



UPPSALA
UNIVERSITET

Rapport nr 2, 2014

Nationella ämnesprovet i geografi och kravgränssättning

**Översyn och resultat av Angoff-metoden för nationella provet
i geografi årskurs 6**

Andreas Alm Fjellborg och Lena Molin

Förord

Rapportserien *Om nationella prov i Geografi* kommer att redovisa studier och diskutera resultat i syfte att kontinuerligt utveckla och kvalitetssäkra ämnesproven i geografi. Under arbetet med att utveckla ämnesproven i geografi fann provgruppen ett behov av en plats där studier utifrån olika aspekter av ämnesproven kan publiceras för en större läsekrets. Vi vill med denna rapportserie även stimulera till diskussion i och om ämnet geografi, dess syften, innehåll och metoder samt stödja ämnesdidaktisk forskning, kunskapsbildning, utbildning och undervisning.

Gruppen för nationella prov i geografi för åk 6 och 9 arbetar sedan 2012 på Skolverkets uppdrag vid Institutionen för pedagogik, didaktik och utbildningsstudier, Uppsala universitet.

Uppsala i november 2013

Lena Molin
Projektledare för Nationella prov i geografi åk 6 och 9
Institutionen för pedagogik, didaktik och utbildningsstudier
Uppsala Universitet

Nationella ämnesprov i geografi och Kravgränssättning

Översyn och resultat av Angoff-metoden för nationella proven i geografi

årskurs 6

Inledning

År 2008 fastställde regeringen att fler ämnesprov i åk 6 och 9 ska genomföras bland annat i de fyra SO-ämnena (geografi, historia, religion, samhällskunskap). Ämnesproven i slutet av årskurs 6 och 9 ska så allsidigt som möjligt pröva elevens ämneskunskaper i respektive årskurs. Syftet med nationella prov är att stödja en likvärdig och rättvis bedömning och betygssättning samt att ge underlag för analys av i vilken utsträckning kunskapskraven uppfylls på skolnivå, på huvudmannanivå och på nationell nivå. De nationella proven bidrar även till att konkretisera kursplanerna (grundskolan) och ämnesplanerna (gymnasieskolan), och till en ökad måluppfyllelse för eleverna. När det gäller ämnesproven i årskurs 6 och 9 kan proven också utgöra ett stöd för betygssättningen (för de fyra SO-ämnena gäller detta från våren 2014). Från den 1 juli 2011 följer grundskolan en ny läroplan, Lgr 11, med nya kursplaner och ett nytt betygssystem. I samtliga nationella prov prövas förmågorna som finns angivna i kursplanen. I SO är ämnesprovet uppdelat på två delprov som genomförs på två för hela landet bestämda dagar. Nya ämnesprov konstrueras varje år. Alla ämnesprov är obligatoriska.

Enligt den överenskommelse som slutits mellan Skolverket och Institutionen för pedagogik, didaktik och utbildningsstudier, Uppsala universitet, åtar sig provinstitutionen att ansvara för konstruktion och resultatanalys av ämnesproven i geografi för årskurs 6 och 9. Ämnesproven konstrueras av en arbetsgrupp vid provinstitutionen i samarbete med geografilärare från hela landet med erfarenhet från undervisning av elevgrupper inom grundskolan. Utprovning av uppgifter engagerar ytterligare ett stort antal lärare och elever i grundskolan. Som en integrerad del av utvecklingsarbetet med de nationella ämnesproven bedrivs ett omfattande arbete med analys och avrapportering av provresultat.

En del i uppdraget till provinstitutionen handlar om att kvalitetssäkra ämnesproven så att resultatet kan användas för att bedöma kunskapsutveckling över tid vilket är en förutsättning för att syftet med ämnesproven ska kunna uppfyllas. Att provresultatet ska vara stabilt över tid innebär bland annat att svårighetsgraden på proven inte ska variera för mycket mellan provåren dvs. kraven för att nå de olika provbetygen ett provår inte ska sänkas eller höjas inför nästkommande prov. För att uppnå stabilitet i proven över tid används såväl kvantitativa som kvalitativa metoder.

En kvalitativ åtgärd för att nå stabilitet är att förtydliga och kvalitetssäkra de bedömningsanvisningar som hör till proven. Det har visat sig, genom Skolinspektionens uppdrag att granska bedömningen av nationella prov genom omdömning, att frågor kring likvärdigheten i bedömning lärare och skolor emellan bör uppmärksammas och diskuteras vidare. Bland annat visar omdömning att de provuppgifter där eleverna förväntas producera längre svar, t.ex. en skrivuppgift där eleven uppmanas att resonera, i högre grad bedöms olika av olika bedömare jämfört med provuppgifter med slutna provformat. Ett kontinuerligt kvalitetsarbete ute på provinstitutionerna med att utvärdera, förtydliga samt utveckla bedömningsanvisningarna är därför av största vikt.

Ett annat kvalitetsarbete för provinstitutionerna är att säkerställa stor överensstämmelse mellan ämnesprovet och gällande kursplan. Av denna anledning läggs stor vikt vid att provet prövar de förmågor som är tydligt framskrivna i styrdokumentet. Eftersom kursplanerna kräver att undervisningen ska utveckla komplexa förmågor måste bedömningen också ta hänsyn till komplexiteten i förmågan.

Andra åtgärder för att stärka kvalitetsarbetet är att i lärarinformation och bedömningsanvisningar betona fördelarna med att lärare tillsammans bedömer elevsvar för att en större samstämmighet och likvärdighet ska uppnås. Layouten på bedömningsanvisningarna är också föremål för utvecklingsarbete. I geografi arbetar vi med uppslag – ett uppslag innehåller information om uppgiften, vilka bedömningsaspekter som är centrala, en bedömningsmatris där kraven för nivå E, C och A tydliggörs samt exempel på autentiska elevsvar som exemplifierar de olika betygsnivåerna. Till elevsvaren finns kommentarer som lyfter fram vad det är i elevsvaret som ger belägg för att elevens kunskaper håller en viss kvalitet. Överskådliga bedömningsanvisningar ökar tydligheten och blir ett kvalitativt stöd för lärares bedömning. Det är vidare viktigt att i samband med att provuppgifterna utprövas och granskas i olika skeden i provutvecklingen att bedömningsanvisningen, de bedömda elevsvaren och kommentarerna är samstämmiga för att säkerställa kvaliteten.

Skolverket lyfter fram och förordar två möjliga vägar för att uppnå stabilitet i proven över tid. Den ena möjligheten är att utöka antalet av varandra oberoende belägg (poäng) i proven genom att utöka antalet uppgifter. Provens täckning av kursplanerna är av avgörande betydelse för att ge legitimitet till proven och samtidigt får proven inte bli för omfattande. Den andra möjligheten är att stärka själva kravgränssättningsarbetet som innebär att besluta om betygsgränserna för proven (s.k. kravgränser). Provinstitutionernas insatser har främst riktats mot att stärka själva kravgränssättningsarbetet, bland annat har konkreta åtgärder som utökade utprövningsvolymmer för stabilare information till kravgränssättning eller utbildning av kravgränssättare vidtagits.

Den fortsatta texten syftar till att utvärdera och diskutera kravgränssättningsförfarandet i anslutning till de nationella proven i geografi. Texten inleds med en schematisk bild av hela kravgränssättningsarbetet, detta följs av en introduktion till kravgränssättningsmetoder i allmänhet och Angoff-metoden i synnerhet. Merparten av texten centreras kring en beskrivning av arbetet med Angoff-metoden inom gruppen för nationella prov i geografi. Resultat och Angoff-metodens användbarhet presenteras och texten avslutas med reflektioner kring utveckling av arbetssätt, alternativa eller kompletterande metoder samt en kortare utblick mot framtida prov, utprövningar och kravgränssättning.

Vad är en kravgräns?

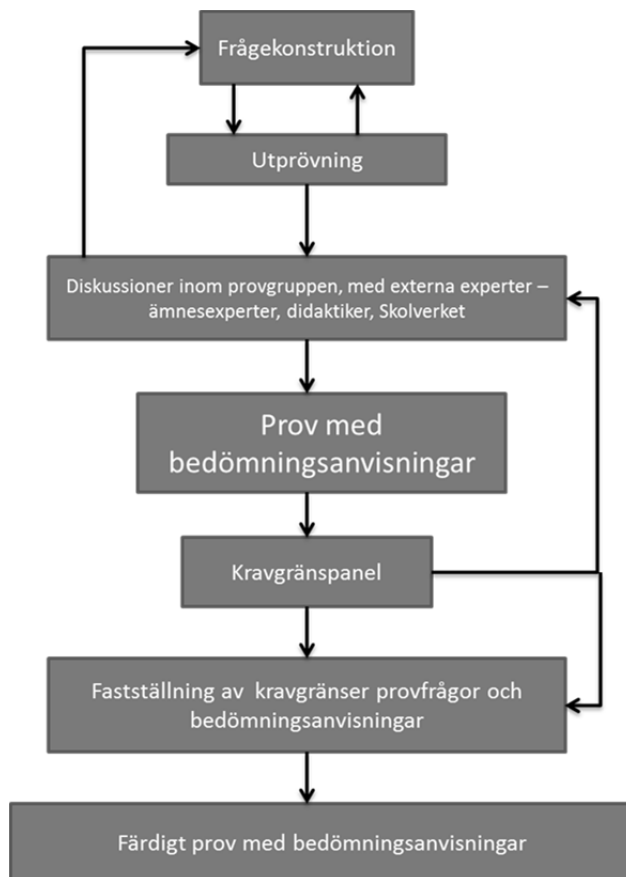
”En kravgräns är en gräns som innebär att kravet för godkänt eller ett visst betyg är att provtagaren får minst den poäng som anger gränsen” (Wikström, 2013: 138). Kravgränser sätts ofta godtyckligt, exempelvis en procentsats rätt svar av provet ger godkänt resultat eller genom att sätta gränserna i efterhand då rättningen redan är genomförd. I båda fallen uppstår det problem. En godtyckligt satt gräns för olika provbetyg, särskilt om gränsen används genomgående, är att 50 eller 75 % rätt svar på ett prov inte har samma betydelse för ett annat prov då frågornas svårighetsgrad eller andra omständigheter skiftar. Proven mäter således inte samma kunskapsnivåer även fast kravgränsen (ex. 50 %) är densamma. Att sätta en relativ kravgräns gör givetvis att det är lättare att få ett högt betyg i en lågpresterande grupp och vice versa oavsett kunskapsnivån hos provtagaren. Kravgränserna har således stor betydelse för utfallen av proven. För låga kravgränser tenderar att godkänna provtagare som inte har tillräckliga kunskaper medan för högt satta gränser felaktigt underkänner elever med tillräckliga kunskaper (Wikman 2013).

Kravgränssättningsstudier används inom gruppen för nationella prov i geografi för att utreda lämpliga kravgränser för nivå E, C- och A. Det finns flera metoder för att genomföra dylika studier och inom nationella proven i geografi 2013 och 2014 användes en *modifierad Angoff-metod med binärt utfall*. Denna text behandlar metoder för kravgränssättning översiktligt samt diskuterar arbetet i gruppen för nationella prov på djupet. Syftet med texten är att dels diskutera metoden som för närvarande används och belysa förtjänster och utvecklingspotential, dels att diskutera de faktiska utfallen av metoden för proven 2013 i relation till de slutgiltiga resultaten av proven. De huvudsakliga frågorna som kommer att diskuteras är:

- Är det en rimlig metod att använda och vad finns det för utvecklingsmöjligheter?
- Utgör det insamlade materialet en god grund för kravgränssättningen inom nationella proven i geografi?
- Vad behöver utvecklas för att metoden skall bli mer tillförlitlig?

Arbetet med kravgränser för nationella prov i geografi – en översikt

Figur 1



Ämnesprovet i geografi konstrueras genom en process i flera steg. Arbetet startas med konstruktion av frågor som sker inom provgruppen, i dialog med lärare och andra ämnesexperter. Frågorna testas i utprövningar där flertalet klasser besvarar frågorna. Eleverna och deras lärare får också besvara en enklare enkät. Bedömningar av elevsvaren, enkäterna och ytterligare kommentarer från lärare ligger till grund för utveckling av potentiella provfrågor. När ett mer eller mindre komplett prov med tillhörande bedömningsanvisningar finns att tillgå tar kravgränssättningen vid. En panel av lärare (experter) inbjuds att delta och bedöma provets svårighetsgrad enligt Angoff-metoden, samt att diskutera, implementera och utvärdera tillhörande bedömningsanvisningar. Erfarenheter från dessa möten ligger till grund för ytterligare bearbetningar och justeringar inför fastställandet av det slutgiltiga provet.

Kravgränssättningsförfarandet är en del av arbetet med de nationella proven som kommer sent i processen men också den del som ligger till grund för E-, C- och A-nivåerna i provet. Det finns en kontinuerlig kvalitetsgranskning genom hela processen, i denna text koncentrerar vi oss på kvalitet och arbetssätt i själva kravgränssättningen. Innan arbetssättet med kravgränser diskuteras mer ingående görs här en kortare genomgång av några relevanta kravgränssättningsmetoder som bedöms kunna vara rimliga alternativ till den modifierade Angoff-metoden som för närvarande används.

Kravgränssättningsmetoder

Lärare som konstruerar prov för att utreda sina elevers kunskapsnivåer har sällan möjlighet att använda sig av mer standardiserade kravgränssättningsmetoder. I större prov, till exempel nationella prov, finns dessa resurser och kravgränssättningen kan inte vara godtycklig utan måste utredas och fastställas på empirisk väg. Därmed krävs någon form av studie för att fastställa vilka kravgränser som skall gälla för olika nivåer. Resurser som krävs för detta är professionella bedömare som relativt väl känner populationen som skall genomföra provet (lärare eller andra

experter), tid att genomföra någon form av kravgränssättningsmetod samt tid att analysera resultaten på ett adekvat sätt.

Kravgränssättningsstudier faller inom två kategorier, *item*-baserade (frågebaserade) eller *individ*-baserade (elev-baserade), Angoff-metoden är en *item*-baserad metod. Andra metoder inom samma kategori är *Ebel*, *Nedelsky* och *Bookmark*. Personcentrerade exempel är *Borderline Survey* och *Contrasting Groups* (Wikman 2013, se vidare Cizek och Bunch 2007). Uppdelningen kommer av angreppssättet, de frågecentrerade studierna är baserade på undersökningar av frågor utifrån en given population, medan personcentrerade studier är baserade kring utredning av frågornas svårighetsgrad genom att undersöka individer i relation till en uppsättning frågor. I och med att arbetet med nationella proven utgår från att undersöka hur olika typer av frågor faller ut på en given population (de elever som skall genomföra provet) så berörs här endast frågecentrerade angreppssätt för kravgränssättning.

- Angoff-metoden: Metoden bygger på att en grupp av experter (lärare) utvärderar en fråga genom att uppskatta hur många i en grupp som nätt och jämt når en kravgräns det vill säga klarar frågan. Detta ger en statistisk skattning normalt givet som ett p-värde eller i procent. Exempelvis kan en lärare, eller annan expert, göra skattningen att 25 % (eller 0,25) av de elever som ligger kring en tänkt betygsgräns klarar en fråga. Lärarens skattning vägs sedan samman med andra paneldeltagares skattning och ett medelvärde uppnås som också blir en föreslagen E-, C-, eller A-nivågräns i de nationella proven - det vill säga panelens rekommendation för kravgräns. Resultatet är således en uppskattning av hur väl en del av populationen som är minimalt (men ändå tillräckligt) kompetent skulle prestera. Metoden är givetvis behäftad med problem i och med skattningsfel och överestimat. Lärare kan ha en tendens att överestimera, eller underestimera den faktiska kompetensen hos de elever som skattningen gäller. En kalibrering mot tillgänglig data är mer eller mindre nödvändig. Dessutom krävs det att lärarna är inbegripna med vad som skall skattas, det bör betonas att skattningen gäller de som skulle klara uppgiften, inte *borde, skulle kunna med rätt övning* eller dylikt.
- Modifierad Angoff-metod: Skillnaden mellan den modifierade Angoff-metoden och dess föregångare är att panelen av experter, eller lärare, ges information om processen, utbildas i metoden och diskuterar sina ursprungliga skattningar och eventuellt ändra dessa. Som i Angoff-metoden ombeds lärarna att skatta hur stor andel de elever som nätt och jämt når en nivå skulle klara uppgiften. Till exempel kan det handla om hur många elever som ligger kring gränsen för nivå E som kommer att klara frågan. Estimatens anges i procent eller p-värden av hur stor andel av *gränseleverna* som skulle klara frågan. Sedermera ges möjligheten för diskussion eller modifiering efter att experterna fått ta del av andra panelers skattningar, tillgänglig data eller annan information som kan påverka deras estimat. Oftast används denna metod vid flervalfrågor.
- Binär modifierad Angoff-metod: Som metoden ovan. Dock efterfrågas inte estimat i formen 0,45 alt. 45 % av eleverna klarar en fråga, utan experterna frågas om en provtagare som precis når E-, C-, eller A-nivå skulle klara frågan eller inte. Lärarnas skattning blir således: 0 – frågan klaras ej alternativt 1 – frågan klaras av en elev som

precis når behandlad nivå. En stor fördel med denna metod är att det är lättare för en lärare att tänka på en specifik elev snarare än en population av elever på gränsen till en nivå (Ricker 2006). Även denna metod är behäftad med några problem. En skattning med binärt utfall tenderar att ge överestimat vid relativt enkla prov och underestimat vid svårare prov, detta diskuteras mer utförligt i avslutningen.

- Nedelsky-metoden: Experter skattar vilka svarsalternativ som en precis godkänd elev skulle klara av att eliminera som ett felaktigt svar. Används endast i flervalfrågor.
- Bookmark-metoden: I denna metod rangordnar kravgränspanelen frågorna efter svårighetsgrad. Därefter får varje paneldeltagare "bokmärka" hur många frågor som elever som precis uppnår nivå E, nivå C eller nivå A skulle klara av. Också här tas ett medeltal av de olika paneldeltagarnas skattningar som ett förslag till kravgräns. Denna typ av metod fungerar på olika typer av frågor. Dock kan det givetvis vara svårt att fastställa ordningen det vill säga svårighetsgraden av frågorna, särskilt i ett brett prov som de nationella proven i geografi kan det finnas problem i att rangordna frågorna sinsemellan särskilt när de är av skiftande karaktär (flervalfrågor, essäfrågor, resonerafrågor etc.)

Ovanstående metoder skulle vara lämpliga i större eller mindre utsträckning för nationella proven i geografi, dock lämpar sig vissa metoder för specifika typer av frågor vilket gör att flertalet metoder inte skulle kunna användas på hela provet. Det finns fördelar och nackdelar med alla metoder. Vad som sammanfattningsvis kan fastställas är att den binära modifierade Angoff-metoden som används i kravgränssättningsarbetet inom nationella prov för geografi, ger information kring hur elever med minimala kunskaper för respektive betygssteg skulle klara sig. Problemen blir att metoden inte ger någon information kring hur många som inte skulle klara provet och inte heller en uppskattning kring den vidare populationen. Sammantaget kanske man kan förvänta sig att över en hel panels elevunderlag (det vill säga de deltagande lärarnas klasser) är resultatet nära en normalfördelning av betygen. Detta är dock inte ett underlag som vi får. För att kunna säkerställa likvärdighet över tid kan det också vara lämpligt att låta lärare skatta andelen elever som ligger på gränsen till nivå E, C och A.

Arbetet med kravgränssättning inom gruppen för nationella prov i geografi

Kravgränssättningsarbetet genomförs vanligtvis i flertalet olika steg. Först bjuds en grupp lärare in att delta vid utvärderingen av frågorna i relation till deras elevpopulationer. Vanligtvis deltar cirka 15-20 lärare vid dessa träffar. Lärarna utgår från bedömningsanvisningar som tillhandahålls och skattar sedan hur en av deras elever som precis når upp till nivå E, C eller A skulle prestera på provet.

Lärarna förväntas: (i) vara väl förtrogna med kursplanen och har tillgång till bedömningsanvisningar för varje enskild fråga. (ii) arbeta med geografiämnet och ha aktuell

erfarenhet av undervisning. Dessutom tillser provgruppen för de nationella proven att lärarna kommer från olika skolor med rimlig spridning över landet.

Kravgränssättningsgruppen inför provet 2014 bestod av lärare som närvarat vid liknande kravgränssättningsträffar tidigare (50 %) och den andra halvan av deltagarna genomförde denna typ av kravgränsmöten för första gången. Antalet lärare som deltog var 18 (åk6) och 16 (åk9). Dessa möten genomfördes över två dagar för varje årskurs.

Panelen av lärare ombeds att uppskatta vilka frågor elever som precis når de olika betygsstegen kommer att klara av. För att sedan kunna få fram kravgränser summeras och analyseras de olika lärarnas resultat och diskuteras inom panelen och vidare inom provgruppen. Då 2013 års prov var det första som genomfördes så finns det inte äldre data att tillgå utan jämförelsen gjordes med utprovningmaterial och diskuterades inom provgruppen innan slutgiltiga beslut om kravgränser togs.

Till kommande år finns det stora möjligheter att utveckla arbetssättet kring kravgränssättningsförfarandet. Det mest tillgängliga är att utarbeta verktyg för att jämföra kravgränssättningsträffar med tidigare års träffar och utfall på de faktiska ämnesproven. Dels för att kontrollera för rättningsbias och urvalsfel. Nedan genomförs en utvärdering av kravgränssättningen inför ämnesprovet 2013. Analyser av dess resultat jämförs med resultat från ämnesprovet 2013 med utblickar mot det kommande ämnesprovet för 2014.

Årskurs 6

Ämnesprovet för årskurs 6 2013 ser något annorlunda ut jämfört med det utkast till prov som användes vid kravgränssättningsträffen med inbjudna lärare. Två frågor tillkom i det skarpa provet och en fråga togs bort från versionen som lärarna bedömde i anslutning till kravgränssättningen. Dessa frågor har exkluderats från de vidare analyserna.

Lärarna som bjöds in till kravgränssättningsträffen blev ombedda att bedöma vilka frågor deras elever skulle klara. Metoden som användes var en modifierad Angoff-metod med binärt utfall. Detta betyder att lärarna skattar huruvida en elev klarar frågan eller ej. De elever som de skulle göra skattningen på är den elev i deras respektive klass som precis når upp till E-, C-, respektive A-nivå. Utfallet blir sålunda en skattning av hur deras minst kompetenta elev som ändå når respektive nivå klarar ämnesprovet. Denna metod har visat sig relevant för denna typ av undersökningar (Rickert 2006). Ett relevant alternativ skulle kunna vara att arbeta med skattningar som inte har en binär utfallsvariabel, det vill säga att lärarna uppskattar hur stor andel av eleverna som kommer klara respektive fråga (vi får då en skattning mellan 0 och 1 där 1 är lika med 100 % klarar frågan). I en sådan studie summeras sedermera skattningarna och vi kan utifrån dessa bestämma kravgränserna. Detta förfarande har visat sig mindre tillförlitligt då lärarna måste göra estimat på alla deras respektive elever i stället för en per nivå, särskilt om lärarna inte utbildats i metoden eller inte fått tillräcklig information. Fördelar och nackdelar diskuteras ytterligare i diskussionen nedan.

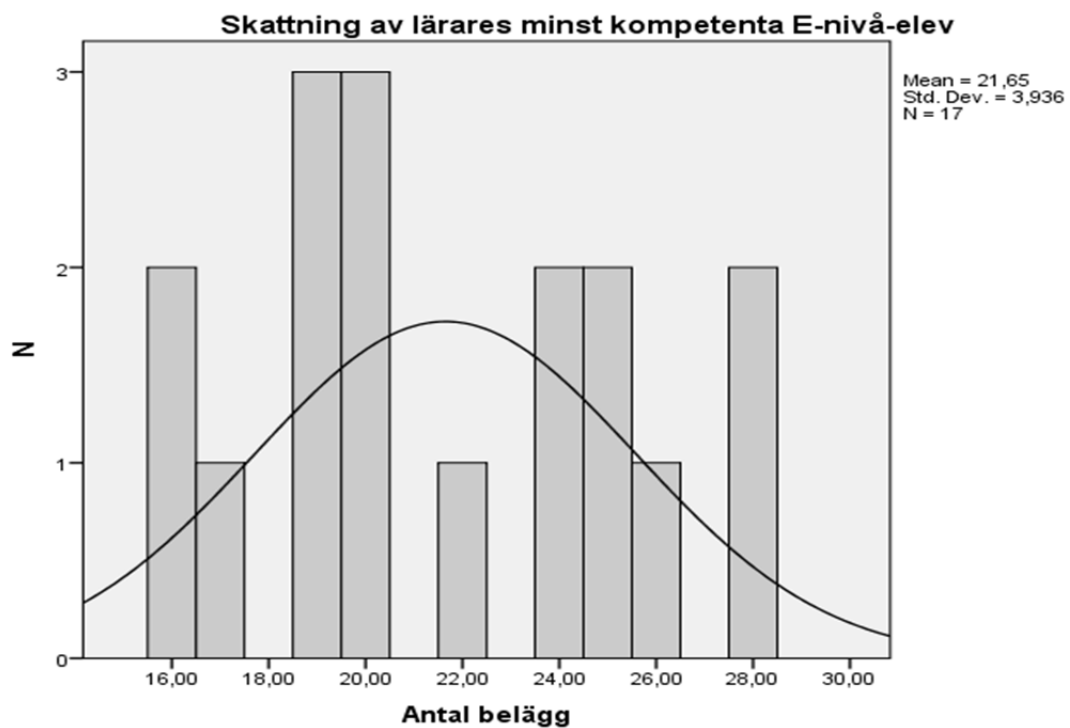
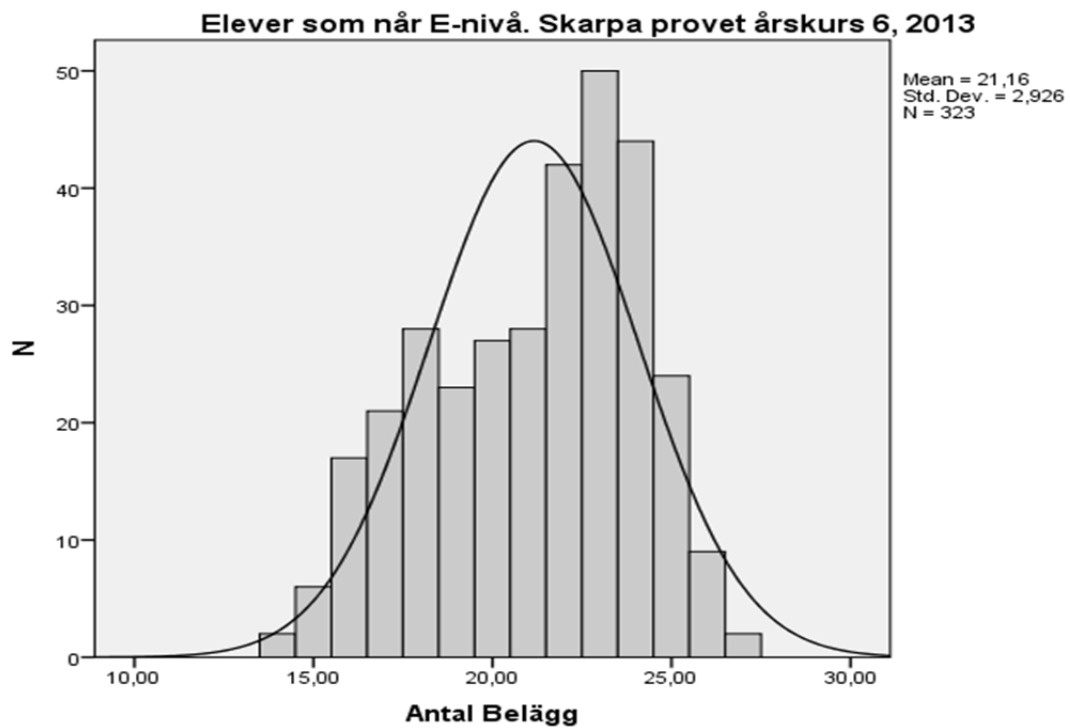
Tabell 1 visar utfallet från kravgränssättningsmötet och utfallet från ämnesprovet 2013. Det finns stora likheter i materialet särskilt för A- respektive E-nivån, lite mindre för C-nivå. Variansen i materialet från kravgränssättningsmötet är stor, där vissa lärare har skattat sina elever mycket lågt eller högt i jämförelse med medelvärde. Detta syns också i standardavvikelsen. Dock är detta mindre problematiskt med tanke på att målet med övningen är att undersöka rimliga kravgränser i relation till lärarnas elevunderlag. Det är således av vikt att ha god förståelse för urvalets betydelse i detta förfarande, något som kan utvecklas och diskuteras nedan. En mer visuell bild av materialet presenteras i en serie histogram (Figur 1-6, nedan).

Tabell 1.

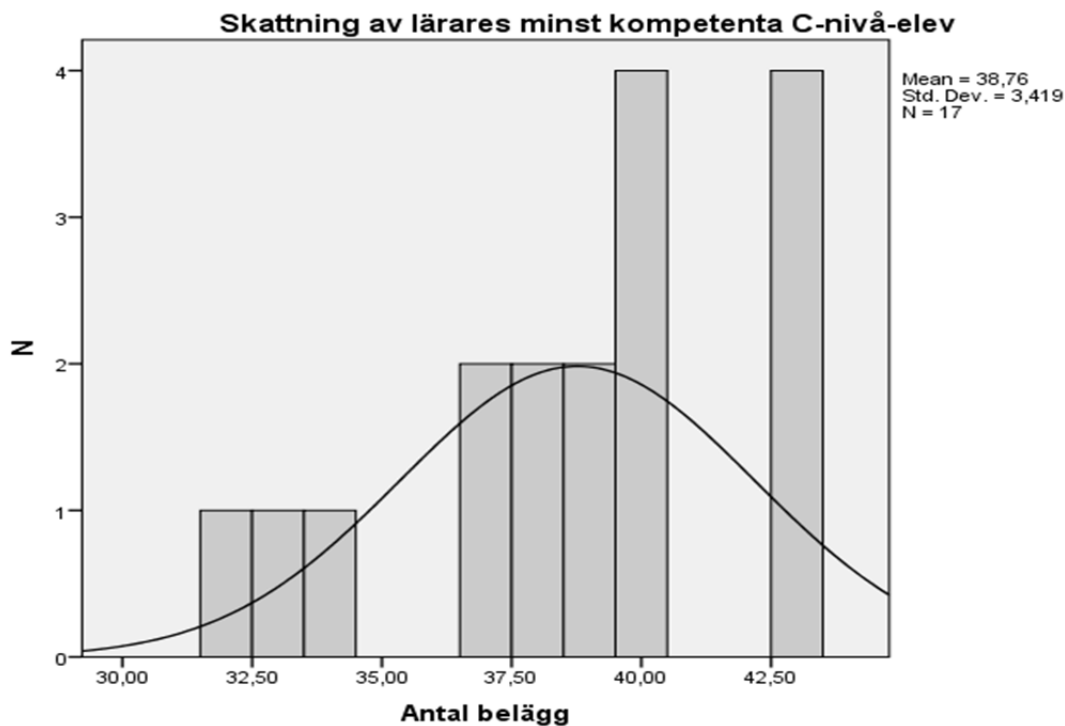
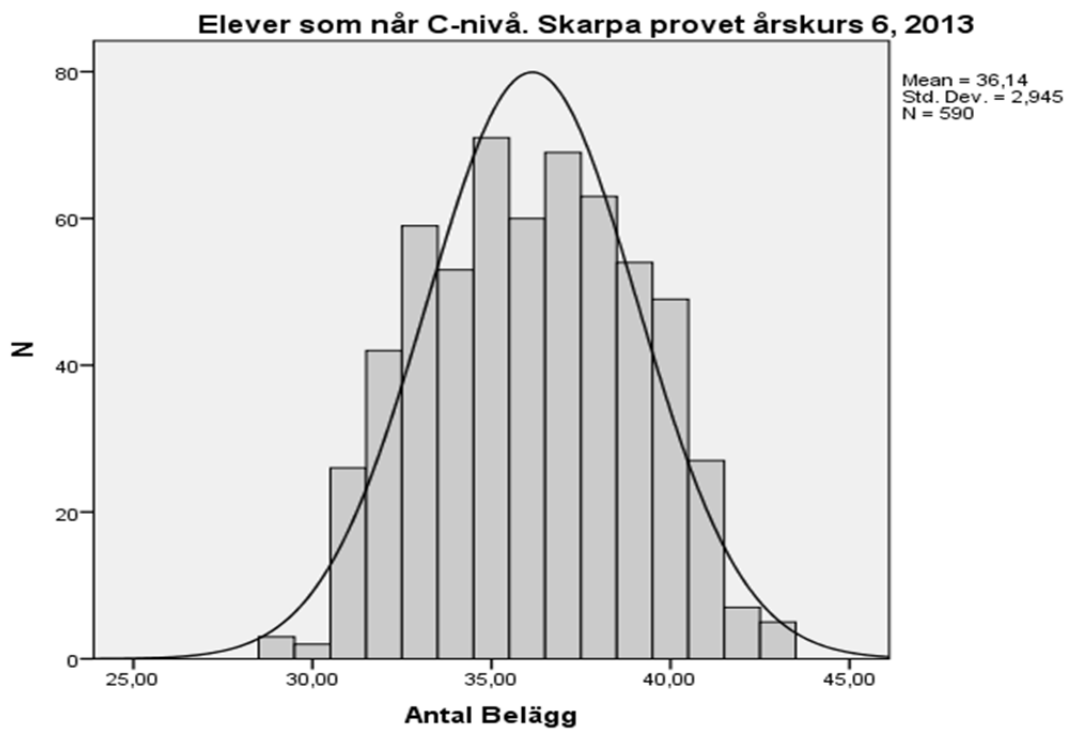
	N	Medel	Median	Std.av	Varians	Min.	Max.
A - Skarpa	159	49,7484	49,0000	2,62164	6,873	45,00	56,00
A-Kravmöte	17	48,5882	50,0000	5,45503	29,757	37,00	57,00
C-Skarpa	590	36,1356	36,0000	2,94464	8,671	29,00	43,00
C-Kravmöte	17	38,7647	39,0000	3,41924	11,691	32,00	43,00
E-Skarpa	323	21,1579	22,0000	2,92608	8,562	14,00	27,00
E-Kravmöte	17	21,6471	20,0000	3,93607	15,493	16,00	28,00

Källa: Författarens beräkningar, data från Nationellt prov geografi åk 6, 2013, jämte data från kravgränssättningsmöte i Uppsala, Mars 2013.

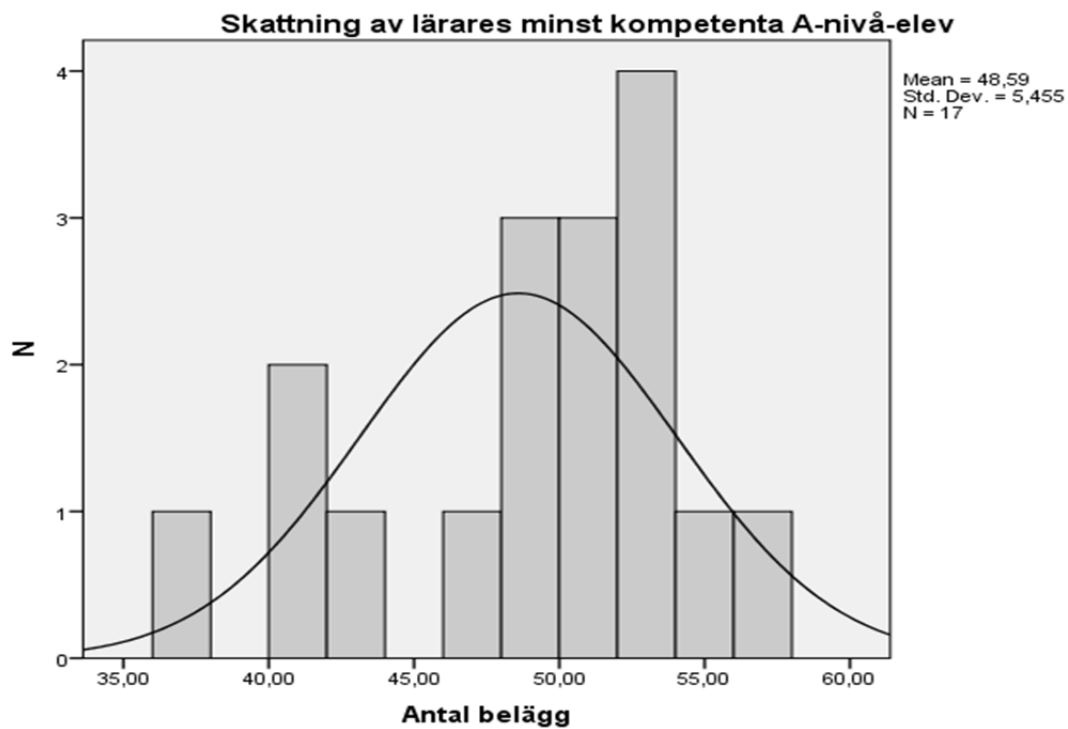
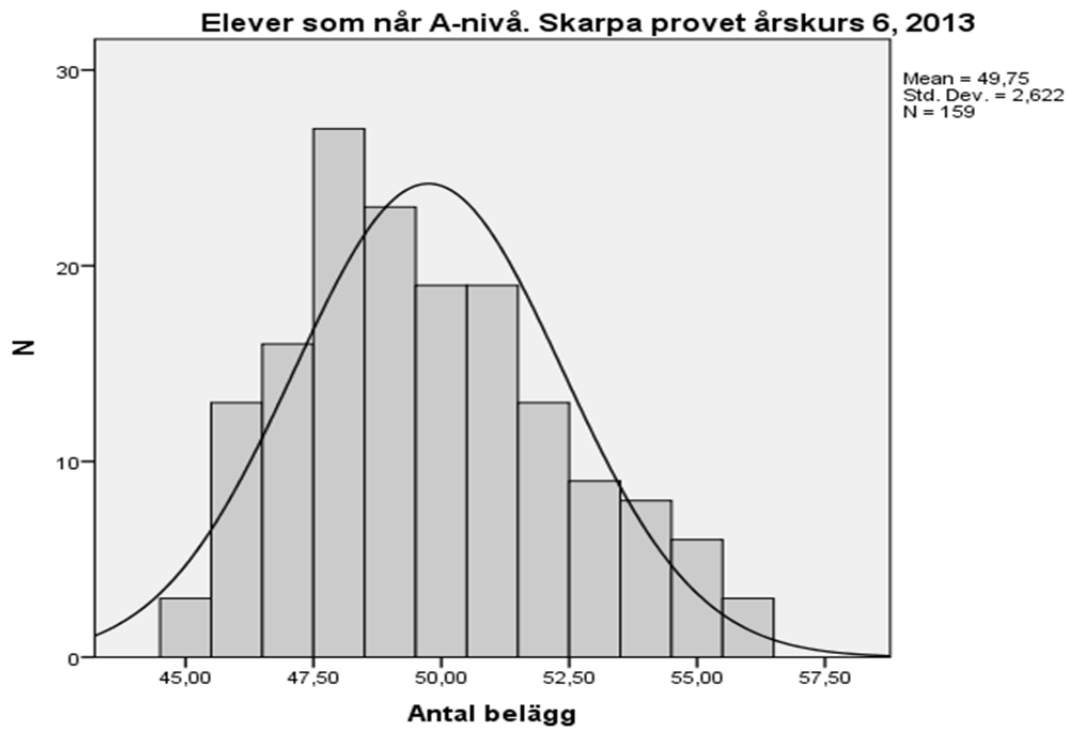
Histogram över elever som nått A-nivå, skarpa provet årskurs 6, 2013, och vid kravgränssättningsmötet, Figur 1-2.



Histogram över elever som nått C-nivå, skarpa provet årskurs 6, 2013, och vid kravgränssättningsmötet, Figur 3-4.



Histogram över elever som nått E-nivå, skarpa provet årskurs 6, 2013, och vid kravgränssättningsmötet, Figur 5-6.



Diskussion av resultat

Resultaten från kravgränssättningspanelen för årskurs 6 visar på relativt högt antal uppnådda belägg. Lärarna har skattat hur en av deras elever som precis når gränserna för nivå E, C respektive A klarar de olika frågorna. Resultaten kan tolkas som att deras minst kompetenta elever i varje betygsgrupp klarar sig relativt bra i en jämförelse med den totala populationen. Eller så består skillnaden i skattningsfel, det vill säga att panelen överestimerar kompetensen hos sina elever. Det finns också fog att fundera kring urvalets betydelse. Om det är relativt kompetenta klasser som lärarna företräder är det inte särskilt överraskande att de som lärarna bedömer som gränselever i sina klasser klarar sig relativt väl i en jämförelse med den totala populationen. Av förklarliga skäl är standardavvikelsen hög för kravgränssättningspanelen vilket endast är ett tecken på spridningen av vad som anses vara en minimalt kompetent elev som ändå når en viss betygsgräns. Det som är intressant att notera är att medelvärdet bland de skattade minst kompetenta i det närmaste sammanfaller med medelvärdet hos de elever som uppnår de olika kravgränserna. Slutresultaten på ämnesprovet kan också resultera i mellanstegen D respektive B men diskuteras inte här i relation till kravgränserna. Den statistik som diskuteras i denna text är av en deskriptiv art, det finns inga anspråk på att göra någon statistisk analys som kan hjälpa oss att förstå, korrigera eller förändra kravgränssättningsarbetet inom provgruppen. Dock visar materialet på att lärarna skattar sina respektive elever relativt högt jämfört med utfallet för den totala populationen. Rimligt vore att kravgränssättningspanelens rekommendation sammanfaller väl med minimivärdet för varje betygsgräns. Dock är detta inte fallet. Panelens skattningar tenderar att ligga en god bit över detta resultat, särskilt för C- respektive E-nivåerna.

De generellt höga skattningarna indikerar att lärarna överskattar elevernas förmåga och därmed kan vi dra slutsatsen att frågorna är relativt lätta, då den binära Angoff-metoden ofta ger överestimat vid enklare frågor (se nedan). Men i relation till detta diskuteras Rackase (2004) ett hypotetiskt exempel och ju bättre experterna är på att skatta desto större blir diskrepansen mellan föreslagna kravgränser och det faktiska utfallet. Därmed är det rimligt att anta att utfallet för kravgränspanelen för årskurs 6 var väl förtrogna med undersökningspopulation (deras egna elever) och därmed satte relativt precisa skattningar vilket resulterade i höga kravgränser.

Diskussion kring metod och utvärdering

Angoff-metoden är vida använd. I forskning kring dess användbarhet finns tecken på att Angoff-metoden och de modifieringar som finns av den är relevanta verktyg för att bestämma kravgränser. Ricker (2006) menar att metoden fungerar särskilt väl när paneldeltagarna kan få ordentlig träning i metoden och där diskussioner kring skattningar kan föras. Dock har den använda Angoff-metoden i denna text visat sig mer lättpåverkad i den mån att det sker större förändringar i skattningarna om lärarna får ta del av diskussioner och empirisk data för att sedan omvärdera sina estimat. Fördelarna ligger i att metoden är lätt att använda och lärarna kan ge rimliga skattningar av flertalet provfrågor vilket är svårare då lärare har en hel population att ta hänsyn till. Både den ursprungliga Angoff-metoden och diverse modifieringar av densamma har visat sig ha blandade resultat (se bland annat: Goodwin 1999, Plake och Impara, Fhermann,

Woehr och Arthur 1991, Chinn och Hertz 2002. För en översikt Ricker 2006). van Nijlen och Janssen (2008) finner i en studie att Angoff-metoden ger höga kravgränser. De menar vidare att detta kan påverkas av, framförallt, två aspekter (i) presentationen av uppgiften samt (ii) de dikotoma utfallet. Det är av vikt att presentationen av *minimalt kompetent elev som når nivån* som undersöks är precis. Det är lätt att lärare skattar vad en elev *borde kunna*, *skulle kunna* prestera i stället för vad eleven faktiskt presterar. Att beskriva vad en lärares elev borde kunna uppnå är skilt från vad de kan bedömas faktiskt uppnå. Är beskrivningen av uppgiften felaktig kan det leda till estimat av en idealbild snarare än det faktiska utfallet. Den binära (klarar/klarar ej) skattningen kan också generera problem. Då en elev faktiskt kanske lyckas på 80 % av frågorna i ett prov med 10 frågor så kan estimaten generera för höga kravgränser då en lärare tvingas välja klarar/klarar inte på varje enskild fråga. Detta kommer då att ge 10 av 10 snarare än 8 av 10 som är en mer precis uppskattning av elevens faktiska kompetens. I ett prov med lätta frågor och estimat av lärare som väl skattar elevpopulationens kompetens kommer en dikotom Angoff-metod att ge för höga kravgränser. I ett prov med svåra frågor får det omvända resultatet (van Nijlen och Janssen 2008, se vidare Reckase 2004). Wikström (2013) berör också detta och menar att det är vanligt att bedömare överskattar provtagarnas kompetens på svåra frågor och underskattar prestationer på lätta frågor när det kommer till den modifierade Angoff-metoden utan binära svarsalternativ, utan när experter/lärare får genomföra skattningar för en större population.

I ljuset av denna metodlitteratur kan vi dra några slutsatser kring materialet presenterat i denna text. Dock är det svårt att på djupet diskutera eventuella över- respektive underestimat. Men C-nivån förefaller vara svårast att skatta. Det är relativt nära mellan medelvärden i både E och A nivåerna mellan kravgränssättningspanelen och det faktiska utfallet på provet. Men vi ska ha i åtanke att det är lärarnas skattningar av deras elever som precis når nivå E, C och A och inte ett medel av eleverna på nivåerna E, C och A. Detta ger vid hand att estimaten inte är särskilt goda. Alla estimat hamnar inom rimliga gränser av respektive nivå men generellt ligger skattningarna i överkant jämfört med vad vi kan förvänta oss från de elever som precis når nivå E, C eller A. Skattningarna från experterna borde sammanfalla med miniminivån i resultaten från ämnesprovet. Det finns dock en diskrepans, minst på A-nivån och störst på C-nivån. Detta kan ha flera förklaringar. Bedömningsanvisningarna spelar roll, det kan också vara lättare för lärare att bedöma mer kompetenta elever, eller så är instruktionerna bristfälliga.

Bedömarna i kravgränssättningspanelen vt 2013 kom fram till gränser som rimligt korresponderade med det slutgiltiga medelvärdet för respektive subpopulation som genomförde ämnesprovet. Dock skall vi ha i åtanke att dessa skattningar också är en del i underlaget för slutgiltiga kravgränser så överensstämmelsen är given från början. Vad som är intressant att fortsatt undersöka är den problematiska C-nivån. Men också de andra nivåerna är av intresse. I lätta frågor underskattas elevernas kompetens, vilket torde medföra att ett lätt prov får för höga kravgränser från panelen. I och med att kravgränspanelens rekommendationer, där de minst kompetenta eleverna för varje nivå är riktmärket torde också medelvärdet för skattningen vara i närheten av minimipöängen för en elev som når ett visst betygssteg. I fallet med ämnesprovet för årskurs 6, 2013 finns det istället en överensstämmelse mellan kravgränssättningspanelens skattningar och medelvärdet inom varje nivågräns på ämnesprovet. Det torde vara ett tecken på att ämnesprovet innehöll relativt svåra frågor, om vi följer Wikströms (2013) argumentation i och med att föreslagna kravgränser är högt satta i relation till elevunderlaget på varje nivå. Men van

Nijlen och Janssen (2008), i ett referat av Rackase (2004), menar att lätta frågor leder till höga estimat av kravgränser om bedömarna är väl insatta i skattningsförfarandet.

En ytterligare komplicerande aspekt ligger i om kravgränssättningen skall vara relativ eller absolut. En relativ kravgränssättning ligger till grund för en rangordning av, i detta fall, elever. Att separera de 10 %, 20 % osv., mest kompetenta eller minst kompetenta från resterande del av populationen. En absolut kravgräns är istället till för att sätta en kravgräns som skiljer på de som uppnått tillräckliga kunskaper från de som inte har det. Det är givet att den senare varianten lämpar sig bäst när kravgränsen skall skilja mellan de elever som når en nivå från de som inte gör det och fördelningen däremellan är av underordnad betydelse (Norcini, 2003). Dock kan detta komplicera bilden i och med att många lärare sätter betyg utifrån ett relativt synsätt i deras respektive klass, vilket framgår av de stora avvikelserna mellan experternas E-, C-, eller A-bedömningar. Detta kan komplicera tolkningen av kravgränssättningspanelens rekommendationer till provgruppen.

Inför kommande kravgränssättningsmöten är det relevant för provgruppen i geografi att fortsätta utveckla och kvalitetssäkra arbetet med kravgränssättning. Det finns exempelvis möjligheter att låta samma lärare, eller en större grupp lärare som är väl förtrogna med arbetet med de nationella proven i geografi delta i kravgränssättningsarbetet. Lärare som tidigare har bedömt prov, varit delaktiga i undervisningen och har kunskap om provkonstruktörernas arbetssätt kan hjälpa till att få en bra skattning av kravgränserna. Om lärarna sedemera anser att provet generellt är svårt eller enkelt kan också ge en indikation om dessa skattningar bör korrigeras. Detta görs till viss del redan idag då kravgränssättningspanelens rekommendation vägs samman med utlåtanden om provens svårighetsgrad från en rad experter inom området. Vidare vore det forskningsmässigt och utredningsmässigt relevant att genomföra någon typ av jämförelse mellan olika skattningsförfaranden för kravgränssättningspanelerna inom ramen för nationella proven i geografi. En sådan undersökning skulle kunna ge god inblick i hur kravgränssättningsmetoder kan användas och vilka problem som kan vara behäftade med olika metoder.

Avslutning

Näsström och Nyström (2008) menar att lärare med högpresterande elever tenderar att sätta högre kravgränser. En instruktion om att utgå från en elev som precis når nivå E, C eller A kan uppfattas olika av olika lärare. Det är därför av vikt att betänka urvalet mer noggrant inför kommande utprovningar. Detta kan göras genom att använda tidigare betyg, meritvärden eller dylikt för att fånga låg- respektive högpresterande skolor, klasser och elever. För att komma åt den relativa uppfattningen om vad som konstituerar kunskapsnivåer hos elever som precis når ett visst betygssteg bör större vikt läggas vid bedömningskriterier och information och kunskap om dessa. Lärarurvalet är av vikt och utbildning och diskussioner lärarna emellan är viktiga inslag i att säkerställa en likvärdig uppfattning av bedömningen och kraven på eleverna.

Därutöver finns det möjlighet att undersöka olika kravgränssättningsmetoder för att få en bättre uppfattning om lägsta gränsen för nivå E. I dagsläget får provgruppen ingen indikation från kravgränssättningsmötena, utöver de mer generella diskussionerna som förs. Lärare har också rest

frågor om vad som bör betraktas som ett elevsvar som ej når E-nivån i de lärarenkäter som lärarna fyller i vid genomförandet av det nationella ämnesprovet i geografi. En utökning av kravgränssättningen i relation till flervalsfrågor bör också beaktas. Att införa Nedsky-metoden för att få bättre kunskap om hur denna typ av frågor fungerar för elever som precis når respektive betygssteg är också relevant för det fortsatta arbetet.

Litteratur

Cizek, G., Bunch, M. (2007): *Standard Setting – A Guide to Establishing and Evaluating performance Standards on Tests*. Thousand Oaks: SAGE Publications.

Chinn, R.N., Hertz N. R. (2002): Alternative approaches to standard-setting for licensing and certification examinations. *Applied Measurement in Education*, 15, pp 1-14.

Fehrmann, M.L., Woehr, D.J., Arthur, W. (1991): The Angoff cutoff score method: The impact of frame-of-reference rater training. *Educational and Psychological Measurement*, 51, pp 857-872.

Grelsson, Gunnel, Bergqvist Göran, Jacobsson, Niklas (2011): Nya kursprov i biologi och kemi för gymnasieskolan. Bakgrund och resultat 2011. *BVM*, Nr 58, 2013.

van Nijlen, Daniël och Janssen, Rianne (2008): Modeling Judgment in the Angoff and Contrasting-Groups Method of Standard Setting. *Journal of Education Measurement* 2008, Vol 45, No1, pp. 45-63.

Norcini, J. J. (2003): Setting standards on educational tests. *Medical Education*, 37: 464–469. doi: 10.1046/j.1365-2923.2003.01495.x

Näsström, G., & Nyström, P. (2008): A comparison of two different methods for setting performance standards for a test with constructed-response items. *Practical Assessment, Research & Evaluation*, 13(9), 1-11.

Plake, B.S., Impara, J.C. (2001): Ability of panelists to estimate item performance for a target group of candidates: An issue in judgemental standard-setting. *Educational Assessment*, 7, pp 87-97.

Rackase, M.D. (2004): What if there were a “True Standard Theory” for standard setting like the “True Score Theory” for tests? *Measurement*, 2, 114-119.

Ricker, Kathryn L. (2006): Setting Cut-Scores: A Critical review of the Angoff and Modified Angoff Methods. *The Alberta journal of Education Research*, Vol 52, No 1, pp 53-64.

Wikström, Christina (2013): *Konsten att göra bra prov – vad lärare behöver veta om kunskapsmätning*. Natur och Kultur. Stockholm.

Rapportserien *Om nationella prov i geografi*

Grubbström, Ann & Alm Fjellborg, Andreas & Molin, Lena (2013) *Lärarnas åsikter om Nationella provet i geografi. Sammanställning av lärarenkät i samband med utvärderingsomgången 2013.*

Alm Fjellborg, Andreas & Molin, Lena (2014) *Nationella ämnesprov i geografi och Kravgränssättning översyn och resultat av Angoff-metoden för nationella proven i geografi årskurs 6.*



UPPSALA
UNIVERSITET