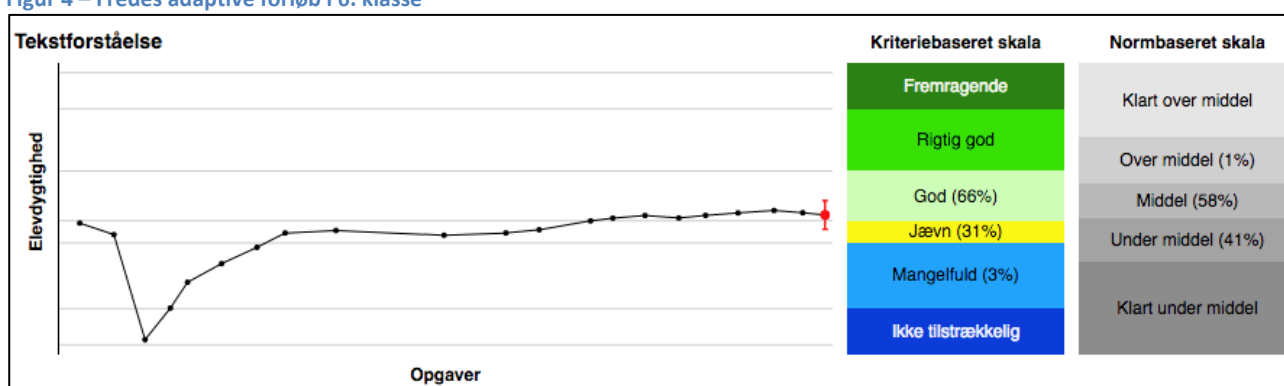


Hvis Frede afsluttede testen midtvejs ville han have i kategorien "God". Som det ses startede Frede fornuftigt ud og svarede rigtigt på de første opgaver, hvorefter han svarer flere forkerte opgaver med enkelte rigtige ind i mellem, og til slut ender han i kategorien "Mangelfuld" på den kriteriebaserede skala. Som det fremgår af figur 3 (øverst) svarer Frede på 13 opgaver og bruger godt 15 min. på det. Dette svarer til, at Frede svarer hurtigere end den normerede tid². Dette illustreres i figur 3 ved at søjlen ligger *under* den stiplede linje. Svarer Frede *langsommere* end normen, er søjlen *over* den stiplede linje. Samlet set ser Fredes besvarelse fornuftigt ud. Dog opgiver han tilsyneladende hurtigt at svare på 3. og 2. sidste opgave.

Testforløbet i 6. klasse

Sammenlignet med Fredes opgavebesvarelse ovenfor, vises i figur 4 en grafisk illustration over hans adaptive testforløb i 6. klasse og figur 5 viser tidsforbruget for besvarelsen.

Figur 4 – Fredes adaptive forløb i 6. klasse



Figur 5 – Fredes tidsforbrug i 6. klasse



² Under afprøvningen af opgaverne er det registreret hvor lang tid, de elever der besvarede opgaven gennemsnitligt brugte på besvarelsen. Denne tid kaldes den normerede tid, og det er under afprøvningen normalt gået noget hurtigere, end når opgaven bruges i drift.

Som det ses, lagde Frede ud med at svare forkert på de første 3 opgaver, hvorefter det beregnes at han ligger i kategorien "Ikke tilstrækkelig". Herefter svarer Frede rigtigt 5 gange i træk og kommer op på niveau igen. Herefter svarer han skiftevis rigtigt og forkert – dog lidt flere rigtige svar, hvorfor han til slut efter 21 opgaver ender i kategorien "God" dog på grænsen til "jævn". Ser vi på Fredes tidsforbrug på opgavebesvarelsen, anvender han stort set den normerede tid (han er en anelse hurtigere end normen) jf. figur 5.

Som det ses, er Frede i begyndelsen (en anelse for) hurtig i opgavebesvarelsen. Herefter bruger han lidt længere tid på opgavebesvarelsen. Her når han op i niveau igen og svarer herefter godt halvdelen rigtigt og bruge lige omkring den normerede tid.

Den statistiske usikkerhed

Den statistiske usikkerhed er et statistisk parameter, som siger noget om, hvor retvisende elevens resultat er, givet at eleven har ydet sit bedste ved besvarelsen af alle spørgsmål. På figurerne 2 og 4 er den statistiske usikkerhed markeret på to måder:

- 1) Den er vist som en rød streg over og under den røde prik, der markerer elevens endelige score. Målet for den statistiske usikkerhed er standardfejlen, også kaldet SEM (Standard Error of Measurement). Længden på hver af de røde streger svarer til SEM.
- 2) Der er nogle procenter i felterne, der markerer kriteriescorene til højre i figurerne. Procenterne markerer konfidensintervaller, hvor der er 95 % sandsynlighed³ for, at scoren vil kunne ligge givet, at eleven har gjort sit bedste hele testen igennem.

Forudsætningen for at usikkerheden er et godt mål for fejl er, at eleverne konsistent gør sit bedste hele vejen igennem testen. Det er erfaringsmæssigt snarere undtagelsen end reglen, hvorfor den statistiske usikkerhed er et usikkert mål for om testresultater er retvisende. Således ses af figur 2, at der iht beregningerne kun var 2 % chance for at Frede ender i kategorien "God", men var testen stoppet tre opgaver tidligere, havde det faktisk været tilfældet. Det er vigtigt at understrege, at den statistiske usikkerhed skal betragtes som en hjælpsom indikator. Men det er samtidig nødvendigt, at læreren forholder sig til testforløbet for at afdække om resultatet er retvisende for elevens faktiske dygtighed. Her kan testsystemets visninger af det adaptive forløb og tidsforbrug være en stor hjælp for dem, der formår at fortolke de oplysninger, der vises heri.

Konklusion

Fredes testforløb i 6. og 8. klasse viser umiddelbart, at han er gået fra "God" til "Mangelfuld". Det ser ud til, at 6. klassetesten rimeligvis kan bruges som mål for Fredes læsefærdigheder ved udgangen af 6. klasse, hvor han starter svagt, men slutter stærkt. Derimod kan testen i 8. klasse meget vel undervurdere Fredes faktiske læsefærdigheder. I størstedelen af testen ligger han faktisk i "God". Han slutter af med fire forkerte svar, heraf to blanke, tilsyneladende uden at læse opgaverne. De to blanke svar var på meget tekstholdige opgaver. Det kan fortolkes som om, at han ender i "Mangelfuld" i stedet for "God" på grund af svigtende motivation hen mod slutningen især i forhold til lange tekster.

³ Svarende til $1,96 \cdot \text{SEM}$.