



Hvordan kan strykprosenten ved eksamen stabiliseres?

ORIGINALARTIKKEL

ANDERS BARLI COLBERG

Fakultet for medisin og helsevitenskap

og

Institutt for klinisk og molekylær medisin

Norges teknisk-naturvitenskapelige universitet

Han har bidratt til utforming av studien, datainnsamlingen, analyse og tolking av data, litteratursøk og utarbeiding av manuskriptet.

Anders Barli Colberg (f. 1989) er medisinstudent.

Anders B. Colberg og Daniel Vatn har bidratt i like stor grad til denne artikkelen.

Forfatter har fylt ut ICMJE-skjemaet og oppgir ingen interessekonflikter.

DANIEL VATN

Fakultet for medisin og helsevitenskap

og

Institutt for klinisk og molekylær medisin

Norges teknisk-naturvitenskapelige universitet

Han har bidratt til utforming av studien, datainnsamlingen, analyse og tolking av data, litteratursøk og utarbeiding av manuskriptet.

Daniel Vatn (f. 1993) er medisinstudent.

Anders B. Colberg og Daniel Vatn har bidratt i like stor grad til denne artikkelen.

Forfatter har fylt ut ICMJE-skjemaet og oppgir ingen interessekonflikter.

RUNE STANDAL

Fakultet for medisin og helsevitenskap

Norges teknisk-naturvitenskapelige universitet

Han har bidratt til idé, utforming av studien, datainnsamling, tolking av data og revisjon av manuskriptet.

Rune Standal (f. 1963) er ph.d., senioringeniør ved IT-seksjonen.

Forfatter har fylt ut ICMJE-skjemaet og oppgir ingen interessekonflikter.

MARIA RADTKE

Institutt for klinisk og molekylær medisin

Norges teknisk-naturvitenskapelige universitet

og

Avdeling for nyresykdommer

St. Olavs hospital

Hun har bidratt til idé, utforming av studien, datainnsamling, tolking av data og revisjon av manuskriptet.

Maria Radtke (f. 1967) er ph.d., spesialist i indremedisin og i nyresykdommer, overlege og førsteamanuensis II.

Forfatter har fylt ut ICMJE-skjemaet og oppgir ingen interessekonflikter.

TOBIAS S. SLØRDAHL

E-post: tobias.s.slordahl@ntnu.no

Institutt for klinisk og molekylær medisin

Norges teknisk-naturvitenskapelige universitet

og

Avdeling for blodsykdommer

St. Olavs hospital

Han har bidratt til idé, utforming av studien, datainnsamling, tolking av data og revisjon av manuskriptet.

Tobias S. Slørdahl (f. 1982) er ph.d., forsker, førsteamanuensis II og lege i spesialisering.

Forfatter har fylt ut ICMJE-skjemaet og oppgir ingen interessekonflikter.

BAKGRUNN

Medisinstudiet ved Norges teknisk-naturvitenskapelige universitet (NTNU) har skriftlig eksamen én gang i året. Grensen for bestått er minst 65 % korrekte svar. Andel stryk varierer fra år til år. Vår hypotese var at varierende andel stryk skyldes forskjellig vanskelighetsgrad på eksamen. Vi undersøkte om relative standardsettingsmetoder ville redusere variasjonen i andel stryk uten å redusere gjennomsnittlig beståttgrense.

MATERIALE OG METODE

Cohens relative standardsettingsmetoder korrigerer for eksamens vanskelighetsgrad. De er enkle i bruk og er et alternativ til å sette en absolutt 65 % grense for bestått. Vi brukte data fra 34 eksamener ved medisinstudiet ved Norges teknisk-naturvitenskapelige universitet i perioden 2010–15 og sammenlignet andel stryk beregnet med dagens bedømmelsesmetode mot Cohens metoder.

RESULTATER

Med dagens beståttgrense på 65 % varierte strykandelen fra 0 % til 13,7 %, og andelen falt utover i studiet. Med unntak av eksamen i første studieår var andel stryk lavere og det var mindre variasjon i strykandelen med både original og modifisert Cohens metode sammenlignet med dagens metode. En av Cohen-metodene ga en strykandel fra 0 % til 10,4 %.

FORTOLKNING

Absolutt beståttgrense på 65 % kan i vårt datamateriale forsvares fordi andelen stryk generelt var lav. Cohens metoder kan være et alternativ ved studiesteder med en høy andel stryk, eller der det er svært store variasjoner fra år til år i andelen stryk på samme eksamen i studieløpet.

Eksamen er en viktig del av medisinstudiet og brukes som et verktøy for læring og for å måle at studentene kan det de skal (1). Siden eksamen fungerer som kvalitetskontroll for videre progresjon i medisinstudiet og for uteksaminering som lege med autorisasjon, er det nødvendig å sette en grense for hva som er godt nok eller grenser for ulike karakterer. Dette kalles standardsetting av eksamen.

Det finnes to former for standardsetting av eksamen, relative og absolutte metoder. Relative standarder tar utgangspunkt i en veldefinert gruppe, og beståttgrensen baseres på denne gruppens samlede prestasjon. Et eksempel på en relativ beståttgrense er gjennomsnittspoengsummen til gruppen minus ett standardavvik. Absolutte standarder

baseres på en forhåndsbestemt grense og er ikke avhengig av gruppens samlede prestasjon. Absolutte metoder egner seg dersom man vil undersøke om studentene har god nok kompetanse for et bestemt formål, som for eksempel studieprogresjon eller arbeid som lege. De to vanligste absolutte standardsettingsmetodene er Angoffs- og Ebels metoder. Begge er basert på at et ekspertpanel anslår vanskelighetsgraden på hvert spørsmål, og summen av disse anslagene settes som beståttgrensen på eksamen (2, 3). Beståttgrensen vil da variere fra prøve til prøve basert på prøvens vanskelighetsgrad. Bruk av ekspertpanel er ressurskrevende, både økonomisk, tidsmessig og organisatorisk. En enklere absolutt standardsettingsmetode er å ha lik beståttgrense fra år til år, men denne tar da ikke høyde for prøvens vanskelighetsgrad.

Ved Norges teknisk-naturvitenskapelige universitet praktiseres hovedsakelig en årlig integrert eksamen, som vil si at alle fag som er undervist i det aktuelle året testes i samme eksamen. Beståttgrensen er forhåndsbestemt til minst 65 % riktige svar, heretter kalt «absolutt 65 %». Andelen som stryker på en eksamen varierer fra år til år, uten at opptakskravene til medisinstudiet indikerer at det er stor forskjell mellom kullene (4). Siden det kun er én eksamen hvert studieår, er fallhøyden stor hvis man ikke presterer på denne eksamen. Utsatt eksamen (tidligere kalt kontinuasjonseksamen) er i august. Dersom studenten også stryker på denne eksamenen, må studieåret tas om igjen. Stryk har derfor store sosiale og økonomiske konsekvenser for studentene, og økonomiske og organisatoriske konsekvenser for fakultetet. Fordi eksamen både er kvalitetskontroll og læreverktøy, og fordi konsekvensene av stryk er store, må det stilles strenge krav til eksamenskvaliteten. For å ha god kredibilitet bør standardsettingsmetodene ta høyde for eksamenens vanskelighetsgrad (5).

Både relative og absolutte standardsettingsmetoder har svakheter. Absolutte metoder som benytter seg av ekspertpaneler, er ressurskrevende og vanskelig gjennomførbare. Relative metoder kan gi uforsvarlig lave beståttgrenser. Dersom noen studenter ikke leser til eksamen, vil dette trekke ned gjennomsnittet og dermed beståttgrensen. Et annet problem med mange relative metoder er at det alltid vil være noen som stryker, noe som kan føre til at også studenter med tilstrekkelig kunnskap vil stryke dersom prestasjonene til gruppen som helhet er sterk. Med utgangspunkt i disse svakheterne utviklet Cohen-Schotanus og van der Vleuten i 2010 en ny metode (5). De hevder at de faglig sterkeste studentene er én stabil faktor i standardsettingsprosessen. Disse studentene har lest og forstått pensum og prioritert studiene høyt, men de vil også påvirkes av en eksamens vanskelighetsgrad. Ved å bruke de faglig sterkeste studentene som referansegruppe utviklet de det som i dag kalles Cohens metode, heretter kalt «original Cohen» (5). Senere er denne metoden blitt revidert av andre (6).

Vi har sammenlignet dagens absolutt 65 %-beståttgrense med to ulike Cohens metoder (original og modifisert Cohen). Vi har undersøkt hvorvidt disse metodene påvirker andelen som stryker og standardavviket for denne andelen. Hypotesen vår var at eksamener med mange stryk var vanskelige, mens eksamener med få eller ingen stryk var enklere. Vi antok derfor at en Cohen metode som tar høyde for vanskelighetsgraden, ville gi færre stryk ved vanskelige eksamener og muligens flere stryk ved enklere eksamener, slik at standardavviket i andelen stryk totalt sett ble redusert.

Materiale og metode

DATASETET

Datasettet i denne studien består av eksamensresultater fra medisinstudiet ved Norges teknisk-naturvitenskapelige universitet. Skriftlig eksamen gjennomføres etter hvert 2. semester de første fire årene, ingen eksamen etter 9. og 10. semester, men etter både semester 11 og 12. Eksamen består av to deler, en del med 100–120 flervalgsoppgaver med 3–5 svaralternativer, og en kortsvars-/essaydel med 3–5 hovedtemaer. Flervalgsdelen teller 60 %, mens kortsvars-/essaydelen teller 40 % (7). Alle ordinære skriftlige eksamener fra og med

2010 til og med 2015 ble vurdert, og alle ble inkludert bortsett fra eksamen for 11. semester i 2010 som ble ekskludert på bakgrunn av manglende data. Hvilke fag som testes i de ulike eksamenene, er angitt i tabell 1. I gjennomsnitt var det 111 kandidater per eksamen og 34 inkluderte eksamenssett. Totalt er 3 779 besvarelser inkludert.

Tabell 1

Oversikt over hvilke fag studentene testes i for hver eksamen

| Studieår | Semester | Basalfag/parakliniske fag | Kliniske fag |
|-------------|---------------|--|---|
| 1. studieår | 1-2. semester | Cellebiologi Biokjemi Genetikk Histologi Embryologi Medisinsk terminologi Medisinsk historie Medisinsk etikk Bevegelsesapparatet Anatomi muskel, skjelett | Lege-pasient-kurs i allmennpraksis |
| 2. studieår | 3-4. semester | Nervesystemets oppbygning og funksjon Anatomi øre, øye, hals, genitalia Embryologi Medisinsk statistikk Genetikk Medisinsk etikk Mikrobiologi Immunologi Endokrinologi Nyrefysiologi Arbeidsmedisin Toksikologi/miljømedisin Farmakologi Patologi | Lege-pasient-kurs (avsluttes i januar) |
| 3. studieår | 5-6. semester | Patologi Mikrobiologi Farmakologi Klinisk kjemi Epidemiologi Atferdsmedisin Bilddiagnostikk Immunologi | Øre-nese-hals Oftalmologi Nevrologi Nevrofysiologi Fysikalsk medisin Onkologi Geriatrici Infeksjonsmedisin Hematologi Kardiologi Lungemedisin Thoraxkirurgi Gastroenterologi Gastrokirurgi |
| 4. studieår | 7-8. semester | Patologi Bilddiagnostikk Tropemedisin Sosialmedisin Mikrobiologi Farmakologi | Akuttmedisin Dermatologi Ortopedi Revmatologi Infeksjonsmedisin Psykiatri Obstetikk Gynekologi Pediatri Endokrinologi Nefrologi Urologi Plastisk kirurgi |

| Studieår | Semester | Basalfag/parakliniske fag | Kliniske fag |
|-------------|--------------|--|--------------|
| 6. studieår | 11. semester | Allmennedisin Arbeidsmedisin Geriatri Miljømedisin Samfunnsmedisin Epidemiologi Medisinsk statistikk Klinisk beslutningslære Helsetjenesteadministrasjon Helsetjenesteøkonomi Kvinnehelse Medisinsk etikk Rettsmedisin | |
| 6. studieår | 12. semester | Oppsummeringssemester | |

BEREGNING AV ORIGINAL OG MODIFISERT COHEN

Original Cohen tar utgangspunkt i poengsummen til studentene på 95-prosentilen og fastsetter beståttgrensen til 60 % av denne. I tillegg korrigerer den for at studentene kan gjette seg til korrekt svar. Formelen for original Cohen er (5): Beståttgrense = $cN + 60(N^* - cN)$, hvor c er estimering av andel korrekte svar som kan tilskrives gjetning, N er maksimal skår, og N^* er skår til 95-prosentilen. Vi har korrigert for gjetning med samme metode som Cohen-Schotanus gjorde i sin studie (personlig korrespondanse Cohen-Schotanus, 2016). Estimering av andel riktige besvarelser som kan tilskrives gjetning (cN) beregnes som følger: $cN = (0,33 \times A) + (0,25 \times B) + (0,20 \times C)$ hvor A , B og C er andelen spørsmål med henholdsvis tre, fire og fem svaralternativer. Original Cohen er laget med tanke på eksamener med bare flervalgsoppgaver. Siden eksamen ved Norges teknisk-naturvitenskapelige universitet både har flervalgs- og kortsvars-/essaydel, korrigerer vi for gjetning kun i flervalgsdelen, mens vi bruker 95-prosentilen av totalpoengsum i hele oppgavesettet (flervalgs- + kortsvars-/essaydelen).

Modifisert Cohen er utregnet med følgende formel (6): beståttgrense = $K \times Px$, hvor K er faktoren man multipliserer skåren til studenten på den gitte prosentilen Px med. Vi la inn ulike K -verdier (0,65, 0,70 og 0,75) og fikk tilhørende beståttgrenser og andeler studenter som stryker. Valg av K -verdier er basert på dagens 65 %-beståttgrense og undersøker derfor andel stryk rundt denne grensen. Taylor fant at poengsummen til studentene på 90-prosentilen var et bedre referansepunkt enn 95-prosentilen som brukes i original Cohen (6). I modifisert Cohen korrigeres det ikke for gjetning.

ANALYSER

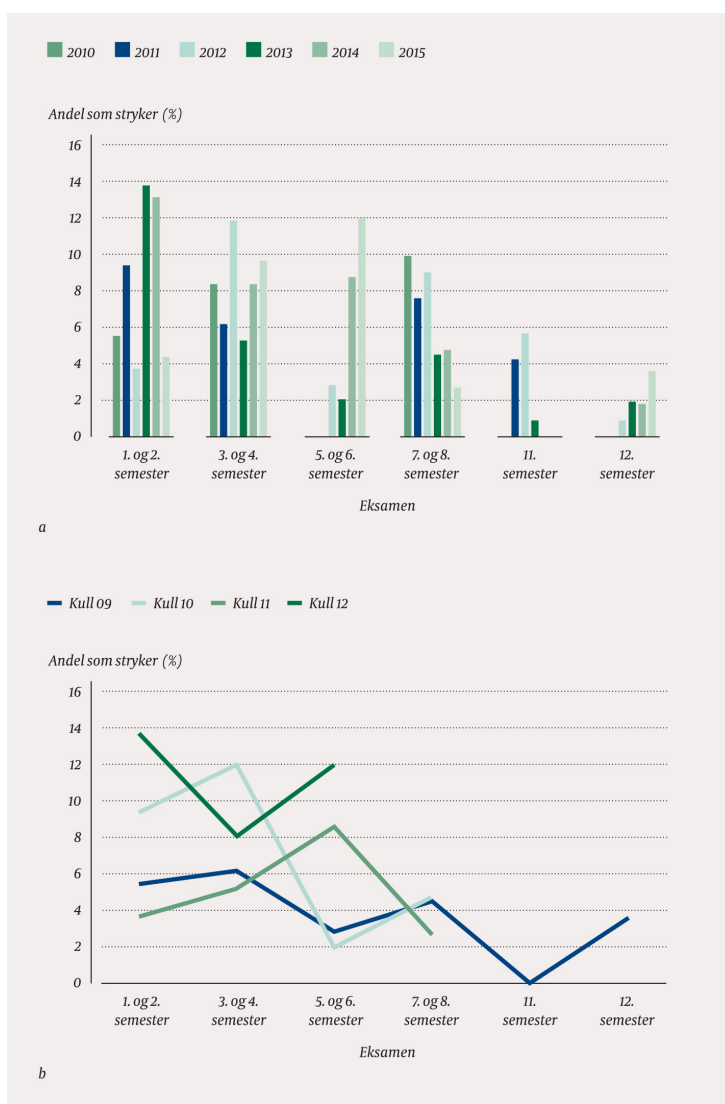
Følgende statistiske analyser og utregninger er gjennomført i Google Sheets 2016: gjennomsnitt, median, andel studenter som stryker, standardavvik (SD) av strykandelen, 90- og 95-prosentilen og korreksjon for gjetning.

ETIKK

Eksamensresultatene foreligger som anonymiserte data, og ingen enkeltpersoner kan identifiseres. Det ble derfor vurdert som ikke nødvendig å søke om tillatelse til å gjennomføre denne studien.

Resultater

Dagens absolutt 65 %-metode ga opptil 12 % differanse i andel stryk på samme eksamen i studieforløpet i den undersøkte perioden (fig 1, tab 2). Eksempelvis var det på eksamen på 5.-6. semester i 2010–11 ingen kandidater som strøk, mens det i 2015 var 11 (12,0 %) som strøk (fig 1a). Andelen studenter som strøk gikk ned underveis i studieforløpet (fig 1b) og var lik for alle standardsettingsmetodene.



Figur 1 a) Andel av medisinstudenter som har strøket til eksamen hvert studieår ved Norges teknisk-naturvitenskapelige universitet 2010–15 ved bruk av dagens metode for standardsetting med 65 % beståttgrense. b) Andel stryk (%) til fire kull fulgt over tid, kull 09 startet i 2009, kull 10 i 2010 osv. X-aksen viser hvilken eksamen data er hentet fra, mens Y-aksen viser antall stryk i prosent

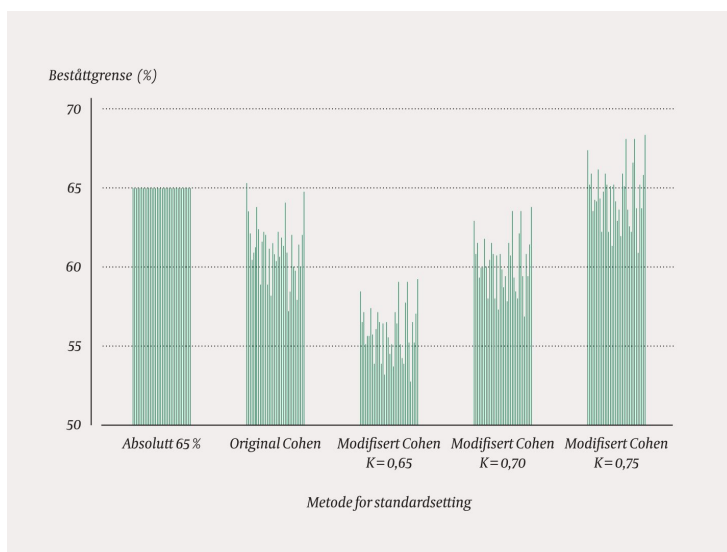
Tabell 2

Oversikt over eksamensdata for hver eksamen som er inkludert i artikkelen

| Semester | Eksamen | Antall kandidater | Gjennomsnittsskår | Medianskår | Antall stryk (%) |
|----------|---------|-------------------|-------------------|------------|------------------|
| 1. og 2. | 2010 | 110 | 7,7 | 80 | 6 (5,45) |
| 1. og 2. | 2011 | 117 | 76,9 | 79 | 11 (9,40) |
| 1. og 2. | 2012 | 108 | 77,9 | 78 | 4 (3,70) |
| 1. og 2. | 2013 | 117 | 78,6 | 76 | 16 (13,68) |
| 1. og 2. | 2014 | 115 | 74,2 | 77 | 15 (13,04) |
| 1. og 2. | 2015 | 114 | 76,3 | 77 | 5 (4,39) |
| 3. og 4. | 2010 | 121 | 74,4 | 74 | 10 (8,26) |
| 3. og 4. | 2011 | 113 | 77,4 | 78 | 7 (6,19) |
| 3. og 4. | 2012 | 118 | 73,7 | 77 | 14 (11,86) |
| 3. og 4. | 2013 | 114 | 77,4 | 78 | 6 (5,26) |
| 3. og 4. | 2014 | 109 | 75,1 | 77 | 9 (8,26) |
| 3. og 4. | 2015 | 114 | 76,1 | 77 | 11 (9,65) |
| 5. og 6. | 2010 | 103 | 80,1 | 81 | 0 (0,00) |
| 5. og 6. | 2011 | 110 | 83,3 | 84 | 0 (0,00) |

| Semester | Eksamen | Antall kandidater | Gjennomsnittsskår | Medianskår | Antall stryk (%) |
|----------|---------|-------------------|-------------------|------------|------------------|
| 5. og 6. | 2012 | 107 | 77,4 | 78 | 3 (2,80) |
| 5. og 6. | 2013 | 99 | 78,5 | 79 | 2 (2,02) |
| 5. og 6. | 2014 | 105 | 77,4 | 79 | 9 (8,75) |
| 5. og 6. | 2015 | 92 | 73,6 | 74 | 11 (11,96) |
| 7. og 8. | 2010 | 112 | 73,3 | 74 | 11 (9,82) |
| 7. og 8. | 2011 | 119 | 75,9 | 77 | 9 (7,56) |
| 7. og 8. | 2012 | 111 | 79,1 | 81 | 10 (9,01) |
| 7. og 8. | 2013 | 112 | 77,4 | 78 | 5 (4,46) |
| 7. og 8. | 2014 | 105 | 77,7 | 80 | 5 (4,76) |
| 7. og 8. | 2015 | 111 | 77 | 77 | 3 (2,70) |
| 11. | 2011 | 118 | 82 | 77 | 5 (4,24) |
| 11. | 2012 | 107 | 75,1 | 76 | 6 (5,61) |
| 11. | 2013 | 113 | 83,2 | 85 | 1 (0,88) |
| 11. | 2014 | 108 | 80,9 | 81 | 0 (0,00) |
| 12. | 2010 | 109 | 82 | 82 | 0 (0,00) |
| 12. | 2011 | 118 | 80,2 | 81 | 0 (0,00) |
| 12. | 2012 | 118 | 80,6 | 81 | 1 (0,85) |
| 12. | 2013 | 106 | 78,6 | 79 | 2 (1,89) |
| 12. | 2014 | 115 | 78,9 | 80 | 2 (1,74) |
| 12. | 2015 | 111 | 79 | 79 | 4 (3,60) |

Beståttgrensen ved original Cohen og modifisert Cohen med K-verdiene lik 0,65 og 0,70, var lavere enn 65 % absolutt. Bruk av modifisert Cohen med K-verdi på 0,75 ga en beståttgrense som svingte rundt dagens grense (fig 2). Ved original Cohen var beståttgrensen 57–65 % og ved modifisert Cohen 53–68 %, avhengig av hvilken K-verdi man brukte (tab 3).



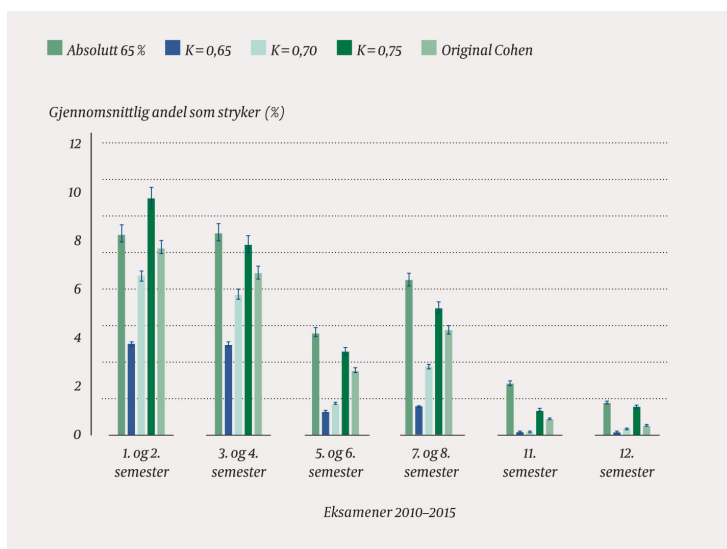
Figur 2 Sammenligning av beståttgrensene ved alle eksamener ved medisinstudiet ved Norges teknisk-naturvitenskapelige universitet 2010–15 ved bruk av de forskjellige metodene for standardsetting. Hver enkeltstøyle representerer en eksamen. Y-aksen viser beståttgrensen, X-aksen viser de ulike metodene

Tabell 3

Sammenligning av beståttgrense og andel stryk med standardavvik ved de ulike metodene for standardsetting (absolutt 65 %, original- og modifisert Cohens metode) for alle eksamener uavhengig av semester ved Norges teknisk-naturvitenskapelige universitet 2010–15

| Metode | Absolutt 65 % | Original Cohen | Modifisert Cohen K = 0,75 | Modifisert Cohen K = 0,70 | Modifisert Cohen K = 0,65 |
|---------------------------------------|------------------|-------------------|---------------------------------|---------------------------------|---------------------------------|
| Gjennomsnittlig beståttgrense (%) | 65,0 | 62,3 | 64,7 | 60,6 | 56,0 |
| Standardavvik beståttgrense | 0,0 | 1,5 | 1,5 | 1,4 | 1,3 |
| Variasjonsbredde beståttgrense (%) | 65,0 | 58,1–64,7 | 61,5–67,6 | 57,4–63,1 | 53,3–58,6 |
| Gjennomsnittlig andel stryk % (n) | 5,2 (6) | 3,9 (4) | 5,0 (5) | 3,0 (3) | 1,7 (2) |
| Standardavvik andel stryk | 4,2 | 3,7 | 4,4 | 3,1 | 2,1 |
| Variasjonsbredde andel stryk (%) | 0–13,7 | 0–13,7 | 0–19,7 | 0–10,4 | 0–8,5 |

Med unntak av eksamen i det første studieåret, var andelen stryk lavere med både original og modifisert Cohen sammenlignet med dagens metode (fig 3, tab 3). Original Cohen ga lik variasjonsbredde i andel stryk som absolutt 65 % (0–13,7 %), men reduserte gjennomsnittet av studenter som strøk totalt fra 5,2 % til 3,9 %. Modifisert Cohen med K-verdi på 0,65 og 0,70 reduserte standardavviket (SD) og gjennomsnittlig andel stryk totalt. Modifisert Cohen med K-verdi lik 0,70 har en gjennomsnittlig andel stryk på 3,0 % (SD 3,1) og en variasjonsbredde på 0–10,4 %, mens 65 % absolutt har gjennomsnittlig stryk på 5,2 % (SD 4,2) og en variasjonsbredde på 0–13,7 %. Modifisert Cohen med K-verdi på 0,75 ga en større variasjonsbredde i andel stryk enn absolutt 65 %, og dermed et høyere standardavvik (tab 3). Standardavviket for andel stryk ble lavere med lavere beståttgrense.



Figur 3 Gjennomsnittlig andel stryk i prosent inkludert standardavviket for hver eksamen ved bruk av de ulike metodene for standardsetting ved Norges teknisk-naturvitenskapelige universitet 2010–15. X-aksen viser hvilken eksamen i studieførløpet, mens Y-aksen viser den gjennomsnittlige andelen stryk

Diskusjon

Vi fant at andelen medisinstudenter som stryker varierer fra år til år på samme eksamen i studieløpet, men at det er en fallende andel studenter som stryker ut over i medisinstudiet. Vi har vist at standardavviket i andel studenter som stryker kan reduseres ved bruk av Cohens metoder, men at dette går på bekostning av en lavere beståttgrense.

Studien viser at det ut over i medisinstudiet er et fall i andelen som stryker. Vi har ikke undersøkt årsakene til dette fenomenet, men det er trolig flere. Medisinstudiet i Trondheim praktiserer spirallæring, det vil si at man i løpet av studiet kommer tilbake til samme tema omtrent hvert andre år. Eksempelvis har man det første studieåret hjertefysiologi med kliniske eksempler, det tredje studieåret kardiologi og det siste studieåret en oppsummering av kardiologien. Studentene vil i løpet av spirallæringen få en økende

forståelse for faget. Videre er studentene fra videregående skole vant til et klart definert pensum med hyppig testing. Medisinstudiet ved Norges teknisk-naturvitenskapelige universitet har derimot læringsmål med flere ulike anbefalte læreverker og kun 1–2 eksamener per år der alt kan testes. Overgangen i studieteknikk vil derfor være stor for mange, men det ser ut til at de fleste lærer seg å mestre dette gjennom studiet. I tillegg er det størst frafall de første to studieårene, da 2–6 må gå om igjen, mens maksimum én må gå om igjen de siste årene. Av de som slutter eller får inndratt studieretten grunnet gjentatte stryk, har 73 % (101 av 137 basert på tall fra 1999–2016) ikke fullført 2. studieår (personlig korrespondanse Mona Dalland Stormo og Marte Laugen, studieseksjonen ved Fakultet for medisin og helsevitenskap, Norges teknisk-naturvitenskapelige universitet). Andre faktorer som kan tenkes å bidra er eksamenserfaring, økende alder og at fagene oppleves mer relevante ut over i studiet, noe som kan øke motivasjonen.

I studien til Cohen-Schotanus sammenlignet de to kohorter ved to ulike medisinstudier i Nederland (5). Ved det ene fakultetet brukte man en referansebasert metode. Her ble gjennomsnittsskår minus et standardavvik benyttet for å sette grensen for bestått. Beståttgrensen varierte da mellom 15 % og 46 %, mens man hadde en relativt stabil strykandel på omtrent 17 %. Ved det andre fakultetet brukte de en forhåndsbestemt beståttgrense på 60 %, og strykandelen var 17–97 % (i gjennomsnitt 53 %). Man kan tenke seg at de studentene som hadde en høyere beståttgrense, ville være mer kunnskapsrike. Likevel var det ingen forskjeller i prestasjoner hos studentene i disse to kohortene på den nasjonale prøven (progresjonstesten) som gjennomføres på seks av åtte medisinstudier i Nederland (5). Studentene på fakultetet med en forhåndsbestemt beståttgrense på 60 % med høy andel stryk brukte imidlertid i snitt ett år lengre på å fullføre studiet. Tatt i betraktning at disse kohortene var like kunnskapsrike på den nasjonal prøven, taler dette for at forhåndsbestemte absolutte beståttgrenser er sløsing av offentlige ressurser, og ikke minst studentenes tid og ressurser (5).

Både original og modifisert Cohens metoder reduserte standardavviket i andelen som strøk. Ved bruk av disse metodene ville færre studenter ha strøket. Det overrasket oss at det motsatte, at flere studenter strøk på eksamener som ingen med absolutt beståttgrense strøk på, aldri var tilfellet. Vi mener at den Cohens metode man burde velge, er den som gir størst reduksjon i standardavviket for andel stryk, men som samtidig gir minst endring av beståttgrensen. Dette for å unngå å senke nivået på eksamen for mye, men samtidig forsøke å redusere variasjonen i strykandelen. I vårt materiale ville det vært modifisert Cohen med $K = 0,70$. Med denne setter man grensen for bestått til 70 % av poengsummen til studentene på 90-prosentilen.

Det er vanskelig å vurdere hvor god en standardsettingsmetode egentlig er, da det er vanskelig å vite nøyaktig hva den «sanne» beståttgrensen bør være for hver enkelt eksamen. Cohens metoder har den fordel at de er forutsigbare for studentene ved at de vet at de aldri trenger flere prosent riktig enn den oppgitte K -verdien (hvis 90 eller 95-prosentilen hadde hatt alt riktig). De vet også at metoden korrigerer for prøvens vanskelighetsgrad og at beståttgrensen ikke er en skjønsmessig vurdering i en eksamenskommissjon. La oss si at fakultetet bestemmer seg for at vi skal bruke Cohen med $K = 0,70$. Har du 70 % av poengsummen til 90-prosentilen, står du. Har du ikke det, stryker du.

En annen fordel med Cohens metoder sammenlignet med andre relative metoder er at den ikke gir en fast andel stryk. Vi synes det er noe problematisk å innføre en standardsettingsmetode som reduserer dagens 65 % absolutte grense for bestått. Selv om det er variasjon i andelen stryk, er det likevel et lite antall studenter som stryker på medisinstudiet hvert år sammenlignet med andre studier (5, 6).

Absolutte metoder med bruk av ekspertpaneler er nok det nærmeste man kommer den «sanne» beståttgrensen ved en medisinsk eksamen (2, 3). Dette benyttes mange steder, inklusive i den amerikanske nasjonale lisensprøven (USMLE) (8). I praksis vil dette være vanskelig å gjennomføre på alle eksamener på hvert enkelt medisinstudium. Ut ifra våre funn mener vi at en absolutt beståttgrense på 65 % kan la seg forsvare, så lenge andelen som

stryker er så lav som i dag. En standardsettingsmetode må ha kredibilitet. Blir variasjonen i strykandelen fra år til år for høy ved testing av en homogen gruppe studenter vurdert etter inntaksgrensene for medisinstudiet, taper eksamen kredibilitet (4). Cohens metoder bør brukes ved studiesteder med en svært høy andel stryk, eller der det er svært store variasjoner i strykandelen på samme eksamen i studieløpet. Ved Norges teknisk-naturvitenskapelige universitet tror vi metodene kan egne seg hvis strykandel ved eksamen avviker betydelig fra hva som er vanlig nå.

HOVEDBUDSKAP

Ved medisinstudiet på Norges teknisk-naturvitenskapelige universitet er det en absolutt grense på 65 % korrekte svar for å bestå eksamen, og andelen som stryker har variert fra år til år

Et alternativ til absolutt grense er Cohens metoder, som er relative og korrigerer for eksamenens vanskelighetsgrad

Forskjellene i andel stryk ble redusert ved bruk av Cohens metoder

Cohens metoder var enkle i bruk, men gikk på bekostning av en lavere grense for bestått

LITTERATUR:

1. Larsen DP, Butler AC, Roediger HL. Test-enhanced learning in medical education. *Med Educ* 2008; 42: 959 - 66. [PubMed][CrossRef]
2. Thorndike RL, Angoff WH. American Council on Education. Educational measurement. Washington, DC: American Council on Education, 1971: 768.
3. Ebel RL. Essentials of educational measurement. Upper Saddle River, NJ: Prentice-Hall, 1972: 650.
4. Samordna opptak. Søkertall og statistikk. www.samordnaopptak.no/info/om/sokertall/ (29.8.2017).
5. Cohen-Schotanus J, van der Vleuten CP, van der Vleuten CPM. A standard setting method with the best performing students as point of reference: practical and affordable. *Med Teach* 2010; 32: 154 - 60. [PubMed][CrossRef]
6. Taylor CA. Development of a modified Cohen method of standard setting. *Med Teach* 2011; 33: e678 - 82. [PubMed][CrossRef]
7. Retningslinjer for eksamen. Regler for obligatorisk undervisning vår/høst 2015. <https://www.ntnu.no/documents/10268/1265317082/Retningslinjer+skriftlig+eksamen+medisinstudiet+DMF+2015.pdf/9d1ce375-11d0-488a-8e53-74fc2a86a6f6> (29.8.2017).
8. Clauser BE, Mee J, Baldwin SG et al. Judges' use of examinee performance data in an Angoff standard-setting exercise for a medical licensing examination: an experimental study. *J Educ Meas* 2009; 46: 390 - 407. [CrossRef]

Publisert: 30. oktober 2017. Tidsskr Nor Legeforen. DOI: 10.4045/tidsskr.17.0025

Mottatt 6.1.2017, første revisjon innsendt 19.5.2017, godkjent 29.8.2017.

© Tidsskrift for Den norske legeforening 2020. Lastet ned fra tidsskriftet.no