Kan Birkebeinerennet lære oss hvordan vi definerer grensen for bestått på eksamen?

"Birkebeinerne" av Knud Bergslien, 1896(1)

Studenter:

Stud. Med Daniel Vatn

Stud. Med Anders Barli Colberg

Veileder/Prosjektleder:

Tobias Schmidt Slørdahl MD PhD, Post doc/Førsteamanuensis II, Institutt for kreftforskning og molekylærmedisin

Biveiledere:

Maria Radtke MD PhD, Førsteamanuensis, Institutt for kreftforskning og molekylærmedisin

Rune Standal PhD, Senioringeniør, DMF fakultetsadministrasjon

BAKGRUNN

Innen medisinsk utdanning brukes forskjellige standardsettingsmetoder for å separere bestått fra stryk, eller for å bestemme en karakter. På medisinstudiet ved Norges teknisk-naturvitenskapelige universitet (NTNU) er det én årlig integrert skriftlig eksamen som består av både flervalgs- og kortsvarsoppgaver. Graderingen er bestått/ikke bestått og går ut i fra en absolutt grense på 65 %. Erfaring har vist at et variabelt antall kandidater stryker fra år til år. Hypotesen var at dette skyldes varierende vanskelighetsgrad på eksamen. Vi ønsket å undersøke om Cohens metoder for standardsetting kunne bekrefte dette og derfor være bedre egnet enn dagens absolutte beståttgrense.

MATERIALE OG METODE

Materialet består av eksamensdata for 34 eksamener på profesjonsstudiet i medisin ved NTNU, avlagt i perioden 2010-2015, og består av 3779 unike eksamensresultater. Vi har sammenlignet resultater ved bruk av standardsettingsmetodene Cohens metode og modifisert Cohens metode, med resultater fra dagens absolutte grense på 65 %.

RESULTATER

Strykraten med en absolutt beståttgrense på 65 % varierte mellom 0 % og 13.7 %, og det var et signifikant fall i strykrate mellom første, og siste studieår ($P<0.005$). Variansen i stryk for hver eksamen var 1.88-24.25 med absolutt beståttgrense på 65 %. Ved å benytte standard Cohen og modifisert Cohen (med $K=0.65$, $K=0.70$ og $K=0.75$) fant vi en varians mellom 0.25-16.14, 0.12-6.09, 0.19-7.50, 0.79-33.41 og en beståttgrense i prosent mellom 58.1-64.7, 53.3-58.6, 57.4-63.1 og 61.6-67.6 respektivt.

FORTOLKNING

Forskjellen i strykrate ble redusert av standard Cohen og modifisert Cohen med $K<0.75$, på bekostning av en lavere grense for bestått. Uten en god kriteriebasert metode for å bestemme korrekt beståttgrense som kontroll, er det vanskelig å benytte disse metodene på våre data. Den absolutte beståttgrensen på 65 % har en akseptabel varians, men det kan være aktuelt å bruke en modifisert Cohens metode på eksamener med ekstreme strykrater. Både standard og modifisert Cohen er kostnadseffektive og enkle metoder om man vil benytte en norm-basert standardsetting som korrigerer for eksamenens vanskelighetsgrad.

Bakgrunn

Hvert år uteksamineres det i overkant av 600 leger fra norske universiteter (2). For å vite om en kandidat sitter med nok kunnskap til å ta gode medisinskfaglige beslutninger, er eksamen et av de beste virkemidlene en utdanningsinstitusjon har. Det er store forskjeller i andelen som stryker på en eksamen fra år til år, uten at opptakskravene til medisinstudiet skulle tilsi at det er noen stor akademisk forskjell mellom kullene. Det er derfor nærliggende å tenke at det er en forskjell i vanskelighetsgraden på eksamen som gir opphav til disse forskjellene. I denne artikkelen vil vi belyse temaet standardsetting og sammenligne to relativt nye metoder for standardsetting; Cohens metode (heretter kalt standard Cohen), og den modifiserte Cohens metode (heretter kalt modifisert Cohen). Standardsetting referer til prosessen der man setter grenser mellom kategorier som bestått og ikke-bestått, eller karakterer (f.eks. A-F). En standard bestemmer om en gitt poengsum er god nok for et gitt formål. Utfordringen med standardsetting er at det ikke finnes noen gullstandard for hvordan dette skal gjøres (3–6). Standardsetting er derfor en prosess hvor man leter etter en grense som ikke er for streng, men likevel fornuftig, rettferdig og reproduserbar.

Det finnes mange ulike metoder for standardsetting, hvor hovedskillet går mellom relativ (normbasert) og absolutt (kriteriebasert) standardsetting (7). Relative metoder for standardsetting tar utgangspunkt i prestasjonene til en definert gruppe, og setter beståttgrensen basert på denne gruppens poengsum. Absolutte standarder er uavhengig av resultatet på en gitt test, og kan settes ved hjelp av et ekspertpanel som går gjennom hvert enkelt spørsmål. På bakgrunn av dette kalkuleres det hvor grensen for bestått skal settes (8). Angoff og Ebel er eksempler på to anerkjente absolutte standardsettingsmetoder (8–10). Bruk av ekspertpaneler er ressurskrevende, og vanskelig gjennomførbart ved en integrert eksamen hvor mange fagfelt testes samtidig. Ved mange utdanningsinstitusjoner brukes derfor en forhåndsbestemt prosentandel som må besvares riktig. Tradisjonelt blir den relative standardsettingsmetoden sett på som den beste metoden for å rangere studenter, mens den absolutte er best på å bestemme om den enkelte student har tilstrekkelig med kunnskap (7). Bruken av de to modellene på samme eksamen kan gi store forskjeller (11,12).

På skriftlig eksamen ved NTNU praktiseres en absolutt metode hvor grensen for bestått (heretter kalt beståttgrense) er satt til 65 %. Ved bruk av en slik forhåndsbestemt beståttgrense har man sett relativt store forskjeller i andelen kandidater som stryker fra år til år. For eksempel var det 5 % av studentene som strøk (heretter kalt strykrate) til førsteårseksamen i 2012, mens det i 2013 var 13 % (13). Ettersom

opptakskravene til medisinstudiet er relativt stabilt fra år til år (66.0 ordinær/59.2 primær i 2011 (14) og 66.2 ordinær/59.0 primær i 2012 (15)), er det mindre sannsynlig at det er en forskjell i studentpopulasjonene som bidrar til variasjonen. Det er derfor rimelig å anta at det er vanskelighetsgraden på eksamen som varierer fra år til år. For studentene vil det i praksis si at noen består eksamen uten å kunne pensum tilstrekkelig, mens andre stryker selv om de kan nok.

På medisinutdanningen ved Universitet i Oslo (UiO) benyttes det en forhåndsbestemt 65 % grense for bestått. Her har de en praksis der de ofte korrigerer på beståttgrensen avhengig av hvordan studentene har prestert (personlig korrespondanse med Kristin Wium, UiO). Ved Universitet i Tromsø (UiT) lages det en sensorveiledning som angir hvor mange poeng som gir bestått, videre vurderes hver enkelt besvarelse av fem sensorer som gjør en helhetsvurdering (16). Ved UiT har man også muligheten til å senke kravene dersom man ser at mange stryker (personlig korrespondanse med Elin Holm, UiT). Ved Universitet i Bergen (UiB) brukes A-F karakterer, hvor E er laveste beståttkarakter, og det er opp til hver enkelt eksamensansvarlig i de ulike fagene å fastsette karaktergrensene (personlig korrespondanse med Arne Tjølsen, UiB). UiB og UiO reviderer for tiden sine eksamensregler og standardsettingen kan derfor bli endret..

Cohens metode ble utviklet med ønske om å kombinere fordelene med absolutte og relative standarderer, samtidig som man eliminerer ulempene. Den kan sammenlignes med Birkerbeinerrennet. I rennet ser man på gjennomsnittstiden til de fem beste, legger til 25 % og regner ut hva tiden på merket blir det aktuelle året (17). Akkurat som at skiføret på Birkerbeinerrennet ikke er likt hvert år, vil en eksamens vanskelighetsgrad, reliabilitet og validitet også variere. Standard Cohen ble utviklet av Cohen-Schotanus og Van der Vleuten i 2010 (3). Under utviklingen av Cohens metode oppdaget de at det var én stabil faktor; de best presterende studentene. De brukte derfor resultatene til de studentene med høyest poengsum som referansepunkt. De flinkeste studentene som har mestret og forstått hele pensum, vil i mye mindre grad bli påvirket av vanskelighetsgraden på eksamen. Cohen-Schotanus og Van der Vleuten tok derfor utgangspunkt i poengsummen ved 95 % persentilen, og beregnet beståttgrensen til 60 % av denne, i tillegg til at de korrigerte for gjetning (3). Denne metoden gir studenter en viktig forutsigbarhet. Ettersom den høyeste poengsummen ikke kan være over 100 %, vet studentene på forhånd at dersom de klarer 60 % korrekt (etter korreksjon for gjetning), vil man bestå.

Taylor CA (4) tok et steg videre og foreslo den modifiserte Cohen i 2011 . Den modifiserte Cohen skiller seg ved at den tar utgangspunkt i poengsummen til 90 % persentil-studenten,og beregner beståttgrensen til en viss prosent (-*K*), av denne. Det korrigeres ikke for gjetning. Taylor CA fant i sitt datamateriale at 90

% persentilen var mer stabil enn 95 % persentilen (4). Det eneste man da mangler for å finne beståttgrensen er konstanten, K . Taylor CA fant K ved å se på tre eksamener hvor det ble brukt Angoff som standardsettingsmetode, og beståttgrensen var lik (4). Ut fra dette kunne man finne gjennomsnittsprestasjon til 90 % persentilen på disse tre eksamenene, og dermed rearrangere og løse likningen for å finne en K - verdi som ga likest mulig strykrate som ved Angoff-metoden. Taylor CA appliserte modifisert Cohen på 32 moduler, og fant en reduksjon av forskjellene i strykrate mellom moduler, her sammenliknet med en fiksert beståttgrense på 50 % (4).

I denne studien sammenligner vi hvordan standard Cohen og modifisert Cohen påvirker forskjellene i strykprosent og beståttgrense på ordinære eksamener på medisinstudiet ved NTNU.

Materiale og metode

Beskrivelse av datasettet

På profesjonsstudiet i medisin har man skriftlig eksamen alle studieår, med unntak av femte. Eksamen skal i hovedsak omhandle læringsmål knyttet til det gjeldende studieåret, men inntil ⅓ av eksamens totale vekt kan omhandle læringsmål fra tidligere semestre. Eksamen skal være faglig integrert og kan omfatte alle basale, klinisk-medisinske, atferdsmessige og mijømessige emner som faller innenfor læringsmålene. Skriftlig eksamenen består av 100-120 flervalgsoppgaver (FVO) med ett riktig svar. Hver oppgave har mellom tre og fem svaralternativer, samt en kortsvars-/essaydel med tre-fem hovedtemaer. Flervalgsdelen teller 60 %, mens kortsvars-/essaydelen vektes 40 % (18). Navnet på eksamener i denne artikkelen viser til hvor i studiet eksamen foreligger. IAB er eksamen etter 1. studieår. Mens IIID er Embedseksamen i 6. studieår.

Datasettet i denne studien består av eksamensresultater ved Det medisinske fakultet ved NTNU. Alle ordinære skriftlige eksamener for profesjonsstudiet i medisin, fra og med 2010 til og med 2015 er inkludert. I gjennomsnitt var det 111 kandidater pr. eksamen. Dersom det forelå klage på et resultat, har vi tatt utgangspunkt i den endelige poengsummen. Alle innleverte besvarelser er tatt med i datasettet, med unntak av IIIC-eksamen i 2010, som ble ekskludert på bakgrunn av manglende data.

Beregning av standard og modifisert Cohen

Vi appliserte standard og modifisert Cohen på eksamensresultatene og beregnet strykrate og varians i strykrate, før vi sammenlignet dette med den den absolutte prefikserte standardsettingen som brukes i dag.

Formelen for standard Cohen er: (3);

$$Beståttgrense = cN + 60(N^{*} - cN)$$

hvor c er en estimering av andel korrekte svar som kan tilskrives gjetning, N er maksimal skår, og N^{*} er skår til 95 % persentilen. Ved standard Cohen har vi korrigert for gjetning med samme metode som Cohen-Schotanus og Van der Vleuten gjorde i sin studie (personlig korrespondanse med Janke Cohen-Schotanus);

$$Cn = (0,33\ x\ A) + (0,25\ x\ B) + (0,20\ x\ C)$$

hvor Cn er den andelen av eksamen som kan tillegges gjetning, og A, B og C er andelen spørsmål med henholdsvis tre, fire, og fem svaralternativ. Partiell kunnskap, altså evnen til å utelukke gale svar, er ikke medberegnet. Standard Cohen er laget med tanke på rene flervalgsoppgaveeksamener. Vi måtte derfor finne en måte å applisere den på en eksamen som kombinerer flervalgs- og kortsvarsoppgaver. Vi gjorde dette ved å korrigere for gjetning på flervalgsdelen og finne 95 % persentilen av totalpoengsum (flervalg+kortsvar). Høy grad av korrelasjon mellom resultatene på flervalgsoppgavedelen og kortsvarsoppgavedelen (median 0.63 alle eksamener) rettferdiggjør en slik tilnærming. (19). Formelen for modifisert Cohen er (4);

$$Beståttgrense = K \times Px$$

hvor K er faktoren man multipliserer skåren til studenten på den gitte persentilen Px med. Vi valgte, som Talyor CA, å bruke 90 % persentilstudenten. Vi laget en kalkyle hvor man legger inn ulike K -verdier og få tilhørende resultater som strykrater og beståttgrense. Resultatet for tre av K -verdiene; 0.65, 0.7 og 0.75 presenters i denne artikkelen. Verdiene er valgt på bakgrunn av at vi ønsket å undersøke hva som skjer med strykrate med en beståttgrense i området rundt dagens grense på 65 %.

Taylor plottet en kumulativ tetthetsfunksjon (cumulative density functions, CDFs) på eksamener, og vurderte om andre studenter responderte på vanskelighetsgraden på samme måte som 95 % persentilen,

både innenfor og på tvers av eksamener. På bakgrunn av dette mente Taylor at et mer hensiktsmessig referansepunkt vil være der det er en tydelig endring i gradienten av den kumulative tetthetsfunksjonen, noe som kan avgjøres ved å identifisere persentilen der den andre ordens deriverte er på et lokalt minimum. På datasettet til Taylor gjaldt dette for 90 % persentilen. Taylor viste at 90 % persentilen var konsistent over tid ved å se på 59 flervalgsoppgaver som var blitt gjenbrukt i seks andreårs moduleksamener, og så at prestasjonen til 90 % persentilen varierte i liten grad fra år til år (4). Taylor kom med forslag om at komponentene i en modifisert Cohen bør vurderes før lokal gjennomføring, om man vil bruke korrigering for å gjette, hvilken persentil man skal ta utgangspunkt i og hva multiplikatoren/konstanten, *K*, skal være. Vi lagde egne CDF kurver og fant tilsvarende resultater som Taylor, og har derfor valgt å bruke 90% persentilen i vårt datasett (supplerende fig 1).

Analyser

Følgende statistiske analyser og utregninger er blitt gjennomført på hver enkelt eksamen: gjennomsnitt, median, strykrate, varians i strykrate, 90 % persentil, 95 % persentil, korreksjon for gjetning, intern konsistens (KR20) for flervalgsoppgaver og Pearsons produkt-moment korrelasjonskoeffisient mellom kortsvar og flervalgsoppgaver.

Kuder-Richardson Formula 20 (KR20), er et mål på intern reliabilitet for binære målinger (det er kun ett riktig svar på hver FVO oppgave), for å se om samme binære (riktig/feil) resultat samstemmer over en populasjon med testobjekter. Med andre ord indikerer KR20 om man vil forvente å få samme resultat om studentene hadde tatt testen igjen. KR20 utregnes med følgende formel:

$$KR20 = (\frac{N}{N-1})(\frac{1-\Sigma(p_i q_i)}{\sigma^2})$$

Hvor N er antall spørsmål, p_i er andel av kandidater som har besvart spørsmålet korrekt, q_i er andel som har besvart spørsmålet feil, og σ^2 er variansen til eksamen.

Samtidig med å se på hele eksamen (FVO og korstsvarsoppgaver) har vi også gjort de samme beregningene på alle eksamenene med utgangspunkt i kun flervalgsoppgavene. Dette innebærer at beståttgrensen for forhåndsbestemt absolutt metode blir på 65 % av totalpoengsummen som flervalgsoppgavene gir. Bakgrunnen for dette valget er at vi ønsker å kunne si noe om standardsetting på rene flervalgseksamener, og fordi standard Cohen opprinnelig er ment for eksamener med kun flervalgsoppgaver.

Etikk

Eksamensresultatene foreligger som anonymiserte data og ingen enkeltpersoner kan identifiseres. Det ble derfor vurdert som ikke nødvendig å søke om tillatelse til gjennomføring av denne studien. Studien innholder en interessekonflikt ved at det er en studentoppgave som potensielt kan påvirke hvordan man skal standardsette eksamener ved Det medisinske fakultet i Trondheim. Dette kan muligens også gjelde forfatternes fremtidige eksamener.

Resultater

Ujevn strykrate med dagens standardsetting

Dagens standardsettingsmetode ga opptil 12 % differanse i strykrate på samme eksamen i den undersøkte perioden. Eksempelvis var det på eksamen i 3. studieår (IIAB) i 2010 og 2011, 0 (0 %) kandidater som strøk, mens det i 2015 var 11 (11,96 %) (fig 1.a). Tilsvarende spredning så vi når vi kun studerte flervalgsoppgavedelen av eksamen (fig 1.b). Korrelasjonen i strykraten mellom hele eksamen og kun flervalgsoppgavedelen var svært god (0.72). Gjennomsnittspoengsum viste også forskjeller på samme eksamen fra år til år, men denne var betydelig mindre enn variasjonen i strykrate (tab 1). Det var en god omvendt korrelasjon (-0.74) mellom gjennomsnittlig poengsum og strykrate.

Strykraten faller utover i studieløpet

Det var en nedadgående tendens i strykrate jo lengre ut i studieforløpet man kommer (fig 1.a og 2.a). Det var et signifikant fall (p=0.04) i strykrate fra 1. (9.5 %) til 6. studieår (0.8 %). Tendensen var tilsvarende hvis man kun studerte flervalgsoppgavedelen (fig 2.a og 2.b).

Standard Cohen og modifisert Cohen gir en lavere grense for bestått

Beståttgrensen når man bruker standard Cohen og modifisert Cohen med de to laveste K-verdiene (0.65 og 0.7), var lavere enn dagens forhåndsbestemte beståttgrense på 65 %. Bruk av modifisert Cohen med K-verdi lik 0.75 ga en beståttgrense som svingte rundt dagens 65 % grense (fig 3.a). For alle eksamener samlet, varierte beståttgrensen mellom 57-65 % ved standard Cohen, 61-68 % ved modifisert Cohen

(K=0.75), 57-64 % ved modifisert Cohen (K=0.70) og 53-59 % ved modifisert Cohen (K=0.65) (tab 2.a). Tilsvarende funn ble gjort på flervalgsoppgavedelen, men her ga standard Cohen en lavere beståttgrense i forhold til de andre metodene.

Ser man på gjennomsnittlig beståttgrense for alle eksamenene sett under ett, var gjennomsnittlig beståttgrense lavere ved kun flervalgseksamen, sammenlignet med tilsvarende metode på hele eksamen (tabell 2.b). Samtidig ser man at variasjonsbredden i beståttgrense er høyere for flervalgseksamen enn for hele eksamen.

Standard Cohen og modifisert Cohen påvirker strykratens varians

Varians i strykrate for eksamen med absolutt metode varierte mellom 1.9 og 24.3 (fig 5.a). Den absolutte metoden ga to topper i varians, både ved 1. års eksamen (IAB) og ved 2. årseksamen (IIAB), mens de andre metodene i mindre grad ga dette. Unntaket var modifisert Cohen med K-verdi lik 0.75 som ga en varians på 33.41 ved 1. årseksamen (IAB). Gjennomsnittlig varians for alle eksamener samlet, var høyest for absolutt metode på 10.78 sammenlignet med de andre metodene, hvor varians var på 10.07, 3.10, 2.14 og 6.31 på henholdsvis modifisert Cohen med K=0,75, K=0.70, K=0.65 og standard Cohen. Jo lavere beståttgrensen ble, dess lavere ble varians i strykrate.

Standard Cohen og modifisert Cohen påvirker gjennomsnittlig strykrate

Tendensen til fallende gjennomsnittlig strykrate utover i studieforløpet var lik for alle standardsettingsmetodene. Med den absolutte metoden falt den gjennomsnittlige strykraten fra 8.28 på første studieår til 1.35 på siste studieår. For standard Cohen falt den fra 7.70 til 0.34 (fig 4.a). Med unntak av eksamen i første studieår, var strykraten lavere med både standard og modifisert Cohen sammenlignet med dagens absolutte metode (fig 4.a og tab 2.a). Standard Cohen ga i dette tilfellet en lik spredning i strykrate som den absolutte metoden (0-13.7 %), men reduserte gjennomsnittet av studenter som strøk totalt (5.21 % til 3.90 %). Modifisert Cohen med K-verdi på 0.75 ga et større sprik i strykrate enn den eksisterende absolutte grensen, samt en høyere standarddeviasjon. En K-verdi på 0.65 reduserte antall stryk, og ga en svært lav gjennomsnittlig varians. En K-verdi på 0.7 reduserer den gjennomsnittlige strykraten totalt, og viser nest laveste varians i strykrate. Sammenlignet med 65 % absolutt metode gir modifisert Cohen med K-verdi lik 0.7 en lavere total gjennomsnittlig strykrate (henholdsvis 2.96 og 5.21), en lavere standardeviasjon (henholdsvis SD=3.06 og SD=4.23), og samtidig en mindre variasjonsbredde i

strykrate (henholdsvis 0-10.4 og 0-13.7). Den gjennomsnittlige strykraten på flervalgsdelen av eksamen falt også gjennom studieløpet (fig 4.b). Til forskjell fra eksamen samlet, følger den gjennomsnittlige strykraten for hver eksamen med modifisert Cohen med K-verdi lik 0.75, den gjennomsnittlige strykraten for dagens absolutte metode (fig 4.b).

Ser man på gjennomsnittlig beståttgrense for alle eksamenene sett under ett, var gjennomssnittlig beståttgrense lavere ved kun flervalgseksamen, sammenlignet med tilsvarende metode på hele eksamen (tabell 2.b). Samtidig ser man at variasjonsbredden i beståttgrense er høyere for flervalgseksamen enn for hele eksamen.

Diskusjon

I denne studien har vi sammenlignet tre standardsettingsmetoder; den tradisjonelle forhåndsbestemte absolutte standardsettingen, og to nye standardsettingsmetoder kalt standard Cohen og modifisert Cohen. Den forhåndsbestemte absolutte metoden brukt ved medisinstudiet ved NTNU viser relativ store forskjeller i strykrate på samme eksamen. Strykraten faller utover i studieløpet. Bruk av standard Cohen og modifisert Cohen med K-verdi lik 0.65 og 0.70 gir en lavere gjennomsnittlig strykrate og en lavere varians i strykrate.

Ved bruk av modifisert Cohen med K-verdi lik 0.75 får man også en lavere gjennomsnittlig strykrate, men en større variasjonsbredde av strykraten. Dette skyldes 1. årseksamen (IAB) i 2013. Det samme ser man igjen på variasjonsbredden i strykrate for K-verdi lik 0.75, som går fra 0-19.66 %. Med forhåndsbestemt 65 % beståttgrense er strykraten på denne eksamenen 13.68 %. Modifisert Cohen med K-verdi lik 0.75 gir en beståttgrense som øker med 0.25 % til 65.25%, dette fører til en økning i strykrate til 19.66 %. For resterende eksamenener er utslaget mindre, og metoden gir en lavere varians for de fire siste eksamenene i studieløpet sammenlignet med absolutt metode.

Når man skal benytte seg av Cohens metode, standard eller modifisert, må man ta stilling til hvem som skal utgjøre referansegruppen. Ved bruk av CDF kurver fant vi tilsvarende resultater som Taylor (4). En svakhet ved denne studien er at vi ikke kan sammenligne med en absolutt og kriteriebasert metode som Angoff (9) eller Ebel (10) . Dermed vil ingen av de brukte metodene fastslå hva som er minste kunnskapsgrense for bestått basert på ekspertuttalelser. Å vite hvor den sanne beståttgrensen bør ligge er derfor vanskelig.

De to Cohen-metodene skiller seg fra hverandre ved at standard Cohen korrigerer for gjetting, mens modifisert Cohen ikke gjør det (20). Integrerte eksamener på medisinstudiet ved NTNU bruker "number right scoring". Dette betyr at man får uttelling for alle korrekte svar, men ikke trekk for feil svar. Dette legger opp til at man kan gjette seg frem til riktige svar (4). Det finnes flere metoder for å korrigere for gjetning, men dette kan være problematisk. Det vil så godt som aldri være 20 % sannsynlig at studenten gjetter riktig på et spørsmål (gitt fem svaralternativer), ettersom studenter ofte vil kunne ekskludere ett eller flere av alternativene. Dette kalles partiell kunnskap (21). Andre metoder enn korrigering for gjetting egner seg trolig bedre for å hindre at studenter tipper seg til bestått på eksamen. Man kan enten øke konstanten K i den modifisert Cohen (4), eller øke antallet oppgaver (7).

I denne studien har vi vist at strykraten på en enkelt eksamen kan være relativt ulik fra år til år. Disse forskjellene er likevel betydelig mindre enn de man hadde i studien til Cohen-Schotanus, hvor det eksempelvis var en variasjonsbredde i strykrate på 17-97 % når de brukte en forhåndbestemt 50 % absolutt beståttgrense (3). Vi fant at enten vi benyttet standard Cohen eller modifisert Cohen, ga dette en lavere gjennomsnittlig strykrate på de enkelte studieår, og på alle eksamener sett under ett. Ved bruk av standard Cohen og modifisert Cohen med K-verdi lik 0.65 og 0.7, ga dette med få unntak lavere varians i strykraten på de enkelte studieårene, men også en lavere beståttgrense. Dette er i tråd med det Cohen-Schotanus og Taylor fant. Når vi følger kullene fra år til år, er det en fallende tendens i antall stryk. Dette kan komme av at svakere kandidater må gå året om igjen, faller fra studiet, at de som strøk i starten er bedre forberedt til neste eksamen, samt at man kanskje lærer seg bedre studievaner gjennom studiet. Selv om vi viser at tendensen er fallende antall stryk gjennom studiet, er det noen unntak. Dette kan muligens tilskrives ulik vanskelighetsgrad på eksamener.

Bruk av K-verdi lik 0.7 reduserer varians i strykrate på den enkelte eksamen, samtidig som den reduserer beståttgrensen minst. Gjennomsnittet av beståttgrensene når man ser på alle eksamener, var på 60.3 % (variasjonsbredde mellom 57.5 % og 63.07 %). Dette gir dermed også en lavere gjennomsnittlig strykrate sammenlignet med dagens metode. Samtidig så vi en reduksjon i varians av strykrate og av standardavvik i strykrate, noe som er en indikasjon for at dette kan være en bedre metode (3,4,15). Med en lavere gjennomsnittlig beståttgrense vil man kunne tenke seg at man slipper igjennom studenter som ikke kan nok.

Målet med studien var å undersøke om det finnes en bedre standardsettingsmetode som jevner ut svingningene i dagens absolutte metode. Standard Cohen og modifisert Cohen med K-verdi lik 0.65 og 0.7 ga en lavere beståttgrense enn 65 %, og følgelig en reduksjon i strykratene. Vi må opp i en K-verdi lik

0.75 før man får en beståttgrense som er over 65 % på noen av eksamenene. Med en K-verdi lik 0.75 svinger beståttgrensen rundt 65 %. Ut fra hypotesen skulle man anta at beståttgrensen ble litt høyere på enkle eksamener, og litt lavere på vanskelige eksamener, og at gjennomsnittlig strykrate dermed skulle jevne seg ut. Vi forventet dermed å finne en noenlunde lik gjennomsnittlig strykrate totalt, men en lavere varians, og en lavere variasjonsbredde i strykrate. Mens gjennomsnittlig strykrate gikk ned fra 5.21 % ved absolutt metode, til 4.98 % med modifisert Cohen med K-verdi lik 0.75, gikk varians og variasjonsbredden opp. Varians i strykrate ved absolutt metode var på 17.92, mens varians i strykrate ved modifisert Cohen med K-verdi lik 0.75 ble på 19.55, og variasjonsbredden gikk fra 0-13.7 (absolutt metode) til 0-19.7 (modifisert Cohen med K-verdi lik 0.75). Ser man på strykraten på hver enkelt eksamen, ser man modifisert Cohen med K-verdi lik 0.75 gir en jevnere strykrate for nesten alle eksamener, med unntak for IAB-eksamen 2013. Ved IAB-eksamen 2013 var det 16 (13.68 %) stryk. Cohen med K-verdi lik 0.75 gir en beståttgrense på 65.25 %, og antall stryk blir da 23 (19.66 %). Det er dermed hele syv kandidater som fikk en poengsum mellom 65 % og 65.25 %, og dette er langt over hva man kunne forvente dersom poengsummene var normalfordelt når median for aktuelle eksamen er på 76 %. Likevel ser man en sterk omvendt korrelasjon (Pearssons produkt koeffisient på -0.73) mellom gjennomsnittspoengsum på eksamen, og antall stryk. Man kan derfor spekulere i om sensorer har en tendens til å lete etter ekstra poeng hos studenter som ligger på grensen til stryk. Denne mulige feilkilden vil man ikke få med en standardsettingsmetode som standard Cohen eller modifisert Cohen, siden man på forhånd ikke vet hva beståttgrensen blir.

Ettersom mange medisinske eksamener i utlandet kun bruker flervalgsoppgaver og at man også i Norge går i den retningen, valgte vi å studere bruken av Cohens metoder på kun flervalgsoppgavedelen av eksamen. Dersom man velger å gjøre dette ved medisinutdanningen ved NTNU, ser man at både standard Cohen og modifisert Cohen er aktuelle. Bruken av metodene ga ikke store forskjeller fra hele eksamen sammenliknet med flervalgsdelen. Den metoden som ga størst forskjell i bruken var standard Cohen, da denne korrigerer for gjetning. Med standard Cohen ble variasjonsbredden på beståttgrensen nærmest lik, men gjennomsnittlig beståttgrense falt med 1.04 % (fra 62.2 % til 61.24 %). Gjennomsnittlig strykrate falt noe, men det største utslaget kom på variasjonsbredden av strykraten, som falt fra 13.7 % til 9.57 %.

En svakhet med en slik type standardsetting som baserer seg på et referansepunkt, er at det må være et visst antall besvarelser. Hvor mange besvarelser som behøves ble ikke undersøkt av Cohen-Schotanus, men de antok at metoden en ville gi gode resultater når antall besvarelser er over 100 (personlig korrespondanse med Cohen-Schotanus). Shoenman gjorde et studie på bruk av standard Cohen på grupper

på ca 50, som viste gode resultater (22). På eksamener med et enda lavere antall kandidater, kan ikke disse metodene brukes. Dette gjelder for eksempel kontuinasjonseksamener.

Uansett standardsettingsmetode må man finne en balanse mellom det å la for mange bestå, og å ha en for hard grense. Å vurdere om en standardsettingsmetode er god eller ikke er vanskelig. Når man sammenligner dagens forhåndsbestemte absolutte metode med standard Cohen og modifisert Cohen i vårt materiale, var dagens metode bedre enn antatt. Den gir en lav total gjennomsnittlig strykrate (5.13 %, og et standardavvik på 4.18) og en variasjonsbredde på strykraten fra 0-13.7 %. Dette er betydelig lavere enn hva man oppnådde med standard Cohen i studien til Cohen-Schotanus. (3). Ettersom få studenter stryker i utgangspunktet skal det kun få studenter til for å påvirke den prosentvise strykraten betydelig, og dermed kan man spekulere i om det faktisk er små forskjeller i studentpopulasjonen fra år til år som er årsak til de forskjellene man ser. Dette på tross av like poenggrenser for inntak til studiet. Hvilke tema som blir gitt på eksamen kan kanskje også påvirke strykrate. Med standard Cohen og modifisert Cohen med K-verdi lik 0.7 og 0.65, ble beståttgrensen alltid lavere enn 65 %, og man får dermed en lavere gjennomsnittlig strykrate, en lavere varians av strykraten og en lik/mindre variasjonsbredde av strykraten.

Flere av medisinutdanningene i Norge har en mulighet til å justere på beståttgrensen dersom man ser at unormalt mange studenter stryker. Det er ikke noe system på hvordan dette gjøres, og det blir da en skjønnsmessig vurdering. En mulighet dersom man beholder den tradisjonelle forhåndsbestemte beståttgrensen, vil være å bruke en Cohens metode i disse situasjonene. På denne måten får man en justering som blir satt i system og ikke basert på synsing.

Standard og modifisert Cohen reduserer hvor mye vanskelighetsgraden påvirker eksamensresultatene, samtidig som de reduserer varians. Resultatet av å bytte standardsettingsmetode, vil være at færre stryker på de eksamenene der mange stryker, samtidig som strykraten endres lite på de eksamenene få stryker. Et spørsmål man sitter igjen med, er om det er riktig at færre stryker på eksamen. Diskusjonen blir derfor om hvor den sanne beståttgrensen bør ligge, noe som er vanskelig å bestemme. Man kan tilnærme seg dette ved brukte kriterie-baserte standardsettingsmetoder.

Konklusjon

Andel stryk på samme eksamen med absolutt metode har variert med opptil 13.7 % fra år til år, og er muligens innenfor hva man må akseptere. Samtidig ser man at sensorer kanskje har en tendens til å lete frem poeng for å løfte kandidater opp over beståttgrensen. Alternativt kan man benytte seg av standard eller modifisert Cohen med $K<0.75$ som reduserer disse forskjellene, men på bekostning av en lavere

beståttgrense. Hva som er en akseptabel beståttgrense vil alltid være vanskelig å fastslå, og vil være avhengig av vanskelighetsgraden på eksamen, noe Cohenmetodene prøver å justere. Et alternativ er å benytte seg av en av Cohenmetodene ved ekstreme strykrater.

Hovedbudskap

Forskjellene i strykrate ble redusert av standard Cohen og modifisert Cohen med K<0.75, på bekostning av en lavere grense for bestått. Den absolutte beståttgrensen på 65 % har en akseptabel varians, men det kan være aktuelt å bruke en modifisert Cohens metode ved ekstreme strykrater. Både standard og modifisert Cohen er kosteffektive og enkle metoder om man vil benytte en norm-basert standardsetting som korrigerer for eksamenens vanskelighetsgrad.

Referanser

1. Birkebeinerne - Wikipedia [Internet]. [cited 2015 Nov 2]. Available from: https://no.wikipedia.org/wiki/Birkebeinerne#/media/File:Birkebeinerne_p%C3%A5_Ski_over_Fjeldet_med_Kongsbarnet_(cropped).jpg
2. Vekst i antall leger [Internet]. ssb.no. [cited 2015 Oct 22]. Available from: http://www.ssb.no/helse/artikler-og-publikasjoner/vekst-i-antall-leger
3. Cohen-Schotanus J, Janke C-S, van der Vleuten CPM. A standard setting method with the best performing students as point of reference: Practical and affordable. Med Teach. 2010;32(2):154–60.
4. Taylor CA. Development of a modified Cohen method of standard setting. Med Teach. 2011;33(12):e678–82.
5. Norcini JJ. Setting standards on educational tests. Med Educ. 2003 May;37(5):464–9.
6. Downing SM, Tekian A, Yudkowsky R. Procedures for establishing defensible absolute passing scores on performance examinations in health professions education. Teach Learn Med. 2006 Winter;18(1):50–7.
7. Downing SM, Yudkowsky R. Assessment in Health Professions Education. Taylor & Francis; 2009. 336 p.
8. Bandaranayake RC. Setting and maintaining standards in multiple choice examinations: AMEE Guide No. 37. - PubMed - NCBI [Internet]. [cited 2015 Oct 12]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19117221
9. Thorndike RL, Angoff WH, American Council on Education. Educational measurement. American

Council on Education; 1971. 768 p.

10. Ebel RL. Essentials of educational measurement. Prentice-Hall; 1972. 650 p.

11. Norcini JJ, Shea JA. The Credibility and Comparability of Standards. Applied Measurement in Education. 1997;10(1):39–59.

12. George S, Haque MS, Oyebode F. Standard setting: Comparison of two methods. BMC Med Educ. BioMed Central Ltd; 2006 Sep 14;6(1):46.

13. Karakterstatistikk [Internet]. [cited 2015 Sep 21]. Available from: http://www.ntnu.no/karstat/login.do?lang=no

14. Poenggrenser [Internet]. Available from: http://www.samordnaopptak.no/arkiv/statistikk/11/poenggrenser_vara_hoved.html

15. Poenggrenser [Internet]. Available from: http://www.samordnaopptak.no/arkiv/statistikk/12/poenggrenser_vara_hoved_12.html

16. [No title] [Internet]. [cited 2016 Jan 20]. Available from: https://uit.no/Content/318862/Utfyllende%20bestemmelser%20eksamen%20MED-1501.pdf

17. Birkebeiner.no [Internet]. [cited 2015 Sep 14]. Available from: http://www.birkebeiner.no/no/MainMenu/Arrangement/Ski1/Birkebeinerrennet/Maksimaltid-og-merket/

18. Retningslinjer for eksamen. Regler for obligatorisk undervisning vår/høst 2015 [Internet]. [cited 2015 Oct 12]. Available from: https://innsida.ntnu.no/c/wiki/get_page_attachment?p_l_id=22780&nodeId=24647&title=Retningslinjer+for+eksamen+p%C3%A5+medisinstudiet+-+DMF&fileName=Retningslinjer%20skriftlig%20eksamen%20medisinstudiet%20DMF%202015.pdf

19. Pearson Product-Moment Correlation - When you should run this test, the range of values the coefficient can take and how to measure strength of association [Internet]. [cited 2016 Feb 3]. Available from: https://statistics.laerd.com/statistical-guides/pearson-correlation-coefficient-statistical-guide.php

20. Chevalier SA. A Review of Scoring Algorithms for Ability and Aptitude Tests [Internet]. 1998 Apr [cited 2015 Sep 19]. Available from: http://files.eric.ed.gov/fulltext/ED417220.pdf

21. Paul Ngee Kiong Lau, Sie Hoe Lau, Kian Sam Hong, Hasbee Usop. Guessing, Partial Knowledge, and Misconceptions in Multiple-Choice Tests. Educational Technology & Society. 2011;(14):99–110.

22. Schoeman FHS. Standard setting for specialist physician examinations in South Africa [Internet]. University of the Free State; 2015 [cited 2016 Feb 8]. Available from: http://hdl.handle.net/11660/1663

Tabell 1.

Semester	Eksamen	Kandidater	Mean	Median	Antall stryk(%)	Antall FVO	KR20 FVO
IAB	2010	110	78.7	80	6(5.45)	120	0.82
IAB	2011	117	76.9	79	11(9.40)	120	0.86
IAB	2012	108	77.9	78	4(3.70)	101	0.78
IAB	2013	117	78.6	76	16(13.68)	99	0.83
IAB	2014	115	74.2	77	15(13.04)	100	0.83
IAB	2015	114	76.3	77	5(4.39)	101	0.79
ICD	2010	121	74.4	74	10(8.26)	120	0.91
ICD	2011	113	77.4	78	7(6.19)	120	0.77
ICD	2012	118	73.7	77	14(11.86)	100	0.84
ICD	2013	114	77.4	78	6(5.26)	105	0.84
ICD	2014	109	75.1	77	9(8.26)	100	0.80
ICD	2015	114	76.1	77	11(9.65)	100	0.84
IIAB	2010	103	80.1	81	0(0.00)	100	0.72
IIAB	2011	110	83.3	84	0(0.00)	100	0.76
IIAB	2012	107	77.4	78	3(2.80)	100	0.66
IIAB	2013	99	78.5	79	2(2.02)	100	0.76

IIAB	2014	105	77.4	79	9(8.75)	100	0.84
IIAB	2015	92	73.6	74	11(11.96)	100	0.74
IICD	2010	112	73.3	74	11(9.82)	120	0.82
IICD	2011	119	75.9	77	9(7.56)	120	0.77
IICD	2012	111	79.1	81	10(9.01)	100	0.82
IICD	2013	112	77.4	78	5(4.46)	100	0.73
IICD	2014	105	77.7	80	5(4.76)	100	0.76
IICD	2015	111	77.0	77	3(2.70)	100	0.72
IIIC	2011	118	82.0	77	5(4.24)	85	0.70
IIIC	2012	107	75.1	76	6(5.61)	100	0.66
IIIC	2013	113	83.2	85	1(0.88)	100	0.67
IIIC	2014	108	80.9	81	0(0.00)	55	0.47
IIID	2010	109	82.0	82	0(0.00)	100	0.71
IIID	2011	118	80.2	81	0(0.00)	100	0.7
IIID	2012	118	80.6	81	1(0.85)	100	0.74
IIID	2013	106	78.6	79	2(1.89)	100	0.59
IIID	2014	115	78.9	80	2(1.74)	100	0.73
IIID	2015	111	79.0	79	4(3.60)	100	0.77

Tabell 1: Oversikt over eksamensdata for hver eksamen som er inkludert i oppgaven. FVO står for Flervalgsoppgaver. KR20 er Kuder-Richardson Formula 20, et mål på intern reabilitet for en eksamen.

Tabell 2.a

Standardsettingsmetode	Gjennomsnittlig beståttgrense	Standardd eviasjon	Beståttgrense variasjonsbredde	Gjennomsnittlig Strykrate	Standardd eviasjon	Strykrate variasjonsbredde (%)
65% absolutt	65	0.00	65	5.21	4.23	0 - 13.7
Standard Cohen	62.25	1.54	58.1 - 64.7	3.90	3.69	0 -13.7
Modifisert Cohen K=0.75	64.66	1.45	61.5 - 67.6	4.98	4.42	0 - 19.7
Modifisert Cohen K=0.7	60.35	1.36	57.4 - 63.1	2.96	3.06	0 - 10.4
Modifisert Cohen K=0.65	56.04	1.26	53.3 - 58.6	1.74	2.10	0 - 8.5

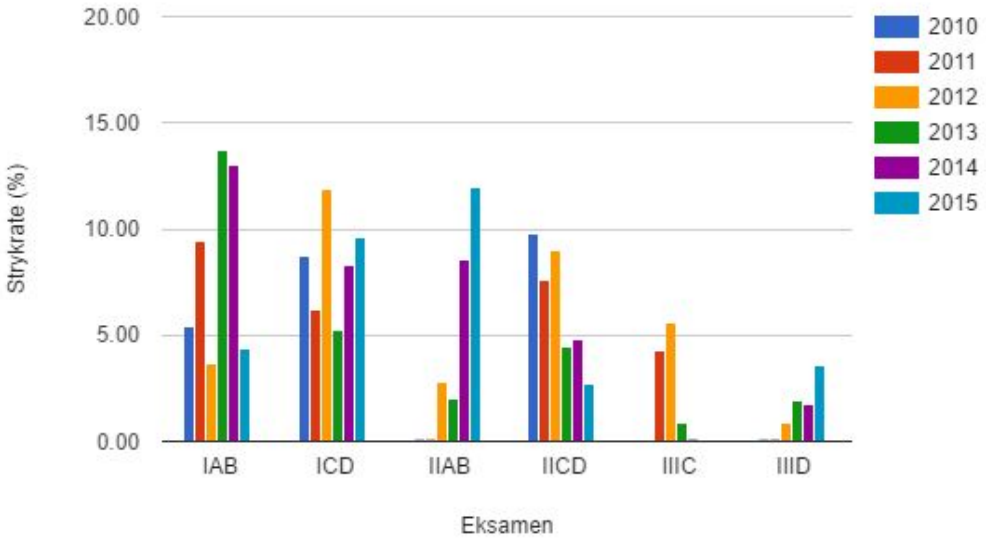
Tabell 2.b

Standardsettingsmetode	Gjennomsnittlig beståttgrense	Standardd eviasjon	Beståttgrense variasjonsbredde	Gjennomsnittlig Strykrate	Standarddeviasjon	Strykrate variasjonsbredde (%)
65% absolutt	65	0.00	65	5.38	4.18	0 - 16.52
Standard Cohen	61.24	1.89	57.2 - 65.3	3.24	2.70	0 - 9.57
Modifisert Cohen K=0.75	64.64	1.93	61.0 - 68.4	5.41	3.61	0 - 15.38
Modifisert Cohen K=0.7	60.33	1.80	56.9 - 63.9	2.96	3.06	0 - 10.43
Modifisert Cohen K=0.65	56.02	1.67	52.9 - 59.3	1.39	1.50	0 - 5.93

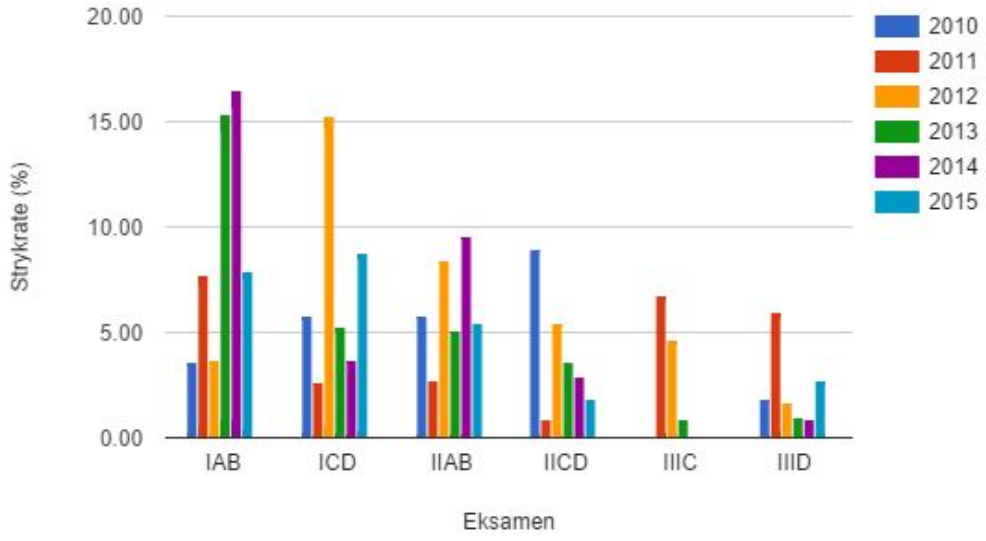
Tabellene sammenlikner de forskjellige standardsettingsmetodene med hensyn til beståttgrense og strykrate. Tabell 2.a tar for seg hele eksamen, mens tabell 2.b tar for seg FVO delen alene.

Figur 1

A


20.00
15.00
10.00
5.00
0.00
Strykrate (%)
IAB
ICD
IIAB
IICD
IIIC
IIID
Eksamen
2010
2011
2012
2013
2014
2015

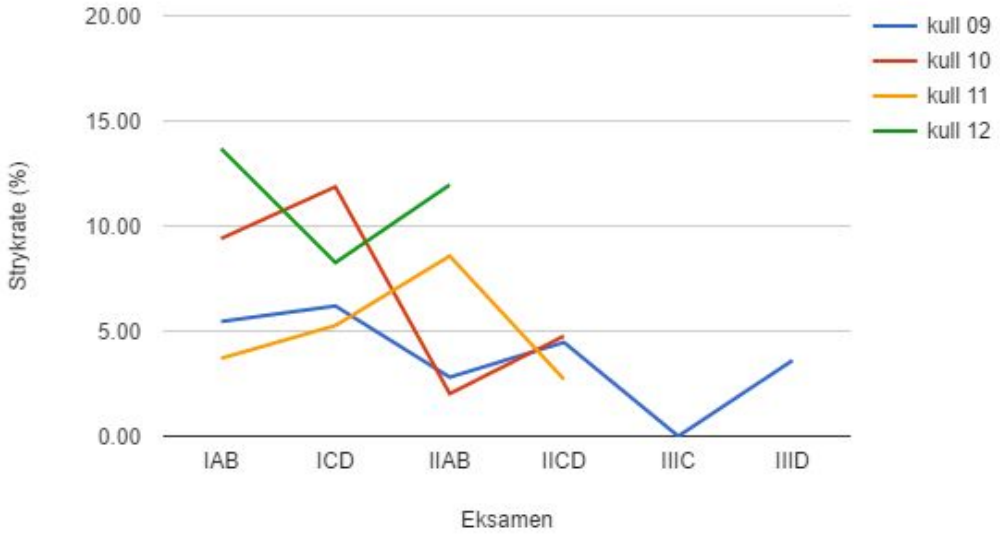
B


20.00
15.00
10.00
5.00
0.00
Strykrate (%)
IAB
ICD
IIAB
IICD
IIIC
IIID
Eksamen
2010
2011
2012
2013
2014
2015

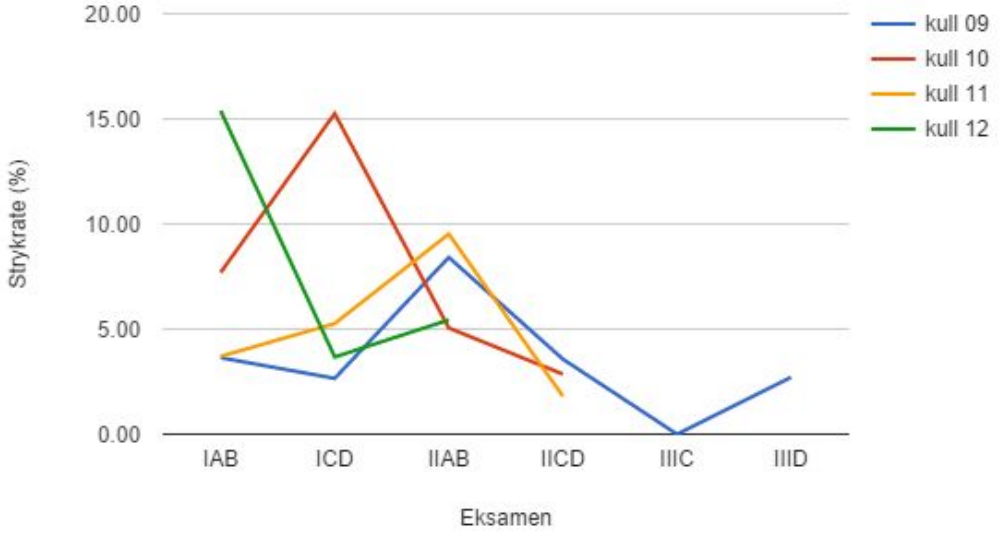
Figur 1: Oversikt over prosentandel av kandidater som har strøket ved hver eksamen på medisinstudiet ved NTNU med dagens ordning med en absolutt grense på 65 %. (A) viser tall for hele eksamen, mens (B) viser resulater hvis man kun ser på flervalgsoppgavedelen av eksamen. X-aksen viser de ulike eksamenene, mens Y-aksen viser antall stryk i prosent.

Figur 2

A


20.00
15.00
10.00
5.00
0.00
Strykrate (%)
IAB
ICD
IIAB
IICD
IIIC
IIID
Eksamen
kull 09
kull 10
kull 11
kull 12

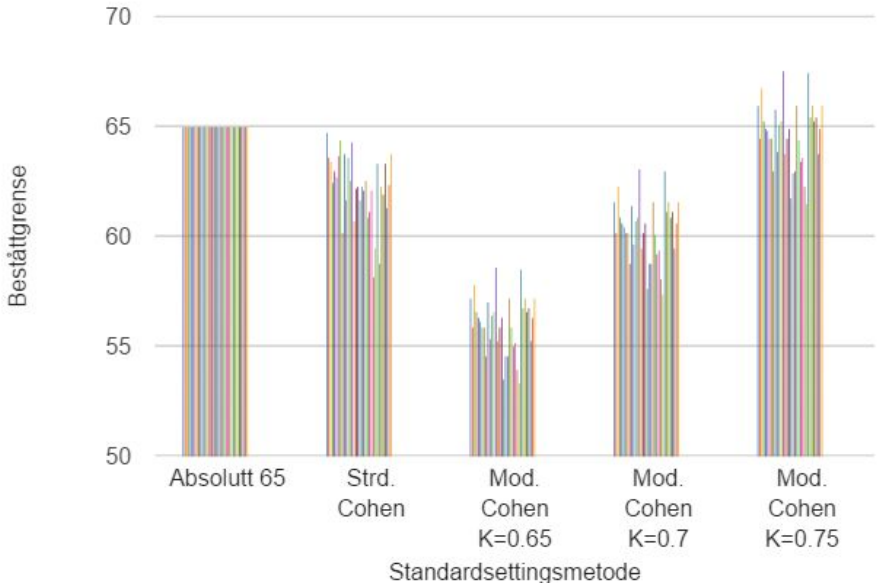
B


20.00
15.00
10.00
5.00
0.00
Strykrate (%)
IAB
ICD
IIAB
IICD
IIIC
IIID
Eksamen
kull 09
kull 10
kull 11
kull 12

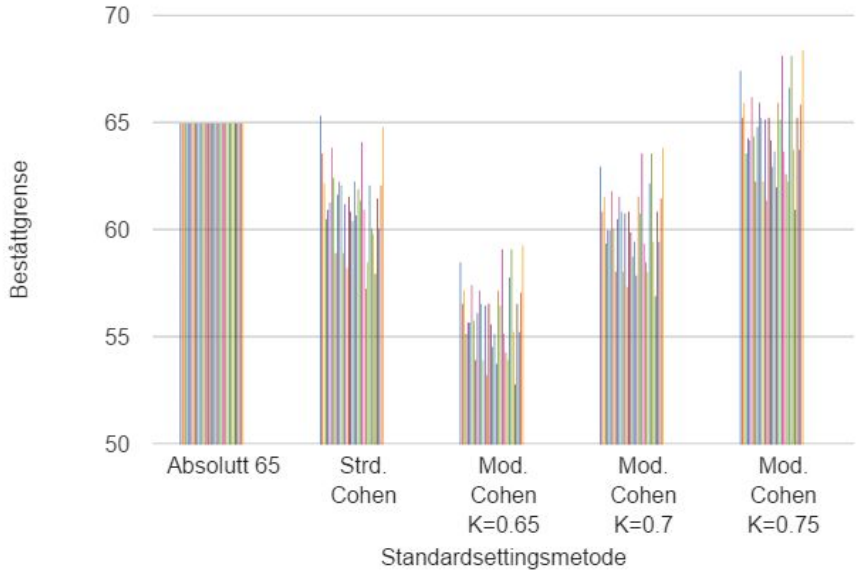
Figur 2: Fire student-kull fulgt over tid, og deres strykrate. Kull 09 startet i 2009, Kull 10 i 2010 osv.(A) viser tall for hele eksamen, mens (B) viser resulater hvis man kun ser på flervalgsoppgavedelen av eksamen. X-aksen viser hvilken eksamen data er hentet fra, mens Y-aksen viser antall stryk i prosent.

Figur 3

A


70
65
60
55
50
Beståttgrense
Absolutt 65
Strd. Cohen
Mod. Cohen K=0.65
Mod. Cohen K=0.7
Mod. Cohen K=0.75
Standardsettingsmetode

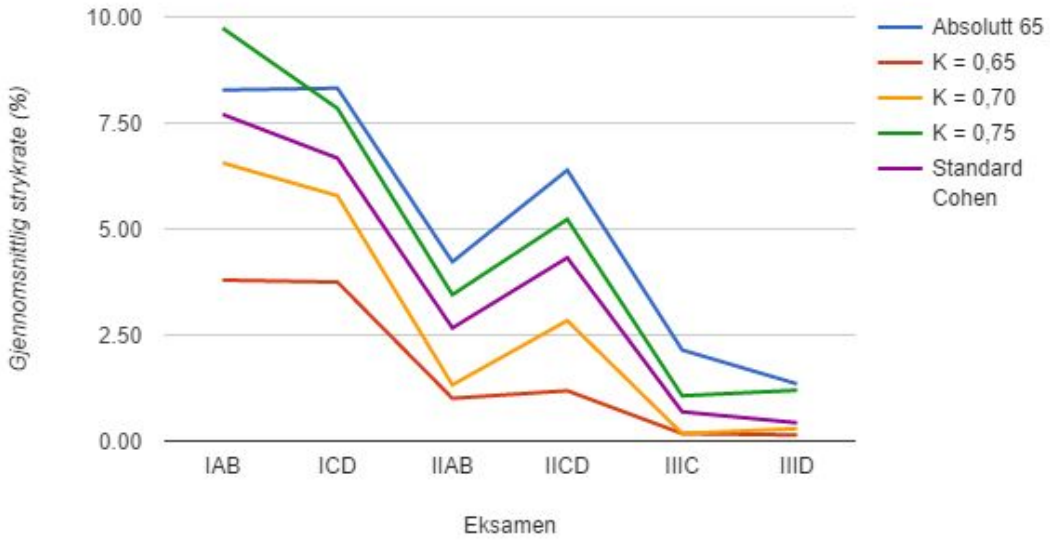
B


70
65
60
55
50
Beståttgrense
Absolutt 65
Strd. Cohen
Mod. Cohen K=0.65
Mod. Cohen K=0.7
Mod. Cohen K=0.75
Standardsettingsmetode

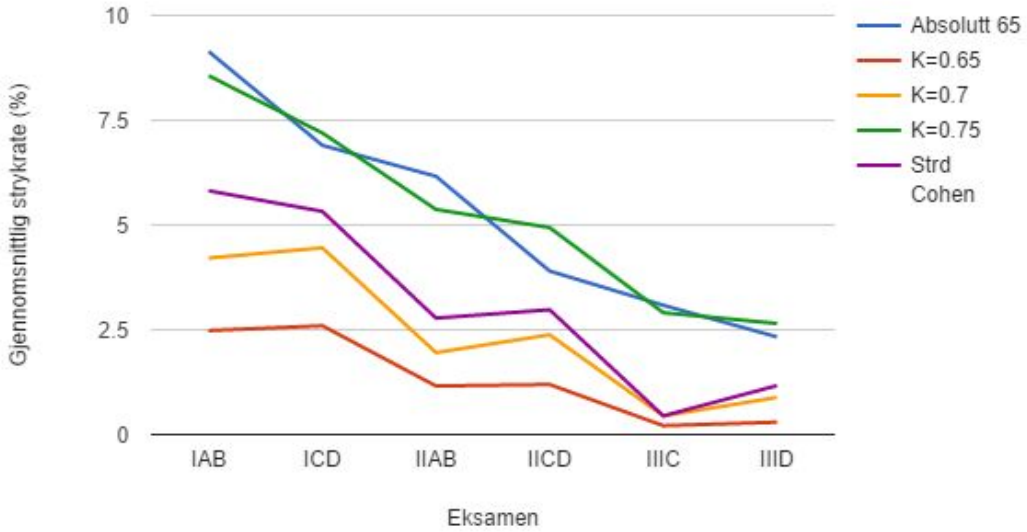
Figur 3: Beståttgrensene bestemt ved de ulike standardsettingsmetodene. (A) viser tall for hele eksamen, mens (B) viser resulater hvis man kun ser på flervalgsoppgavedelen av eksamen. Y-aksen viser beståttgrensen, x-aksen viser de ulike metodene. Hver enkeltsøyle representerer en eksamen.

Figur 4

A


10.00
7.50
5.00
2.50
0.00
Gjennomsnittlig strykrate (%)
IAB
ICD
IIAB
IICD
IIIC
IIID
Eksamen
Absolutt 65
K = 0,65
K = 0,70
K = 0,75
Standard Cohen

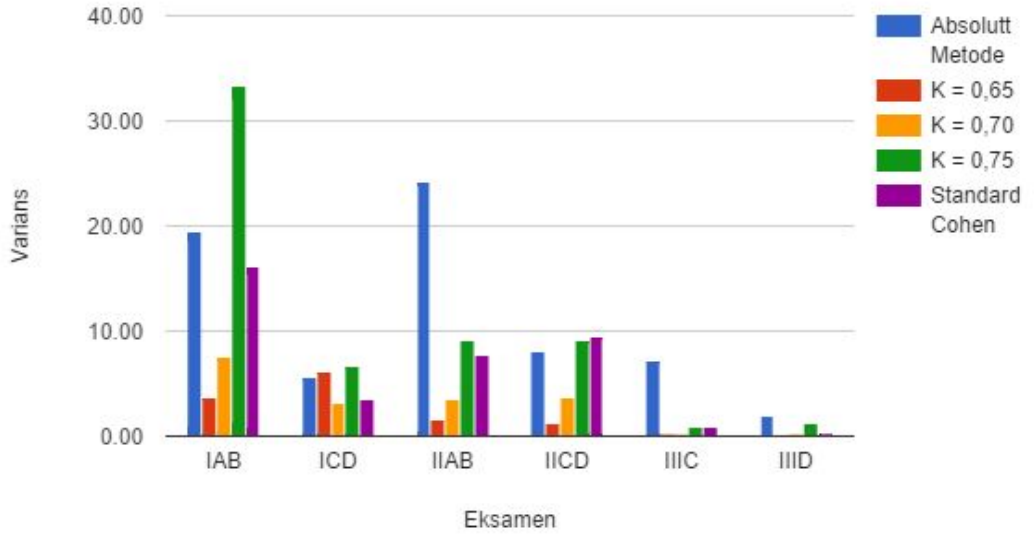
B


10
7.5
5
2.5
0
Gjennomsnittlig strykrate (%)
IAB
ICD
IIAB
IICD
IIIC
IIID
Eksamen
Absolutt 65
K=0.65
K=0.7
K=0.75
Strd Cohen

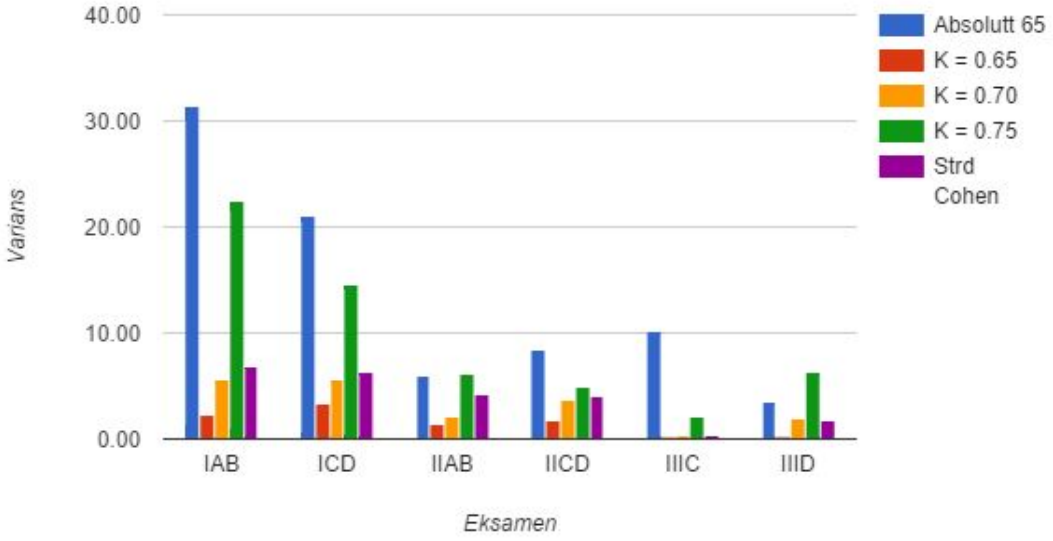
Figur 4: Gjennomsnittlig strykrate for hver eksamen.(A) viser tall for hele eksamen, mens (B) viser resulater hvis man kun ser på flervalgsoppgavedelen av eksamen. X-aksen viser hvilken eksamen, mens Y-aksen viser den gjennomsnittlige strykraten.

Figur 5

A

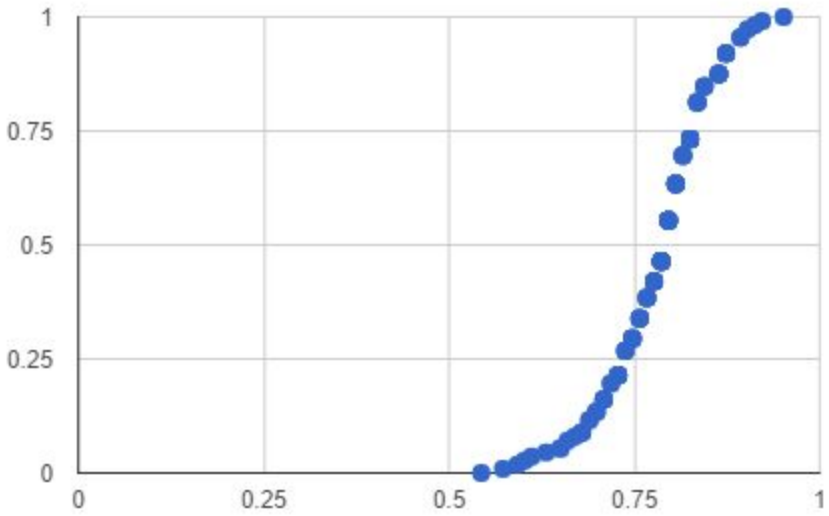

40.00
30.00
20.00
10.00
0.00
Varians
IAB
ICD
IIAB
IICD
IIIC
IIID
Eksamen
Absolutt Metode
K = 0,65
K = 0,70
K = 0,75
Standard Cohen

B


40.00
30.00
20.00
10.00
0.00
Varians
IAB
ICD
IIAB
IICD
IIIC
IIID
Eksamen
Absolutt 65
K = 0.65
K = 0.70
K = 0.75
Strd Cohen

Figur 5: Data fra FVO delen av eksamen. Varians i andel stryk ved bruk av de forskjellige standardsettingsmetodene. Variansen er regnet ut i fra samlet resultat for hver eksamen(2010-2015). (A) viser tall for hele eksamen, mens (B) viser resulater hvis man kun ser på flervalgsoppgavedelen av eksamen. X-aksen viser hvilken eksamen, mens Y-aksen viser varians.

Supplerende figur 1



Supplerende figur 1: Representativ CDF (cumulative distribution function) kurve for ICD eksamen 2013.

Supplerende tabell 1

Studieår	Semester	Basalfag	Kliniske Fag
1. Studieår	IAB	Cellebiologi biokjemi genetikk histologi embryologi medisinsk terminologi Medisinsk historie Medisinsk etikk bevegelsesapparatet Anatomi muskel - skjelett	Lege-Pasient kurs i allmennpraksis.
2.Studieår	ICD	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	IIAB	Patologi	ØNH

		mikrobiologi farmakologi klinisk kjemi epidemiologi atferdsmedisin bildediagnostikk Immunologi	Oftalmologi nevrologi nevrofysiologi fysikalsk medisin onkologi geriatri infeksjonsmedisin Hematologi Kardiologi Lungemedisin Thoraxkirurgi Gastroenterologi Gastrokirurgi
4. Studieår	IICD	Patologi Bildediagnostikk Tropemedisin Sosialmedisin Mikrobiologi Farmakologi	Akutt medisin Dermatologi Ortopedi Revmatologi infeksjonsmedisin Psykiatri Obstretikk Gynekologi Pediatri Endokrinologi Nefrologi urologi Plastisk kirurgi
6. Studieår	IIIC	Allmenmedisin Arbeidsmedisin Geriatri miljømedisin samfunnsmedisin	

		epidemologi medisinsk statistikk klinisk beslutningslære Helsetjenesteadministrasjon Helsetjenesteøkonomi Kvinnehelse Medisinsk etikk Rettsmedisin	
6. Studieår	IIID	Oppsumeringssemester	

Supplerende tabell 1: Oversikt over hvilke emner kanditatene testes i for hver eksamen.