

# Skal vi bruge krudt på at skyde en "smart abe" ned?

Eller bruge det som afsæt til at fokusere mere generelt på de udskældte, udanske multiple choice-test, som nu har vundet indpas i vores evalueringskultur?



Peter Weng, UC Metropol,  
København

*Kommentar til Sunde & Sunde: "Den smarte abe: betydning af og korrektion for gætning ved karaktergivning i multiple choice-test", MONA, 2016-4.*

Denne kommentar vedrørende multiple choice-test tager udgangspunkt i artiklen *Den smarte abe: betydning af og korrektion for gætning ved karaktergivning i multiple choice-test* forfattet af P. Sunde og P.B. Sunde. I artiklen skriver de at i en MC-test er der mulighed for at opnå et forudsigeligt antal svar ved gætning alene, og at denne "skavank" både bør og kan man korrigere for. Dette gættefænomen ved MC-test har der været fokus på gennem mange år; derfor er det nye i artiklen argumentationen for at korrigere for konsekvensen af dette fænomen. Begrundelsen er bl.a. at vi i Danmark, i modsætning til i mange andre lande, har en kriteriebaseret karaktergivning i grundskolen. Der argumenteres i artiklen ud fra en undersøgelse forfatterne har gennemført på grundlag af karaktergivningen ved afgangsprøven i biologi sommeren 2015. Deres undersøgelse viser at muligheden for at gætte i en MC-test medfører at testpersonerne systematisk vil svare rigtigt på et højere antal spørgsmål end der faktisk er dækning for ud fra testpersonens faglige viden. Testpersonerne får altså rigtigt ved spørgsmål som de faktisk ikke selv ved eller selv tror på korrektheden af ved deres afkrydsning. Specielt har denne gætteeffekt en betydning ved tildeling af karakterer til de testpersoner der har den svageste faglige viden.

Forfatterens argumentation for deres budskab er gennemført sobert. Den primære argumentation er "at gætning giver et betydeligt pointbidrag i situationer hvor eksaminander kun kender svaret på en mindre del af spørgsmålene", og argumentationen for hvorledes man kan korrigere for denne gætning ud fra bestemte givne forhold,

er overbevisende gennemført, både teoretisk og gennem beskrivelse af det konkrete MC-test-forløb. Så argumentationen, ud fra de præmisser forfatterne lægger til grund, giver ikke anledning til at drage konklusionerne i tvivl.

Men er gætteskavanken ved MC-test så stor at det er nødvendigt at korrigere for denne?

Eller sagt på en anden måde: Er den uretfærdighed der fremkommer på grund af gætteskavanken ved MC-test der især tilgodeser de fagligt svage elever, de "smarte aber", værd at korrigere for? Disse elever vil med stor sandsynlighed få en højere karakter end de burde have, men er denne skavank så stor i forhold til de mange andre skavanker vi ved der altid er ved test og karaktergivning?

Hvis man som jeg er enig med Peter Allerup i hans udtalelse til Information i marts 2014 – "Danmark er det eneste land, hvor man tror, at 7-trinsskalaen er ligesom et centimetermål, der tillader beregninger af et gennemsnit, som signalerer, hvor dygtig en studerende er" – er svaret på ovennævnte spørgsmål: Nej! Retfærdigheden/målesikkerheden ved 7-trinsskalaen er tvivlsom og kan svinge meget uanset typen af prøve/test og anvendte omsætningstabeller, såvel ved skriftlige som mundtlige prøver/test.

Der er altså mange usikkerhedsfaktorer ved karakterskalaer som 13- og 7-trinsskalaen vi har brugt i Danmark i de seneste mange år. Det betyder for mig at en påvisning af gætningens betydning for karakteren til de fagligt svageste elever i en MC-test bliver tvivlsom, ikke mindst når man som forfatterne vælger at se bort fra alle andre faktorer der kan give usikkerhed ved MC-test. I sig selv er det problematisk at korrigere usikkerhed (gætning) på noget der i sig selv er usikkert (7-trinsskalaen).

Men artiklen om "den smarte abe" kan måske være med til at vi får et større fokus på den stigende anvendelse af MC-test i den danske evalueringskultur, ikke mindst inden for naturvidenskab og matematik.

Af mulige fokuspunkter knyttet til MC-test som jeg mener er relevante, er følgende to områder. Det første er en belysning og diskussion af danske piger og drenge tilgang til MC-opgaver. Det andet område jeg også mener har interesse at få belyst betydningen af, er den ikke ringe andel af simple – og sammensatte – MC-opgaver der anvendes i de internationale undersøgelser TIMSS og PISA som danske elever deltager i.

Igennem flere årtier har der været flere undersøgelser der fokuserer på kønsforskelle i test. En del af fokuseringen har vedrørt betydningen af opgaveformaterne i test. Udgangspunktet i flere af disse undersøgelser har været spørgsmålet om korrektheden af udsagnet: Drenge får mere ud af at gætte i MC-test end piger. I de internationale undersøgelser TIMSS 2015 og PISA 2015 gælder det at der er signifikante forskelle i drengenes favør når det gælder matematik, og for naturvidenskab i PISA 2015 er billedet det samme på de fleste delområder. Generelt klarer drengene sig bedre end

pigerne i disse test. Har dette noget at gøre med at en ikke ringe andel af opgaver i de to internationale test er MC-opgaver?

Generelt er det måske en god idé at få drøftet anvendelse af MC-test i sammenhæng med både en norm- og kriteriebaseret karakterskala. En sådan drøftelse kunne fx tage udgangspunkt i følgende opstilling af fordele og ulemper fra Roberts, 2006, i min oversættelse:

Fordele:

- Teste viden hurtigt i store grupper
- Bruges til at give hurtig feedback
- Kan scores automatisk
- Kan analyseres for sværhed og gruppering af resultater
- Kan lagres i en opgavebank og genbruges efter behov.

Ulemper:

- Er tidskrævende at konstruere
- Tester kun hukommelsesviden og fakta
- Tester aldrig evnen til at kunne afkode, forstå og anvende tegn og evnen til at analysere
- Tester aldrig kreativitet eller unik tænkning
- Opmuntrer eleverne til overfladelæring.

Afslutningsvis kan man overveje om Niels Egelund har ret og fået ret i følgende udsagn som han fremkom med til Ekstra Bladet i maj 2013 (Ekstrabladet 2013) på baggrund af en på det tidspunkt ny undersøgelse der viste at 20 forskellige lærere havde bedømt en skriftlig dansk opgave med karakterer fra 02 til 12.

*Det er en forbandelse, der er ved de eksamensformer, vi har, at der er så stor forskel på bedømmelsen. Men det har vi lært at leve med, og det har vi levet med i lang tid.*

*Jeg kunne jo godt tænke mig, at der var en højere grad af standardisering, og at flere af eksamenerne i folkeskolen for eksempel foregik ved multiple choice og lignende. Så ville man i højere grad kunne udelukke den slags tilfældigheder.*

*Det er meget udansk at gøre det, og der har vi en tradition for i Danmark at lave meget bløde vurderinger, som er indbygget i hele vores system. Så når jeg foreslår mere multiple choice, er jeg udmærket klar over, at jeg har hele det pædagogiske apparat i nakken bagefter.*

Det kan jo godt være at lidt mere forskning om hvordan de før så udanske multiple choice-test har fået betydning og anvendelse i vores uddannelsessystem, kan give os alle et grundlag der kan kvalificere diskussion om deres anvendelse i både en formativ og summativ evaluering.

## Referencer

- Ekstrabladet (2013): Lokaliseret 10.01.2017 på <http://ekstrabladet.dk/nyheder/samfund/article4475031.ece>
- Information (2014): Lokaliseret 10.01.2017 på <https://www.information.dk/debat/2014/03/karakterer-dovne-umotiverede>
- Roberts, T.S. (2006). *The Use of Multiple Choice Tests for Formative and Summative Assessment*, Proceeding – ACE '06 Proceedings of the 8th Australasian Conference on Computing Education – v.52, s. 175-180, fundet den 6. januar 2017 på <http://www.crpit.com/confpapers/CRPITV52Roberts>.