

# Den smarte abe:

## betydning af og korrektion for gætning ved karaktergivning i multiple choice-tests



Peter Sunde, Institut for Bioscience, Aarhus Universitet



Pernille Bødtker Sunde, DPU, Aarhus Universitet, og VIA Læreruddannelsen i Aarhus

**Abstract:** *I forbindelse med eksamener baseret på multiple choice-tests er der en mulighed for at opnå et forudsigeligt antal rigtige svar ved hjælp af gætning alene. I forbindelse med kriteriebaseret karaktergivning (karaktergivning baseret på læringsmål) er det derfor nødvendigt at tage højde for at eksaminander systematisk vil svare rigtigt på et højere antal spørgsmål end de egentlig kender svarene på. Med udgangspunkt i et konkret eksempel (folkeskolens afgangsprøve i biologi, maj 2015) demonstrerer vi hvorledes gætning har en markant effekt på antal rigtige svar og karakterfastsættelsen i den lave del af karakterspektret. Vi viser endvidere hvorledes man kan korrigere for dette.*

### Introduktion

Multiple choice-tests (MCT's) er udbredt i mange uddannelsessystemer verden over og har i de senere år også fundet stigende anvendelse i Danmark, herunder i Undervisningsministeriets (UVM's) egne afgangsprøver for folkeskolen.

En MCT er kendetegnet ved at eksaminanden skal vælge et eller flere rigtige svar blandt en række mulige, og karaktergivningen baseres på andelen af rigtige svar. Det betyder også at det er muligt at gætte et svar hvis man ikke kender det i forvejen. Dermed vil en eksaminand altså i testen kunne få flere svar rigtige end det han/hun egentlig kender svaret på eller pga. tidsnød ikke kan nå at tænke over svaret til. Medmindre der i testen opereres med minuspoint for forkerte svar ("negative marking"), vil tilfældig gætning på spørgsmål man ikke kender svaret på, være en smart strategi som enhver eksaminand bør benytte sig af da det er en strategi man kun kan vinde på. I virkeligheden vil der også være en gråzone mellem det sikre svar og den blinde gætning idet eksaminanden ofte vil "chance" et svar han/hun ikke er 100 % sikker på er rigtigt.

Problemstillingen med gætning har været kendt og diskuteret lige så længe som MCT'en, og mange forskellige løsninger og korrektionsmetoder har været foreslået (Betts et al., 2009; Lesage et al., 2013). I denne artikel vil vi indskrænke os til at påpege

nogle af de problemstillinger som kan opstå når MCT's bruges i en dansk kontekst, nærmere betegnet i forbindelse med karaktergivning efter 7-trinsskalaen.

I angelsaksiske uddannelsessystemer hvor MCT's er vidt udbredt, er karaktergivning på basis af testens resultat ofte *normbaseret*, forstået således at den enkelte præstation vurderes i forhold til den samlede pointfordeling i eksamenspopulationen. Dvs. at en elevs besvarelse sammenholdes med alle elevers samlede pointfordeling.

Ved normbaseret karaktergivning er der derfor ikke behov for at korrigere de opnåede pointtal for at en del af svarene er opnået gennem gætning da de ekstra point opnået gennem gætning ikke ændrer på de forskellige eksamenspræstationers indbyrdes placering (forudsat alle eksaminander gætter svar på spørgsmål de ellers ikke kan svare på).

Den danske 7-trinsskala adskiller sig på et væsentligt punkt fra ovennævnte karaktergivningsprincip ved at være *kriteriebaseret*, forstået således at karakteren skal fastsættes ud fra i hvilken grad præstationen svarer til en given opfyldelse af fagets læringsmål. Beståelseskarakteren 02 ("den tilstrækkelige præstation") skal således gives for den præstation som demonstrerer den lavest acceptable grad af målopfyldelse (Undervisningsministeriet, 2007). Ligeledes er det i vejledningen præciseret at de øvrige karakterer over beståelseskarakteren 02 i eksamenssituationen skal gives efter faglige kriterier og ikke efter populationsnormen. Heri ligger (i det mindste implicit) også at graden af fagets målopfyldelse for en given præstation skal kunne udtrykkes i form af tilstræbt objektive kriterier og ikke blot som en relativ placering i pointfordelingen. I forbindelse med MCT's, som i sagens natur tester deklarative læringselementer ("viden": Kender man svaret på et lukket spørgsmål, eller gør man ikke?), skal karaktergivningen altså baseres på hvor stor en del af den adspurgte stofmængde eksaminanden bør kunne svare rigtigt på. Og da bliver andelen af rigtige svar som kan opnås gennem gætning, betydende, ikke mindst for de eksaminander som kun kender svaret på en mindre del af spørgsmålene. Det er derfor vigtigt at sondre mellem hvor stor en reel viden der ønskes for at opnå en given karakter, og hvor stor en del af de stillede spørgsmål denne viden vil resultere i rigtige svar på når den forventede "gætningsbonus" lægges til.

I det følgende vil vi redegøre for problemstillingen med rigtige svar baseret på gætning samt vise hvorledes man i forbindelse med fastsættelse af læringsmålsbaserede pointkriterier for karaktergivning kan korrigere for det forventede antal rigtige svar pga. gætning. For at illustrere problemstillingen har vi som praktisk eksempel undersøgt en af UVM's egne MC-prøver, folkeskolens afgangsprøve i biologi, maj 2015. Disse prøver er tilgængelige på ministeriets hjemmeside, og ud fra deres rettevejledning (Undervisningsministeriet, 2015) er det muligt at kvantificere præcis hvor stor en del af den testede viden en eksaminand skal besidde for at opnå karaktererne -3, 00, 02, 4 osv., altså videnskriterierne for at opnå de forskellige karakterer.

I analysen og diskussionen vil vi udelukkende forholde os til de matematiske og statistiske aspekter i forbindelse med gætnings betydning for andel rigtige svar i forhold til eksaminandens reelle paratviden og hvilken indflydelse dette kan have for den kriteriebaserede karaktergivning. Vi vil således ikke forholde os til andre relevante aspekter i forbindelse med MCT's såsom emnevalg, faglig sværhedsgrad eller muligheder for misforståelser i forbindelse med besvarelsene som tidligere har været diskuteret i dette tidsskrift (fx Andersen & Linderoth, 2012; Lauritsen, 2006; Allerup, 2012). Vi vil heller ikke forholde os til hvorvidt den reelle andel af svar som en eksaminand skal kende (uden at gætte) for at opnå en given karakter, er fagligt acceptabel fra et samfundsmæssigt synspunkt. Vores anliggende er alene at kvantificere og illustrere betydningen af tilfældig gætning for sandsynligheden for opnåelse af de forskellige karakterer samt foreslå praktiske løsninger til hvorledes man kan tage højde for dette i karakterudmålingen i forbindelse med kriteriebaseret karaktergivning.

## Teori

Rent matematisk vil det gennemsnitlige antal rigtige svar ( $N$ ) som en eksaminand opnår i en MCT, kunne beskrives som:

$$N = n \left( p + \frac{1}{k} (1 - p) \right)$$

hvor  $n$  er antal spørgsmål,  $p$  er andelen af spørgsmål som eksaminanden reelt kender svaret på, og  $k$  er antallet af svarmuligheder per test (hvis alle testspørgsmål har det samme antal svarmuligheder hvoraf ét er rigtigt). En tipskupon med 13 spørgsmål og tre svarmuligheder (1, X, 2) repræsenterer således en MCT hvor  $p$  er 0 (eksaminanden har ikke nogen forudgående forudsætninger for at vide hvad det rigtige svar er),  $n$  er 13, og  $k$  er 3. Ved hjælp af sypigetips (tilfældig gætning) kan man i gennemsnit forvente 4,333 rigtige svar per korrekt udfyldt tipskupon. Da udfaldet af hvert gæt er enten rigtigt eller forkert (dvs. at udfaldet er en binomialfordelt stokastisk variabel), er det principielt muligt at få alt fra 0 til 13 rigtige. Sandsynligheden for 0 rigtige er således  $(1-1/3)^{13} = 0,0051$  og for 13 rigtige  $(1/3)^{13} = 0,0000000063$ . Fra et evaluerings-synspunkt repræsenterer det forventede antal rigtige svar ved hjælp af sypigetips en "nul-fordeling" eller en "abe-fordeling", dvs. den fordeling af tilfældigt rigtige svar man vil få hvis man sætter en stor gruppe eksaminander helt uden faglige forudsætninger til at løse opgaven. Sat på spidsen kunne dette være en chimpanse som er trænet i at sætte kryds i en computerstyret menu, men som ingen idé har om hvad den svarer på. Ud fra principperne for 7-trinsskalaen skulle en sådan eksaminand takseres med karakteren -3. Følgelig bør en stringent karakterfastsættelse også tage udgangspunkt i at selv en "abe" vil forventes at opnå væsentligt mere end 0 point.

Det samme princip gør sig gældende i afgangsprøven i biologi. Forskellen er blot at antallet af svarmuligheder er større, og sandsynlighederne for at gætte sig frem til et rigtigt svar varierer fra 0,33333 til 0,04167 i de forskellige items (appendiks 1).

## Materiale og metoder

### *Folkeskolens afgangsprøve i biologi*

Folkeskolens afgangsprøve i biologi består af en elektronisk MCT hvor eksaminanden inden for den afsatte tid skal sætte kryds ved de rigtige svarmuligheder. Opgavesættet for maj måned 2015 bestod af i alt 20 opgaver med i alt 53 items (<http://www.uvm.dk/Uddannelser/Folkeskolen/Folkeskolens-proever/Censur-og-evaluering/Rettevejledninger/Tidligere-rettevejledninger?smarturl404=true>). Det maksimale pointtal var altså 53 rigtige svar. Ifølge UVM's omsætningstabel blev 0 rigtige svar tildelt karakteren -3, 1-18 rigtige svar blev tildelt karakteren 00, 19-24 rigtige svar blev tildelt karakteren 02, 25-30 rigtige svar blev tildelt karakteren 4, 31-37 rigtige svar blev tildelt karakteren 7, 38-45 rigtige svar blev tildelt karakteren 10, og 46-53 rigtige svar blev tildelt karakteren 12. En eksaminand skulle altså kunne give korrekt svar på mindst 36 % (19/53) af de stillede svarmuligheder for at opnå karakteren 02.

### *Simulering af eksamensforsøg med forskellig grad af viden og gætning*

Ved hjælp af en computersimulering hvor udfaldet af hver item-besvarelse (rigtigt eller forkert svar) afhang af sandsynligheden for at gætte rigtigt, simulerede vi resultatet af 25.000 opgavebesvarelser under forskellig grad af gætning, dvs. at vi lod "eksaminanden" kende svaret på 0 til 50 af de 53 items og gætte svaret på de resterende. Da chancen for at gætte sig frem til den rigtige svarmulighed i de forskellige items varierede fra 0,042 til 0,33 (appendiks 1), lod vi (for at lade testen være så konservativ som muligt) de simulerede elever starte med at kende svarene på de spørgsmål hvor den relative gevinst ved at gætte spørgsmålet var mindst, dvs. hvor der var færrest valgmuligheder. De første 13 items som en eksaminand kunne svare på i denne simulering, havde således sandsynligheden 0,333 (én rigtig ud af tre svarmuligheder) for at gætte rigtigt, det 14.-16. item havde sandsynligheden 0,29 (to rigtige ud af syv muligheder), det 17.-41. item havde sandsynligheden 0,25 (én ud af fire) for rigtigt svar ved en tilfældighed osv. Eller sagt med andre ord: 25.000 gange lod vi en elev gætte alle spørgsmål, 25.000 gange lod vi eleven kende ét svar på et item med én af tre valgmuligheder og gætte svarene på de resterende 52 items, 25.000 gange lod vi eleven kende svaret på to items med én af tre valgmuligheder og gætte svarene på de resterende 51 items og så fremdeles. Dermed genererede vi for hvert antal items eksaminanden kendte svaret på, en pointfordeling bestående af antal items med kendt svar plus de point som eksaminanden opnåede ved at gætte.

For di gætning er en tilfældig (stokastisk) proces, varierede antallet af point opnået ved gætning fra "eksamensforsøg" til "eksamensforsøg". Den absolutte pointspredning var naturligvis omvendt proportional med antallet af items der blev gættet. Som naturlig følge heraf fandtes den største pointspredning eksamensforsøgene imellem når alle 53 items blev gættet, hvor pointfordelingen var tilnærmelsesvis normalfordelt (fig. 1). Omvendt var pointspændet mellem forskellige eksamensforsøg pga. gætning indskrænket til 3 point når simuleringen var baseret på at eksaminanden kunne svaret på 50 af de 53 items.

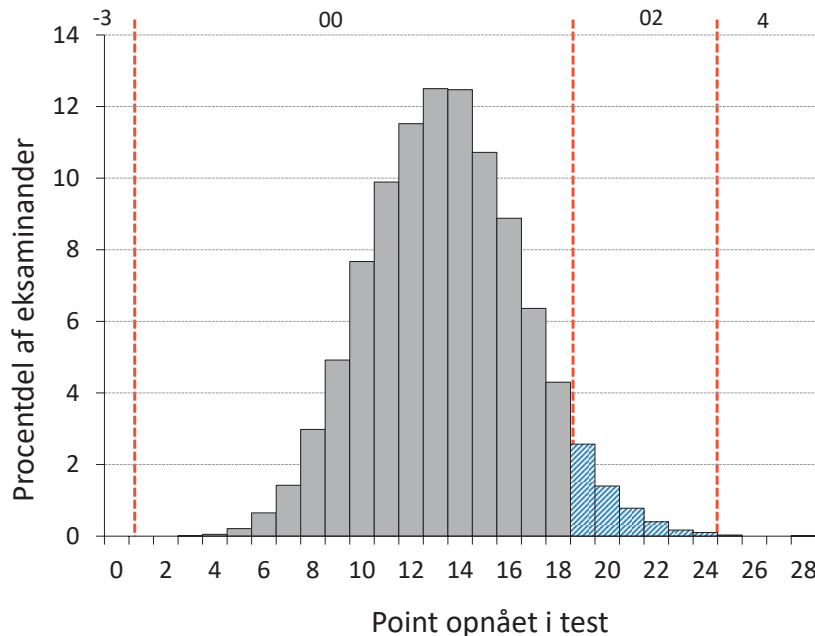
For at inkludere den af tilfældigheder fremkomne spredning (som følge af gætning) i evalueringen af eksaminandernes pointopnåelse angiver vi for hvert simuleringstrin den nedre percentil (dvs. det pointtal som skiller de nederste 1% fra de øverste 99% eksamensforsøg med identiske antal kendte svar), den nedre kvartil (det pointtal som skiller de nederste 25% fra de øverste 75%), medianværdien (det pointtal som skiller de nederste 50% fra de øverste 50%), den øverste kvartil (det pointtal som skiller de nederste 75% fra de øverste 25%) og den øverste percentil (det pointtal som skiller de nederste 99% fra de øverste 1%). Vi vil især lægge vægt på medianpointtallet da dette angiver hvilket pointtal halvdelen af eksamensforsøgene som minimum vil resultere i ved et givent vidensniveau.

Dermed kan vi sammenholde andelen af items med rigtige svar i testen med andelen af items eksaminanden reelt kunne svaret på. Ud fra dette kan vi estimere hvor stor en andel af de stillede items en eksaminand behøvede at kunne svaret på (eksaminandens "reelle" viden) for at have mindst 50% chance for at opnå en given karakter.

## Resultater

En simpel summation af sandsynlighederne for at gætte sig til det rigtige svar tilsiger at en eksaminand helt uden faglige forudsætninger som gættede på alle spørgsmål (en "smart abe"), i gennemsnit ville opnå 13,4 point i biologiprøven i maj 2015. En multiplikation af sandsynlighederne for at svare forkert på samtlige spørgsmål giver tilsvarende en sandsynlighed på  $1,6 \cdot 10^{-7}$ . En "abe" løb med andre ord en risiko på ca. én til fem millioner for ikke at opnå et eneste point og dermed opnå karakteren -3.

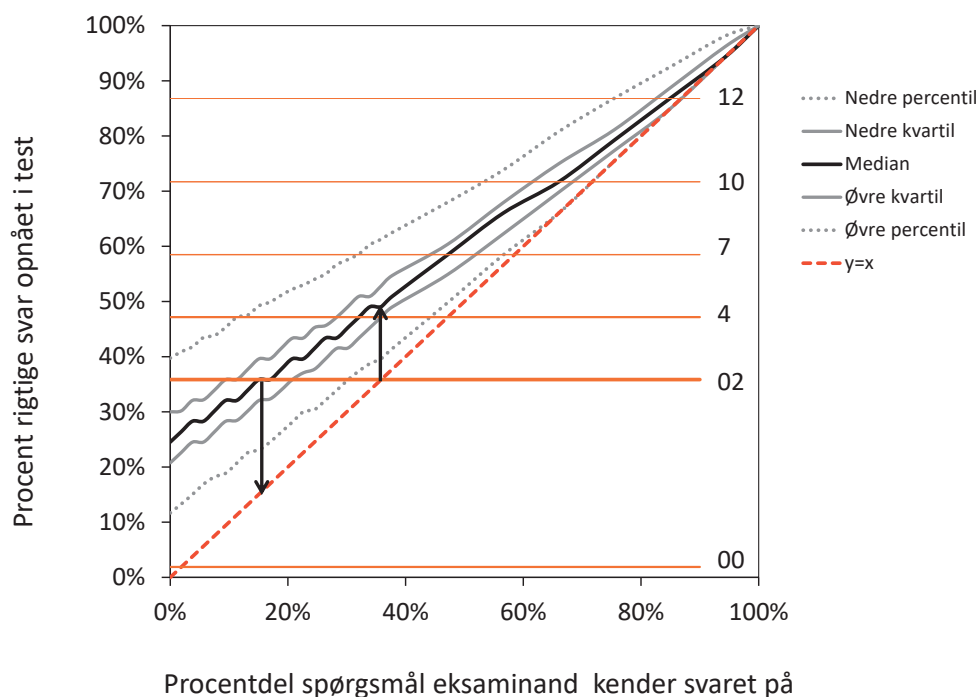
I computersimuleringen (fig. 1) opnåede 1.365 ud af 25.000 "aber" (5,5%) mindst 19 rigtige svar hvilket ville have udløst karakteren 02. Omtrent hver 18. "abe" som blev sat til at løse eksamensopgaven i biologi, ville med andre ord have held til at bestå prøven. I 11 tilfælde (0,4% eller én gang ud af 2.273) var "aben" endvidere så heldig at opnå mindst 25 rigtige svar og dermed et 4-tal. Oddsene for at en "abe" vil opnå karakteren 4, var dermed ca. 2.700 gange højere end for at få -3.



**Figur 1.** Pointfordeling af 25.000 simulerede eksamensforsøg i folkeskolens afgangsprøve i biologi, maj 2015, opnået alene gennem tilfældig gætning. De vertikale linjer angiver pointkriterier for opnåelse af karaktererne 00 til 4.

I figur 2 vises antal opnåede point for forskellige kvartiler af eksamensforsøgene som funktion af hvor stor en del af spørgsmålene de simulerede eksaminander kendte svarene på. Ved i grafen at aflæse x-værdien for medianpointlinjens skæring med  $y = 19/53 = 36\%$  (som er 9 ud af 53 svarmuligheder = 17%) fremgår det at en eksaminand for at have mere end 50 % chance for at opnå 19 point og dermed opnå karakteren 02 kun behøvede at kende det rigtige svar på 9 ud af 53 svarmuligheder (og gætte resten). Paratvidenskriteriet for at opnå karakteren 02 i folkeskolens afgangsprøve i biologi, maj 2015, var altså 17% (9/53) og ikke 36% (19/53) som UVM's pointfordelingsnøgle umiddelbart kunne give indtryk af (tabel 1).

Fra grafen kan man også aflæse at i fald det tiltænkte kompetencekrav i virkeligheden var ca. 36% korrekt viden, og UVM i sin pointnøgle ikke har korrigeret for gæt (vi har ingen viden om hvorvidt dette er tilfældet), skulle pointkriteriet for opnåelse af karakteren 02 have været 26 point eller en korrekt svarprocent på 49 (aflæses i figur 2 som medianlinjens y-værdi ved  $x = 36\%$ ) hvilket under de gældende karakterkriterier ville være tilstrækkeligt til et 4-tal (kriterium:  $25/53 = 47\%$  rigtige svar). Da også den nedre kvartil-linje kryber over x,y-koordinatet (36%, 47%), vil over 75% af alle eksaminander som kender svaret på 36% af spørgsmålene, altså opnå mindst 47% rigtige svar ved at gætte resten og dermed score et 4-tal.



**Figur 2.** Sammenhængen mellem procent rigtige svar i test (hvor eksaminanden gætter svar han/hun ikke kender) og procent svar som eksaminanden reelt kender svaret på (uden at gætte) i 25.000 simulerede eksamensforsøg i folkeskolens afgangsprøve i biologi, maj 2015. Linjen  $y=x$  angiver sammenhængen mellem antal rigtige opnåede svar og andelen af svar som eksaminanden reelt kunne svaret på uden at gætte.

De forskellige linjer angiver spredningen i pointfordeling pga. vekslende held med at gætte rigtigt: Den øvre percentil angiver således pointtallet som skiller de øverste (heldigste) 1% fra de underliggende 99%, den øvre kvartil angiver pointskellet mellem de øverste 25% og de underliggende 75%, medianen skiller de øverste og nederste 50%, den nedre kvartil skiller de nederste 25% fra de overliggende 75%, og den nedre percentil angiver pointgrænsen mellem de nederste 1% og de overliggende 99% af eksamensforsøgene. Intervalzonerne mellem den øvre og nedre kvartil og øvre og nedre percentil angiver således pointspændet for henholdsvis 50% og 98% af alle eksamensforsøg. De horisontale linjer angiver pointkriterierne for opnåelse af karaktererne 00-12. Det faktuelle vidensniveau (andel svar eksaminanden reelt kan svaret på) for at opnå en given rigtig svarprocent i testen med 1, 25, 50, 75 eller 99% sandsynlighed ( $y$ -værdien) kan aflæses på  $x$ -aksen. Omvendt kan andel rigtige svar i testen for et givent vidensniveau ( $x$ -værdi) aflæses på  $y$ -aksen.

Den vertikale pil ned angiver forskellen mellem andel rigtige svar i testen (36%) som skal til for at opnå karakteren 02, og andelen af spørgsmålene en eksaminand reelt be-



høver at kunne svare på (17%) for at have 50 % chance for at opnå 36 % rigtige svar ved at gætte.

Den vertikale pil op angiver hvor stor en andel rigtige svar en eksaminand i gennemsnit skulle have i testen (49 %) for at modsvare en paratviden svarende til 36 % af spørgsmålene (uden brug af gætning).

Som en naturlig konsekvens af at antallet af spørgsmål som gættes, aftager proportionalt med antallet af spørgsmål som eksaminanden kender svaret på (jf. formel), mindskes spændet mellem andelen af rigtige svar og andelen af svar som eksaminanden rent faktisk kan svare på, for de højere karakterer (tabel 1, figur 2).

Mens en eksaminand med 50 % chance for at få 02 kunne forvente at opnå 112 flere rigtige svar ved hjælp af gætning ( $[19/9] - 1$ ), var den tilsvarende relative "gættebonus" på 47 % for en eksaminand på vippen til et 4-tal ( $[25/17] - 1$ ), 23 % for en 7-tals-kandidat, 13 % ved grænseværdien for et 10-tal og 2 % ved grænseværdien for et 12-tal (tabel 1).

Eller sagt med andre ord, så er det de svageste eksaminander som opnår den største pointfordel ved gætning, mens 12-talseleven som kender næsten alle svarene på spørgsmålene, kun har minimal gevinst af gætning (af den simple årsag at der næsten ikke er nogen spørgsmål han/hun behøver at gætte svaret på).

Karakter	Minimum rigtige svar i test		Reel viden	
	Antal point	Andel	Antal svar	Andel
-3	0	0 %	0	0 %
00	1	2 %	0	0 %
02	19	36 %	9	17 %
4	25	47 %	17	32 %
7	31	58 %	25	47 %
10	38	72 %	34	64 %
12	46	87 %	45	85 %

**Tabel 1.** Pointtabel for karaktergivning i folkeskolens afgangsprøve i biologi, maj 2015, i form af antal og andel opnåede point (ud af 53 mulige) og den reelle viden (antal og andel rigtige svar ud fra viden) en eksaminand skal besidde for i gennemsnit at opnå dette pointtal når der gættes på spørgsmål man ikke kender svaret på (se også figur 2).



## Diskussion

To vigtige pointer kan udledes af denne analyse.

For det første viser resultaterne klart at gætning giver et betydeligt pointbidrag i situationer hvor eksaminanden kun kender svaret på en mindre del af spørgsmålene. Dermed vil fagligt svage elever kunne få et betydeligt utilsigtet karakterløft hvis der ikke tages højde for gætning i pointudmålingen. I det konkrete tilfælde betød dette at det i praksis er umuligt at opnå nulkarakteren -3 medmindre man som eksaminand gør en aktiv indsats for at svare forkert. UVM's egne tal (Undervisningsministeriet, 2016) viser da også at ingen eksaminander fik karakteren -3 i biologi ved majeksamen 2015, ligesom dette heller ikke var tilfældet for UVM's øvrige prøver hvor karaktergivningingen var baseret på en ren MCT. Derimod var det fuldt muligt (hver 18. eksamensforsøg) for en "abe", som per definition har et kundskabsniveau der skulle takseres til -3, at opnå beståelseskarakteren 02.

Hvad angår tærskelværdien for opnåelse af beståelseskarakteren 02 ("den tilstrækkelige præstation"), fremgår det også af analysen at det reelle kompetencekrav for at have mere end 50 % chance for at bestå eksamen svarer til at eksaminanden skal kende svaret på ca. 17 % af den paratviden der eksamineres i, og ikke 36 % som testens omsætningstabel umiddelbart giver udtryk for. I den helt lave ende af karakterudmålingen burde det stringente pointkriterie for at opnå 00 således rettelig have været 13 point og ikke 1 point da en "abe" i over halvdelen af tilfældene ville få mindst 13 point ved blind gætning.

Ud over at påpege at gætningsfaktoren giver et betydeligt pointbidrag for karaktergivningingen i den nedre ende af præstationsfordelingen, kan eksemplet også bruges til at illustrere hvorledes det er muligt at beregne hvor stor en reel viden en given pointsum svarer til og vice versa. For tilrettelæggere af MCT's er hermed også angivet hvorledes man i pointkriterierne for karaktergivning kan korrigere for forventet gætning. For brugere af testen vil en sådan beregning også give transparens i forhold til hvad de reelle kompetencekrav er for opnåelse af en given karakter. Dette kan være nyttig viden for de lærere som forbereder fagligt udfordrede elever til folkeskolens afgangsprøve i biologi, såvel som alle dem der måtte have en undervisningspolitisk interesse i at kende de faglige beståelseskrav til en given uddannelse.

Vi vil i det følgende knytte nogle supplerende bemærkninger til disse to diskussionspunkter. Om ikke andet så for at foregribe de mest forventede kritikpunkter som vores konklusioner vil blive mødt med.

Af forventelige modargumenter som også har været fremført tidligere (fx Downing, 2003; Nørgård, 2006), kan nævnes (1) at meget få eksaminander vil møde frem til en eksamen helt uden faglige forkundskaber, (2) at eksaminander ikke nødvendigvis vil gætte et svar de ikke kan, samt (3) at pointkorrektion for gætning ikke ændrer på den relative rangering af eksaminandernes pointfordeling.

Til det første modargument vil vi blot påpege at en stringent karaktergivning skal tage højde for de (sjældne) tilfælde hvor en eksaminand med utilstrækkelige faglige forudsætninger skulle forsøge sig mod de faglige odds. Hvad vigtigere er, så viser vores eksempel at gætning også giver et betydeligt merbidrag til pointsummen for de eksaminander som på et meget sparsomt fagligt grundlag er på vippen til at opnå karakteren 02. Dette skyldes det simple forhold at gættebonusen er omvendt proportional med andelen af spørgsmål eksaminanden reelt kender svaret på (fig. 2). Følgelig er det de eksaminander som møder op med de svageste faglige forudsætninger som vinder mest ved at gætte. Hvis man ikke er sig dette bevidst, betyder det at eksaminander består eksamen med meget ringere kundskaber end fastsat ud fra de egentlige målopfyldelseskriterier. Og jo lavere beståelseskravet er sat i forhold til andel rigtige svar, jo større bliver afvigelsen mellem det eksaminanden reelt kan, og det han/hun svarer rigtigt på.

Den markante pointeffekt af gætning på chancen for at opnå 02 (112 %) og 4 (47 %) i det aktuelle eksempel skyldes netop at pointkriterierne for opnåelse af disse to karakterer var mindre end 50 % rigtige svar (tabel 2). I eksamenssystemer hvor beståelseskriteriet typisk er 60-95 % rigtige svar (fx højere uddannelser eller den danske teoriprøve for kørekort eller indfødsret), er evalueringsproblemet pga. gætning tilsvarende mindre, grænsende mod ikkeeksisterende.

Til det andet modargument vil vi anføre at som en del af arbejdet med at forberede elever til den gældende eksamensform er det i princippet enhver lærers pligt at instruere sine elever i at gætte svar de ikke kender. Det er derfor op til testens tilrettelægger at sørge for at tage højde for at eksaminanden agerer optimalt under de givne testrammer. Som alternativ til at korrigerer for gætning (som her anvist) kan testen også tilrettelægges således at der gives strafpoint ved forkerte svar. Dette har dog den negative effekt at eksaminander ofte underpræsterer ved at afholde sig fra at besvare spørgsmål som de ikke er helt sikre på er rigtige (Betts et al., 2009). Det er nemlig vigtigt at huske at der i virkelighedens verden er en glidende overgang fra det stensikre svar givet uden tvivl over det kvalificerede gæt til det helt ukvalificerede gæt.

Hvad angår det forhold at pointkorrektur for gætning i forbindelse med pointgivning ikke ændrer på den normbaserede rangordning af eksaminandernes pointsum, er dette fuldstændig korrekt. Ifølge UVM's bekendtgørelse skal karakterer på 7-trinsskalaen imidlertid tildeles på grundlag af specifikke læringsmål (Undervisningsministeriet, 2007). Selv hvis man i praksis (og i givet fald mod bekendtgørelsens retningslinjer) tilpasser karakterfordelingen på basis af rangordning, så giver en ukorrigeret pointfordeling ingen information om hvor i pointfordelingen grænsen går mellem det minimalt fagligt tilstrækkelige og det utilstrækkelige (beståelseskriteriet). Helt konkret gælder dette for opnåelse af den laveste beståelseskarakter (02: "den tilstrækkelige præstation") som i princippet skal fungere som kvalitetssikring for hvor ringe

kompetencer en eksamineret studerende må sendes ud i samfundet med. Manglende opmærksomhed omkring det præcise kompetencemål for at kunne bestå en eksamen i en MCT vil dermed i værste fald kunne betyde at eksaminander består eksamener de rettelig ikke burde have bestået ud fra deres faglige standpunkt.

## Afsluttende bemærkninger

Vi har her demonstreret hvorledes man bør tage forbehold for gætningsfaktoren ved en kriteriebaseret karaktergivning baseret på MCT's, og hvor karakterkriterierne er baseret på mindre end 50-60 % rigtige svar. Denne analyse har taget udgangspunkt i en specifik case (folkeskolens 9. klasses afgangsprøve i biologi, maj 2015), men problemstillingen er alment gældende i alle sammenhænge hvor MCT's benyttes til kriteriebaseret karaktergivning (herunder også UVM's eksamener for 9. klasse i skriftlig fransk og geografi).

Vores anbefaling for dem som ønsker en stringent kriteriebaseret karaktergivning på grundlag af MCT's (hvilket alle som benytter 7-trinsskalaen, i princippet er forpligtet til), er at man korrigerer for antal rigtige svar opnået ved gætning i forbindelse med fastsættelse af pointkriterier for opnåelse af de forskellige karakterer. Dette kræver naturligvis at man i tråd med principperne for kriteriebaseret karaktergivning rent faktisk deklarerer hvor stor en reel vidensmængde (andel af spørgsmål som eksaminanden rent faktisk skal kende svaret på) der skal til for at opnå en given karakter. Når dette er besluttet, kan man så for den konkrete test (som i fig. 2) beregne eller simulere hvor mange ekstra point eksaminanden med en given sandsynlighed (fx 50 % svarende til medianværdien) kan regne med at score ved hjælp af gætning i tillæg til denne vidensmængde. Ud fra dette kan man så i karakterfastsættelsen korrigere for antal forventede rigtige svar opnået gennem gætning, som demonstreret i det konkrete eksempel.

Afslutningsvis skal det endnu en gang pointeres at karaktergivning baseret på inddeling i pointfraktiler efter rangordning, hvilket er gængs praksis i udlandet, reelt er et udtryk for normbaseret karaktergivning og dermed i modstrid med de oprindelige principper for karakterfastsættelse efter 7-trinsskalaen.

## Referencer

- Allerup, P. (2012). Folkeskolens centralt stillede test. *MONA*, 2012(3), s. 84-87.
- Andersen, P.U. & Linderoth, U.H. (2012). Undervisning og centralt stillede test i folkeskolen. *MONA*, 2012(2), s. 23-36.
- Betts, L.R., Elder, T.J., Hartley, J. & Trueman, M. (2009). Does correction for guessing reduce students' performance on multiple-choice examinations? Yes? No? Sometimes? *Assessment & Evaluation in Higher Education*, 34(1), s. 1-15.

- Burton, R.F. (2005). Multiple-choice and true/false tests: Myths and misapprehensions. *Assessment & Evaluation in Higher Education*, 30(1), s. 65-72.
- Downing, S.M. (2003). Guessing on selected response examinations. *Medical Education*, 37, s. 670-671.
- Lauritsen, H.J. (2006). En prøve i bakgear. *MONA*, 2006(1), s. 1-10.
- Nørgaard, K. (2006). De digitale afgangsprøver har høj kvalitet. *MONA*, 2006(3), s. 86-92.
- Lesage, E., Valcke, M. & Sabbe, E. (2013). Scoring methods for multiple choice assessment in higher education – Is it still a matter of number right scoring or negative marking? *Studies in Educational Evaluation*, 39(3), s. 188-193.
- Undervisningsministeriet (2007). *Bekendtgørelse om karakterskala og anden bedømmelse. BEK nr 262 af 20/03/2007*. Lokaliseret den 11. august 2016 på <https://www.retsinformation.dk/Forms/R0710.aspx?id=25308>
- Undervisningsministeriet (2015). *Omsætningstabeller biologi og geografi 2015*. Lokaliseret den 7. august 2016 på <https://www.uvm.dk/Uddannelser/Folkeskolen/Folkeskolens-proever/Censur-og-evaluering/Rettevejledninger/Tidligere-rettevejledninger>
- Undervisningsministeriet (2016). *Karakterfordeling i prøvefag 9. kl. (FP9), fordelt på fag og fagdisciplin, 2014/15*. Lokaliseret den 7. august 2016 på <http://statweb.uni-c.dk/Databanken/uvmdataweb/fullClient/Default.aspx?report=KGS-antkar-fag-kar&res=1280x819>

**Appendiks 1.** Sandsynlighederne for at svare rigtigt på de 53 spørgsmål og delspørgsmål (items) i folkeskolens afgangsprøve i biologi, maj 2015, (Undervisningsministeriet, 2015) ved hjælp af simpel gætning. Listen skal læses således: Item<sup>1</sup>: sandsynlighed for at svare rigtigt ved hjælp af gætning(antal rigtige svarmuligheder i item/samtlig antal svarmuligheder i item).

1: 0,2(1/5); 2: 0,2(1/5); 3-1: 0,25(1/4); 3-2: 0,25(1/4); 3-3: 0,25(1/4); 3-4: 0,25(1/4); 3-5: 0,25(1/4); 4: 0,2(1/5); 5-1: 0,25(1/4); 5-2: 0,25(1/4); 5-3: 0,25(1/4); 5-4: 0,25(1/4); 5-5: 0,25(1/4); 6: 0,04167(1/24); 7-1: 0,2857(2/7); 7-2: 0,1667(1/6); 8-1: 0,25(2/8); 8-2: 0,1429(1/7); 9-1: 0,25(1/4); 9-2: 0,25(1/4); 9-3: 0,25(1/4); 10-1: 0,25(2/8); 10-2: 0,1429(1/7); 11: 0,2(1/5); 12-1: 0,3333(1/3); 12-2: 0,3333(1/3); 12-3: 0,3333(1/3); 12-4: 0,3333(1/3); 13-1: 0,25(1/4); 13-2: 0,25(1/4); 13-3: 0,25(1/4); 13-4: 0,25(1/4); 13-5: 0,25(1/4); 14: 0,2(1/5); 15-1: 0,33333(1/3); 15-2: 0,33333(1/3); 15-3: 0,33333(1/3); 16-1: 0,2857(2/7); 16-2: 0,1667(1/6); 17-1: 0,3333(1/3); 17-2: 0,3333(1/3); 17-3: 0,3333(1/3); 17-4: 0,3333(1/3); 17-5: 0,3333(1/3); 18-1: 0,2857(2/7); 18-2: 0,1667(1/6); 19-1: 0,25(1/4); 19-2: 0,25(1/4); 19-3: 0,25(1/4); 19-4: 0,25(1/4); 20-1: 0,3333(1/3); 20-2: 0,25(2/8); 20-3: 0,1429(1/7).

1 "3-1" betyder første item i opgave 3 osv.

### English abstract

*Multiple choice tests provide the possibility of attaining a predictable number of answers correct by mere guesswork, which may inflate grading based on fulfillment of learning objectives unless corrected for. On the basis of a concrete example, we demonstrate how correct answers obtained by guesswork can have a substantial impact on the number of correct answers and grading for students knowing the answers of less than half of the questions posed. Simple guidelines are given for how grading criteria based on the number of correct answers can be adjusted in order to align to a given learning objective.*