

Aber eller elever: en kommentar



Svend Kreiner, Biostatistisk afdeling, Institut for folkesundhedsvidenskab, Københavns universitet

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

Indledning

Jeg har sagt det før, men lad mig bare gentage det: Der er forskel på elever og aber. Det betyder ikke at multiple-choice (MC)-opgaver ikke kan være et problem for folkeskolens prøver, men det er formodentlig et mindre problem end Peter og Pernille Sunde antyder.

Der er to årsager til at jeg ikke kan følge dem. For det første at karaktererne i prøverne ikke er kriteriebaserede karakterer således som forfatterne siger, og for det andet at den abe-adfærd som forfatternes model forudsætter, har meget lidt med virkeligheden at gøre.

Om pædagogiske test med MC-opgaver

MC-opgaver optræder hyppigt i mange forskellige pædagogiske test, og gætteri har derfor også været diskuteret i årtier i forbindelse med sådanne test. Hvis der gættes på den måde som forfatterne antyder, vil det afsløre sig tydeligt ved to forskellige forhold. For det første ved at svarene på disse opgaver er svagere korreleret til det som prøven skal måle, end svarene på opgaver hvor der ikke gættes. Og for det andet at chancerne for et korrekt svar aldrig når helt ned på 0 for selv uendeligt svage elever. Begge ting afprøves rutinemæssigt i forbindelse med afprøvningen af testene og vil føre til at opgaverne elimineres når det forekommer.

Det frikender naturligvis ikke afgangsprøverne fordi disse prøver benyttes uden forudgående afprøvning af om der er problemer med nogle af opgaverne. Det er derfor nødvendigt både at se på forfatternes præmisser og på deres argumenter for at vurdere om problemerne er store nok til at reagere på.

Konsekvenser af MC-opgaver

Ud over at MC-opgaver lægger op til en vis form for gætning, er det måske nyttigt at minde om at MC-opgaver har andre konsekvenser end gætterier. Den væsentligste er at opgaverne bliver meget lettere fordi eleven får noget hjælp til at løse dem. Det betyder at det ikke kun er indholdet af opgaven, men også de svarmuligheder man får stillet til rådighed, der afgør hvor vanskelig en opgave er. Mine erfaringer fra forcensuren i forbindelse med afgangsprøverne er at opgavestillerne er dygtige til at formulere opgaver hvor prøvens samlede sværhedsgrad er næsten den samme fra år til år, men at der ofte er forskel på opgavernes forventede sværhedsgrader og de sværhedsgrader der afspejles af hyppigheden af korrekte svar. Da slutresultatet kun burde afhænge af elevens dygtighed og indholdet af opgaverne, er det klart at MC-formatet kan være en forstyrrende faktor. En meget vanskelig opgave der skulle afsløre at eleven "udtømmende opfylder alle fagets mål med ingen eller få uvæsentlige mangler", kan fx gøres meget let hvis alle de forkerte svarmuligheder grænser til meningsløse.

Er afgangsprøverne kriteriebaserede?

Forfatterens præmis er at karaktererne i forbindelse med afgangsprøverne er kriteriebaserede. De citerer ministeriet korrekt, men det er faktisk ikke korrekt. Hvis man ser nærmere på hvad ministeriet også fortæller om karaktererne, er det let at se at ministeriet ubetænksomt taler med to tunger, og at der er tale om karakterer der er tættere på at være normbaserede end kriteriebaserede.

Ministeriet forklarer i et notat fra 2007 (senest opdateret den 26.2.2015, jf. Ministeriet for Børn, Undervisning og Ligestilling, 2015) at:

"Et af formålene med 7-trinsskalaen er at sikre en fordeling, som svarer nogenlunde til den internationale ECTS-skala."

Og at:

"Når man sammenligner ECTS-skalaens karakterfordeling for bestående præstationer med 7-trinsskalaen, forventes det, at af de eksaminander, der på landsplan og over tid består en given prøve, får:

- 10 procent karakteren 12
- 25 procent karakteren 10
- 30 procent karakteren 7
- 25 procent karakteren 4
- 10 procent karakteren 02."

Eller med andre ord: Da karakterskalaen blev indført i 2007, var der, på trods af ministeriets snak om kriteriebaserede karakterer, tale om et normbaseret karaktersystem. At man dengang hægtede etiketter på karaktererne der fx sagde at "Karakteren 7 gives for den gode præstation, der demonstrerer opfyldelse af fagets mål, med en del mangler", gør dem ikke kriteriebaserede. Karaktererne blev defineret ved en opdeling af pointskalaer der angav hvor mange opgaver eleverne kunne svare rigtigt på, således at ca. 30 % fik karakteren 7, ca. 35 % fik en dårligere karakter, og ca. 35 % en bedre karakter – uanset indholdet af opgaverne og den hjælp som eleverne fik i form af forskellige svarkategorier.

Jeg har i flere år (og senest i 2015 og 2016, jf. Damvad, 2016) hjulpet ministeriet med forcensuren og kan oplyse at det der har været centralt, har været at sikre at karaktererne i forbindelse med den seneste prøve svarer til karaktererne fra tidligere år. At et 10-tal i matematisk problemløsning i 2016 svarer til et 10-tal i samme fag i 2014. Af den grund er karaktererne hverken fugl eller fisk. Hverken normbaserede eller kriteriebaserede. De var normbaserede i 2007, men man har siden forsøgt at fastlægge karaktererne så de bliver sammenlignelige henover tiden. Hvis elevernes faglige niveau ændrer sig (og det er jo det som nogen forventer vil ske pga. skolereformen), skal dette afspejles i karakterfordelingerne. Fordelingen af karaktererne skal altså ikke længere svare til ECTS-fordelingen på 10-25-30-25-10 %.

Responsadfærd i forbindelse med MC-opgaver

Hvis gættteri i forbindelse med MC-opgaver foregår på den måde som forfatterens model påstår, er det let at forudse hvorledes svarene vil fordele sig. Det rigtige svar vil forekomme med større hyppighed end hver af de øvrige svarmuligheder fordi de rigtige svar dels forekommer hos elever der kan løse opgaven, og dels forekommer som konsekvenser af "heldige" gæt. De øvrige svarmuligheder vil til gengæld forekomme med nogenlunde samme hyppighed fordi de alle har den samme sandsynlighed for at blive valgt.

Det er her kæden ryger af hvis man tror at abe-modellen beskriver elevadfærd. Enhver der har arbejdet med MC-opgaver, vil vide at svarene aldrig fordeles sig på den måde. Ud over at den korrekte løsning typisk forekommer hyppigst, er der nogle enkelte forkerte svarmuligheder der forekommer med en vis hyppighed, mens andre svarmuligheder næsten aldrig benyttes. Sådanne svarfordelinger kan bedst forklares på følgende måde.

1. En del elever kan løse opgaven og vælger derfor det korrekte svar.
2. En anden del tror at de kan løse opgaven, men tager fejl og vælger derfor et forkert svar.

3. En tredje del kan ikke løse opgaven, men kan se at de forkerte svarmuligheder er forkerte. De vælger den korrekte løsning (Sherlock Holmes er vist kendt for at bruge denne metode).
4. En fjerde del kan ikke løse opgaven, men kan se at nogle af de forkerte løsninger ikke kan bruges. De gætter blandt det der er tilbage, ud fra subjektive vurderinger af hvad der er mest sandsynligt.
5. En femte del kan heller ikke løse opgaven, men udelader den korrekte løsning sammen med andre muligheder. De gætter derefter blandt det der er tilbage, men får i sagens natur et forkert resultat.
6. Og så er der til sidst aberne som kaster en mental terning og træffer et helt tilfældigt valg.

Gætteri kan føre til korrekte løsninger i det tredje og sjette tilfælde. I det sidste er det naturligvis forkert at kreditere eleven for et korrekt svar, men i det tredje kan man argumentere for at der er tale om kvalificerede gæt som skal honoreres. Og det er præcis det der sker. Hvis antallet af svarmuligheder der er tilbage, er begrænset, og hvis den subjektive sandsynlighed for det rigtige svar er stor fordi eleven har nogenlunde styr på hvad der foregår, vil der være relativt mange rigtige svar. Hvis der er flere svarmuligheder tilbage, og hvis de subjektive sandsynligheder for det rigtige svar er små, vil der scores relativt få point. De dygtige elever scorer altså bedre end de mindre dygtige, hvilket er hele idéen med prøven.

Det andet og femte tilfælde fører til forkerte svar, og gætteriet i det femte tilfælde hjælper ikke fordi gætteriet er "ukvalificeret".

Herefter er der kun den sidste mulighed. De elever der bruger denne strategi (som lærerne forhåbentlig har fortalt dem er en dårlig strategi) vil have samme sandsynlighed for alle svarmuligheder. Hvis der derfor er svarmuligheder der (næsten) aldrig bruges, betyder det at det kun er meget få elever der har brugt strategien, og at Sunde og Sundes bekymringer er ubegrundede.

Afsluttende kommentarer

Selvom gætteri ikke er et stort problem for folkeskolens prøver, betyder det ikke at MC-opgaver er uproblematisk. Men årsagerne er nogle helt andre end dem som artiklen diskuterer.

Det vil føre for vidt her at gå nærmere ind på det. Jeg vil nøjes med at henvise til en artikel af Ferrara og DeMauro (2006) som bl.a. diskuterer både problemer og muligheder i forbindelse med MC-opgaver. Ferrara og DeMauro giver bl.a. eksempler på MC-opgaver i matematik og viser hvorledes man kan udtrække viden om elevernes færdigheder ud fra forkerte valg i stedet for kun at registrere om svarene var rigtige

eller forkerte. Sådanne muligheder findes ikke i forbindelse med IT-baserede selvrettende prøver, og det er efter min mening et meget større problem for folkeskolens prøver end gætterier. Prøverne fungerer i realiteten som ikke-validerede pædagogiske test med den form for usikkerhed som sådanne test altid indebærer. Konsekvensen er bl.a. at der kan være en ikke ubetydelig risiko for at en elev der tager to forskellige prøver, vil få to karakterer der afviger med to eller flere trin på karakterskalaen (jf. Damvad Analytics, 2016).

Referencer

- Damvad Analytics. (2016). Forcensur ved folkeskolens 9. klasses afgangsprøver 2015. Lokaliseret 6. januar 2017 på <http://www.damvad.com/wp-content/uploads/2016/05/Endelige-rapport-Forcensur-2015-08.04.2016.pdf>
- Feerrara, S. & Demauro, G.E. (2006). Standardized AssesSment of Individual Achievement in K-12. I: Brennan, R.L. (red.), *Educational Measurement* (s. 579-622). Westport: American Council of Education & Praeger Publishers.
- Ministeriet for Børn, Undervisning og Ligestilling. (2015). Anvendelse af 7-trinsskalaen. Lokaliseret 6. januar 2017 på <http://www.uvm.dk/Uddannelser/Paa-tvaers-af-uddannelserne/7-trins-skalaen/Anvendelse-af-7-trins-skalaen?allowCookies=on>