

Folkeskolens centralt stillede test



Peter Allerup, AU, DPU

Kommentar til "Undervisning og centralt stillede test i folkeskolen", MONA 2012(2)

Artiklen er skrevet i anledning af at der er indført nye prøveformer i faget biologi, de såkaldte *digitale prøver*. Man skal som læser der ikke omgåes prøver og test i det daglige, huske at adskille tre typiske prøveformer som har været anvendt og stadigvæk anvendes i folkeskolen: (1) en *gammeldags* med en fast samling opgaver skrevet ned på papir, stillet til alle elever med besvarelse inden for en fastlagt tid, (2) en *digital*, som også er en fast samling opgaver (ca. 20), men nu formidlet via computer, som eleverne besvarer inden for et fastsat tidsrum, og endelig (3) *nationale adaptive it-baserede* test, hvor eleverne besvarer en række opgaver formidlet via en computer på en sådan måde at opgavernes sværhedsgrad løbende tilpasses elevens dygtighedsniveau (beregnet løbende ved at se på de afgivne svar). Der er ikke nogen fast tidsramme for testen, idet den stopper når statistiske beregninger af elevdygtigheden har vist at usikkerheden på beregningen falder under en bestemt fastsat grænse (den såkaldte *standard error of measurement*).

Inden for faget biologi er prøveformen (2) indført i 2005, og (3) i 2010, og den gammeldags er dermed trådt lidt i baggrunden. De to forfattere har undersøgt de to nye prøveformer ved at spørge en række lærere i faget om deres meninger og holdninger og gennemføre en SWOT-analyse på de indsamlede data (SWOT står for Strengths, Weaknesses, Opportunities og Threats). Det er fint nok at spørge lærerne, eftersom det jo er dem der skal stå for skud i den daglige undervisning og føre eleverne frem til evalueringen, men man risikerer at få et snævert billede af fordele og ulemper når man spørger ind til holdninger og meninger ved at spørge så bredt og lidet præcist som de to forfattere har gjort i deres hovedspørgsmål: "Hvordan opfatter du indførelsen af afgangsprøven (multiple-choice) i biologifaget?", "Hvordan opfatter du indførelsen af de nationale test i biologifaget?" og "Hvilken centralt stillet prøveform afdækker bedst elevernes udbytte af din biologiundervisning?".

På den plads jeg har fået til at kommentere de to forfatters undersøgelser, vil jeg, punktvis, tage udgangspunkt i de *kvalitative* aspekter som svar fra lærerne har rejst,

idet undersøgelsen med kun 100 respondenter set fra et kvantitativt synspunkt ikke tillader at man konkluderer noget generelt om læreres holdninger og meninger om prøverne. Mine kommentarer vil være at pege på nogle, i mine øjne, åbenlyse *fordele* og *ulemper* ved prøveformerne (2) og (3).

Washback-effekter er jeg enig i det betænkelige ved – det er indskrænkende for den daglige undervisning hvis der øves for meget i prøveopgaver. Det kan dog undertiden være vanskeligt at undgå inden for curriculumorienterede fag (fag med læringsmål) og med prøveformer der forventes at afspejle kravene i disse læringsmål. At spektret er bredt, kan man illustrere ved at tænke på de “opgaver” som Bjarne Riis’ cykelryttere øver sig på for at løse opgaven “først over Alpe d’Huez” – ganske få af dem handler om at øve selve opgaven, som er strækningen – og, i den anden ende, de “opgaver” der skal løses undervejs for at indfri et læringsmål om at kunne de første 300 primtal udenad; her er total washback lig med total succes! Det siger sig selv at ved læringsmål og prøveformer der retter sig mod *kompetencer* (literacy-orienterede udfordringer a la PISA), bliver washback-effekter mindre synlige og sværere at identificere. Om orienteringen er curriculum eller literacy-baseret, afhænger ikke af faget, men af rammerne for beskrivelse af læringskrav og evalueringer.

Nationale adaptive it-baserede test opfatter jeg ikke, som forfatterne gør det, som et eksempel på en *formativ* evaluering. Der er jo tale om tre til fire normative “ben” i selve opgørelsen, hvor den underliggende skala bag hvert ben er langt fra det formative, men hvor selve *sammensætningen* til en “profil” bygget på benene måske kan siges at udgøre et formativt udtryk. Når man fx gør resultatet op i matematik, repræsenterer det ene “ben” færdigheder i “algebra”, det andet færdigheder i “geometri”, og det sidste ben (i matematik er der kun tre) gælder “matematisk anvendelse”. Den underliggende skala repræsenterer de såkaldt latente færdigheder og er en skala af samme type som fx PISA-skalaen (som har internationalt gennemsnit ved 500 med 100 som spredning). Problemet med denne prøveform set fra lærernes synspunkt ligger i mine øjne i at der ikke er to elever i en klasse der har besvaret de samme opgaver under samme testsession; alle har fået forskellige opgaver som resultat af den adaptive teknik. Læreren er dermed berøvet velkendte “trick” med at udnytte sammenligninger *mellem* eleverne for kunne udtale sig om en *bestemt* elev. Den adaptive teknik spiller set fra en statistisk teknisk synsvinkel (dvs. en psykometrisk) smukt sammen med den måde opgavebanken er bygget op på, med opgaver som i testsessionen kan sammensættes vilkårligt og alligevel i enden udgøre et homogent grundlag for elevvurderinger. Det smukkeste ved dette samspil var fra starten tænkt som et ønske om at give eleven en positiv testoplevelse (dvs. det adaptive der jo sikrer at opgaverne hele tiden er på “niveau med eleven”) og ikke at teste eleven i unødigt lang tid (stopkriteriet sikrer at

der ikke stilles flere opgaver end der teknisk set behøves for at beregne elevens niveau præcist). Men måske drukner skønheden i den daglige omgang med resultaterne af denne prøveform?

De digitale prøver fra 2005 afløser gammeldags papirprøver selvom disse måske ikke på landsplan har været anvendt i faget biologi. Til forskel fra de nationale adaptive test er der tale om et fast sæt opgaver (ca. 20, besvaret inden for 30 minutter) som ligger klar på nettet i en computer til 9.-klasses-eleverne. Det lå i starten som et "lokkende" element at eleven kunne få sin karakter umiddelbart efter afslutningen. Prøven er ikke adaptiv og opererer med et retteprincip der betyder at ikkebesvarede opgaver betragtes som fejl. Prøveformen ligner lidt de gammeldags papirprøver og lidt de nationale test, men er egentlig hverken fugl eller fisk. Måske mest fisk, vil nogle mene, og en slem én af slagsen: Når eleverne ser at tiden er ved at være brugt, og der endnu mangler besvarelser, så er det bare om at komme i gang med *at gætte* helt vildt (i multiple-choice-formatet er der ca. 25 % chance for rigtigt svar). Når du rejser dig op og *er glad* for karakteren 10 efter at maskinen har slået op i den omsætningstabel som jo nødvendigvis må være udarbejdet for forhånd, fx

Omsætningstabel biologi 2012	
Karakter	Antal rigtige svar
-3	0
0	1-15
2	16-19
4	20-25
7	26-32
10	33-41
12	42-50

så dæmper det måske glæden lidt ved senere at få at vide at ca. 70 % af kammeraterne også har fået karakteren 10 – alene fordi omsætningstabellen fra starten var "skæv" i forhold til prøvebesvarelsenerne! Så vidt vides, er denne lokkende del af tilbagemeldingen nu rent procedurermæssigt rullet tilbage til fordums *forcensur* hvor man kigger på ca. 1.000 besvarelser for at kunne lægge omsætningstabellen, så der kommer et mere beskedent antal på karakteren 10! Efter ca. 2 uger i spænding kan eleven så se hvad karakteren blev.

Supplerende mundtlige prøver er et forslag formuleret af lærere i undersøgelsen og videreformidlet af de to forfattere som jeg selv støtter ud fra en grundantagelse om at det er godt når lærere kan tage ejerskab over for et evalueringsredskab – det øger man chancen for succes med hele vejen rundt.

Som medforfatter til Dansk Clearinghouses (Nordenbo et al., 2009) påpegning af hvor lidt testresultater egentlig bliver anvendt i pædagogisk sammenhæng, må man bestandigt spørge sig selv om nytten af at gennemføre nye prøveformer. Er det for *systemets skyld* ud fra ønsker om at kende landsniveauet? Det kunne måske klares mere hensigtsmæssigt med en international prøve. Er det for *elevens skyld*? Studievalg efter 9. klasse afhænger jo ikke af resultaterne ved hverken nationale test eller digitale prøver, men nogle siger at det øger motivationen til at læse op mod slutningen af folkeskolen. Eleven kan sjældent bruge selve resultatet til at sige noget om succes eller ej til at gennemføre en ungdomsuddannelse (se Allerup, 2012). Lærerne har uden tvivl glæde af prøveresultaterne, især hvis de formidles som et *formativ feedback* til deres egen undervisning – og så gør det jo heller ikke noget at de nye digitale prøver er ekstremt billige at gennemføre sammenlignet med gammeldags papir, med censorer osv.

Referencer

- Allerup, P. (2012). Hvad kan folkeskolen gøre for at mindske frafaldet ved ungdomsuddannelserne? *Økonomi og politik*, 85-2, s. 43-58.
- Nordenbo, S.E., Allerup, P., Andersen, H.L., Dolin, J., Korp, H., Larsen, M.S. et al. (2009). *Pædagogisk brug af test – et systematisk review*. Aarhus Universitetsforlag.