
Ämnesproven i grundskolans årskurs 9 och specialskolans årskurs 10

Provrapport Samhällskunskap

Årskurs 9

Vårterminen 2014



GÖTEBORGS UNIVERSITET
INST FÖR PEDAGOGIK OCH SPECIALPEDAGOGIK

Skolverket

Innehåll

Inledning.....	Sid 3
Beskrivning av ämnesprovet och utprövningar.....	sid 3
Bakgrund.....	Sid 3
Utseende.....	sid 4
Bedömningsanvisningar.....	Sid 4
Resultat.....	Sid 4
Resultat fördelat på faktorerna kön och undervisning i SvA.....	sid 5
Validitet och reliabilitet	Sid 8
Validitet	Sid 8
Reliabilitet	sid 9
Lärarenkäten	sid 12
Slutsatser	sid 16
Referenser	sid 17
Bilaga 1: Prövningar inför årets prov	

Inledning

Skolverket har i uppdrag av regeringen att ansvara för och genomföra nationella prov i SO-ämnena (dvs. geografi, historia, religionskunskap och samhällskunskap) i årskurs 9. Skolverket har i sin tur gett i uppdrag åt Göteborgs universitet att konstruera och pröva ut de nationella proven i samhällskunskap för åk 6 och åk 9. Denna rapport behandlar ämnesprovet i åk 9. Provet genomfördes 15 och 16 maj 2014.

Lärarna har digitalt inrapporterat provbetyg och uppgiftsresultat för ett urval av elever. Lärare har också besvarat en digital enkät. Ett urval av hela elevprov har också skickats in.

Denna rapport bygger på de inrapporterade provresultaten, de hela elevproven, lärarenkäterna och resultaten från SCB. Skillnaden mellan SCBs statistik och vårt urval redovisas också. Det är också viktigt att hålla i minnet att det är knappt hälften av lärarna för de aktuella åldrarna som har besvarat enkäten och att det kan påverka utfallet.

I materialet nedan nämns flera statistiska metoder och begrepp. Man behöver inte ha kunskaper om dessa för att kunna följa resonemanget. De är ändå angivna så den intresserade kan läsa vidare om våra metoder och statistiska överväganden.

Beskrivning av ämnesprovet och utprövningar

Bakgrund

Ämnet samhällskunskap förekommer bara i de nordiska länderna. I övriga världen finns ofta ett ämne som påminner om statvetenskapsdelen inom samhällskunskap, men inget med vårt ämnes alla olika delar: statsvetenskap, nationalekonomi, sociologi, kulturgeografi och idag ofta internationella studier, kriminologi osv. Det finns inga andra nationella prov i ämnet. Utvecklingspotentialen är stor.

Vid provkonstruktion talar man ofta om god mätsäkerhet. Det kan låta tekniskt men är samma sak som att eleven skall ges en så rimlig chans som möjligt att visa sina kunskaper och att provet speglar elevernas faktiska kunskaper så bra som möjligt. God mätsäkerhet kräver att elever prövas med fler frågor inom varje delkunskapskrav¹. Om inte så beror provresultatet alltför mycket på slump. Detta innebär också att verksamheten att utprova uppgifter blir stor.

Förutom mindre utprövningar av enstaka uppgifter har fyra större utprövningar genomförts. Under framtagandet av ämnesprovet har tusentals elever och hundratals lärare medverkar vid dessa prövningar (3589 elever och 117 lärare från 100 skolor över hela landet för åk 9-provet). Se bilaga 1.

Ämnesprovet måste ge möjlighet för eleverna att visa kunskaper trots att det innehåller man arbetat med, t.ex. begreppet demokrati, kan vara applicerat på olika områden. Det får inte bli för ”områdesberoende”. Efter provet 2013 så genomförde vi en större undersökning av elevsvaren till de uppgifter där eleven fått F på enskild uppgift. Svaren kan då vara tomma eller av sådan kvalitet att de försökt besvara uppgiften, men inte nått godkänt-nivå för uppgiften. I den senare kategorin har de ändå försökt, och inte blivit avskräckta av uppgifterna. Vi fann då att eleverna i mycket hög grad försökt besvara uppgifterna (klart mer än 90 %). Vi har genomfört mindre undersökningar på året inkomna elevlösningar och sett samma trend. Eleverna vågar sig alltså på uppgifterna, vilket är den första tröskeln som man måste passera.

Under arbetet med att sälla fram bra uppgifter ur de prövade har vi arbetat mycket medvetet med olika varianter av IRT, Item Response Theory, s.k. modern mätteori. Vi har med hjälp av olika program analyserat varje prövning i detalj, hur de olika uppgifterna fungerat i olika elevgrupper osv.

En stor del av vårt arbete handlar om reliabilitet och validitet. Vi redovisar detta under ett senare kapitel.

¹ Kunskapskraven i samhällskunskap åk 9 kan lätt ses som sex olika innehållsmässiga delar. Vi kallar detta delkunskapskrav.

Liksom förra året har också arbetat mycket med så kallade stödstrukturer i provet. Vi har noterat att stödstrukturer kan ”stödja” åt fel håll varför vi försökt hitta en lagom nivå. Inom ämnesprovet i samhällskunskap för år 6, så prövade vi redan förra året att flera av de öppna uppgifterna följts av uppmaningar från Sam, en uggle som kan ge eleverna viss hjälp med att strukturera uppgifterna. I år så prövade vi i år 9 en ”tankebubbla” med liknande syfte. Enligt lärarenkäten (se nedan) har detta fallit väl ut.

Utseende

Siffrorna inom parentes nedan avser provet 2013.

Ämnesprovet i samhällskunskap i år 9 bestod av två delprov. Det fanns ingen principiell skillnad mellan delproven. Samhällskunskap år 9 har relativt många uppgifter, 30 stycken. Eleverna har haft relativt många möjligheter att visa sina kunskaper. Detta för att få måtsäkerhet (fler och ”olika” uppgifter) och för att ge eleverna möjlighet att svara på och att pröva sina kunskaper på olika frågor både vad det gäller format och innehåll.²

Np sh år 9	Antal uppg./item	Upp till nivå			Antal strukturerade
		E	C	A	
2014	30	30	30	19	12
2013	29	30	29	16	19

Tabell 1. Fördelning av uppgifter, nivåer och format 2014/2013

Bedömningsanvisningar

Inga provbetyg ges direkt av elevens svar på uppgifterna. De måste först bedömas av en lärare utifrån våra bedömningsanvisningar. Arbetet med att skapa bra bedömningsanvisningar är ofta minst lika svårt som att skapa bra uppgifter. Vi har arbetat mycket med läsbarheten av dessa och har använt samma upplägg som vid 2013 års prov, ett tvåspaltigt upplägg där det – om möjligt – gick att få en uppgifts bedömning på ett uppslag. Enligt enkäterna och följande intervjuer med bedömande lärare så har detta fallit väl ut.

Resultat

Resultaten från provbetygen återrapporterades av SCB den 25 november för år 9 och den 6 dec för år 6. Sedan förra året har de nationella proven inom So en digital inrapportering där lärare har rapporterat in ett urval av elever (födda den 6:e, 16:e och 26:e i varje månad). 1835 elever har inrapporterats digitalt, vilket gjorde att vi tidigt hade preliminära resultat att arbeta med (18 juni). Detta möjliggör ett halvt års bearbetning av resultat på många olika sätt som vi annars inte hade haft möjlighet att göra. Denna obligatoriska inrapportering av ett statistiskt urval är helt avgörande för analys och framtagning av proven, då den ger oss information på uppgiftsnivå, vilket SCB’s rapportering inte gör. I texten nedan används beteckningen ”urval” för dessa elever.

² I internationell litteratur så kallas en bedömning för ett ”item”. Ordet är numera helt försvenskat, alltså item. Ett item kan bestå av flera uppgifter, t.ex att uppgift 2,3 och 4 bedöms tillsammans i ett ”paket” och blir då ett item. Eller tvärtom kan t.ex. uppgift 2a och 2b bedömas var för sig och blir då två item. I NpSh år 9 2014 motsvarade alltid en uppgift ett item.

Nedanstående diagram visar i ordning provbetyget för vårt urval i juni, SCBs provbetyg i november och SCBs siffror för de slutgiltiga betygen (kom i september). Man ser att vårt statistiska urval står sig mycket bra med avseende på det slutgiltiga resultatet för provbetyget, därav syns också vikten av att dessa insamlingar kan göras.

Som jämförelse är också de slutgiltiga betygen medtaget. Dessa grundar sig på en mycket bredare bedömning av eleverna än vad de nationella proven gör.

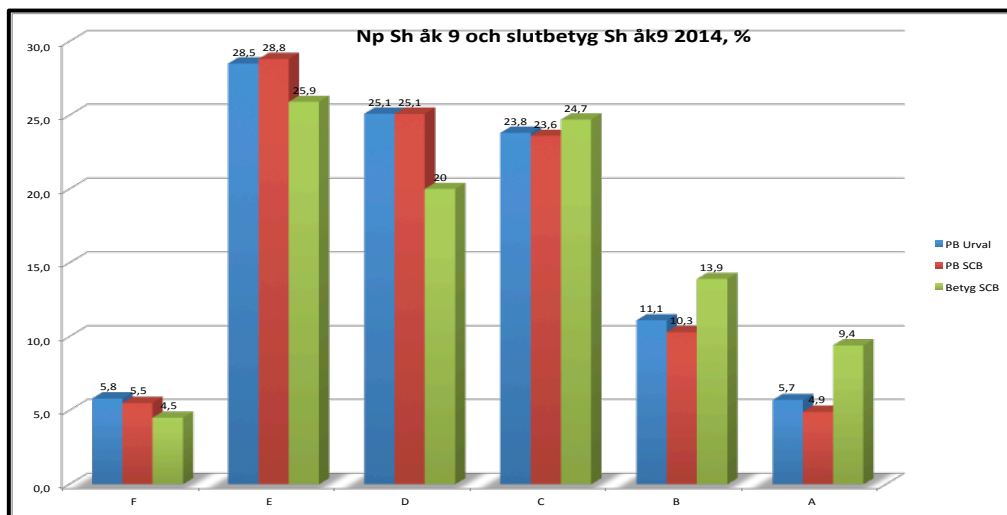


Diagram 1. Fördelning av provbetyg och slutbetyg samhällskunskap i ämnesproven i årskurs 9: /Urval provbetyg/ SCB provbetyg/ slutbetyg 2014.

Betygen uppvisar en liten s.k. ”hängmatta” (D är lägre än omgivande E och C), det finns många betyg E.

I enkäterna tycker lärarna i snitt att provbetygen är ”rimligt” satta. Lärarna uttrycker dock ofta att provbetygsnivån för F/E är för lågt satt. Denna synpunkt återspeglar sig inte direkt i lärarnas egen betygssättning.

Resultat fördelat på faktorerna kön och undervisning i svenska A

Skillnader mellan pojkar och flickor

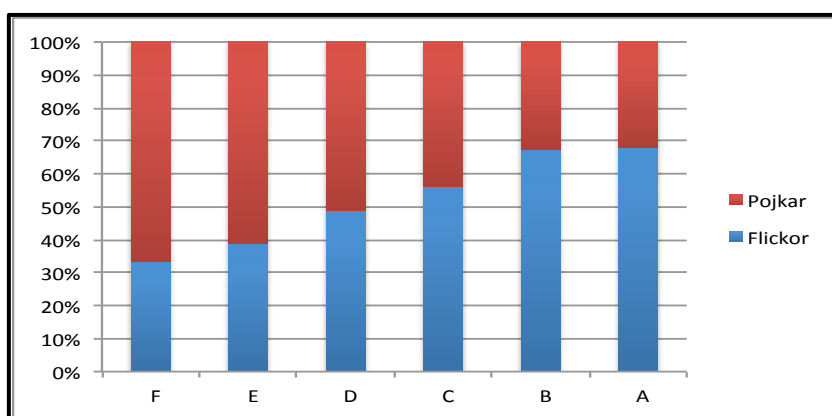


Diagram 2. Fördelning av provbetyg med avseende på kön/SCB. Åk9 2014

Flickor presterar bättre än pojkar på provet. Vad gäller slutbetyget i åk 9 har flickor i många år haft bättre resultat i ämnet än pojkar. Skillnaden mellan åk 6 och åk 9 är mycket små under 2014.

Vi har liksom i förra årets provrapport analyserat om provet missgynnar eller gynnar endera könet enbart utifrån variabeln ”kön”. Vi kan genomföra s.k. undergruppsanalyser för att se på hur de två undergrupperna

vi har data på i vårt urval, kön och om man följer kursplanen för SvA. Vi använder Mantel-Haentzels DIF-analys (Differential item functioning) för detta. Vi vill gärna se om enskilda uppgifter missgynnar/gynnar någon.

En mycket förekommande teori internationellt är att alla uppgifter skall vara neutrala avseende undergrupper. Detta kan vara mycket svårt att uppnå inom prov av samhällskunskapsämnets typ, eftersom uppgifterna ofta adresserar områden som kan var mycket engagerande för olika grupper. Att då åstadkomma helt neutralt ”uppförande” av uppgifterna skulle ibland ge väl ”uttunnade” uppgifter som inte vore så engagerande. Men motsatsen gäller också, dvs. uppgifterna får heller inte vara sådana att de berör någon grupp ”för mycket”, då blir de orättvisa ur provsynpunkt.

Fler uppgifter detta år än förra året är ”neutrala”. För att se på skillnade så används Mantel-Haentzels DIF-analys. Resultatet anges i steg: neutralt, gynnar svagt eller gynnar kraftigt (alternativt missgynnar).

Om man inte når att få helt igenom neutrala uppgifter, så kan en angreppspunkt vara att uppgifter

- 1) inte bör missgynna/gynna starkt, och
- 2) om det finns uppgifter som gynnar en grupp, så bör det finnas andra som gynnar den andra gruppen.

Vad gäller kön så ser tabellen för provet 2014 ut som följer (provet är sekretessbelagt, men för analysen så ge vissa indikationer angående uppgiftens beskaffenhet):

"Uppgift"	DIF	Format	CI
1	Gynnar flickor svagt	Öppen	Blogg
2	Gynnar flickor svagt	Öppen	Kränkningar
3	Gynnar flickor svagt	Kort öppen	Demokrati
4	Gynnar flickor svagt	Öppen	Framtid
5	gynnar pojkar starkt	Strukturerad	Val
6	Gynnar pojkar svagt	Strukturerad	Demokrati/Begrepp
7	gynnar pojkar svagt	Kort öppen	Källkritik
8	gynnar pojkar svagt	Strukturerad	Källkritik
9	Gynnar pojkar svagt	Strukturerad	EU
10	Gynnar pojkar svagt	Strukturerad	Internationellt
11	Gynnar pojkar svagt	Strukturerad	Media
12	Gynnar pojkar svagt	Strukturerad	Ekonomi

Tabell 2. DIF-analys avseende kön, Np Sb åk 9 2014

Uppgifterna är ”anonymiserade”, dvs. tilldelat uppgiftsnummer är slumpmässigt. Fördelning pojkar/flickor är jämn, dvs. det finns uppgifter där båda grupperna gynnas. Pojkarna gynnas av strukturerade svarsformat, vilket är känt sedan tidigare. De strukturerade är (till viss del) mer begreppsorienterade. Dock är uppgift 9, 10 och 11 inte utpräglad begreppsorienterad, utan mer sambandsbeskrivande. En fundering kan vara att innehållet i uppgift 1 och 2 ovan är mer engagerande för flickor?

Således, som helt prov betraktad så bör ingen grupp vara förfördelad, mixen av engagerande öppna uppgifter, och strukturerade uppgifter ser ut att ha fungerat väl ur denna infallsvinkel. En slutsats är också att provet således inte kan bestå av endast ett format.

Skillnader mellan de som läser och inte läser svenska som andraspråk (SvA)

På samma sätt som ovan har vi gjort analys på uppgifterna med avseende på om de gynnar eller inte gynnar de som läser SvA. Man kan tro att många uppgifter missgynnar de som läser SvA. I denna analys så jämförs de i de olika grupperna som har samma ”kunskapsnivå” i ämnet, på enklare svenska, de som är lika duktiga i ämnet, här definierat som provresultat bearbetat med de antaganden som görs i statistiska bearbetningen. De elever som läser SvA har i snitt betydligt sämre resultat på provet. Detta gäller för alla nationella prov, dock med viss skillnad mellan proven. Men med denna statistiska möjlighet, att jämföra de med samma förmåga i ämnet, och den enda särskiljande variabeln är att de läser svenska som andraspråk, så kan vi analysera uppgifterna i provet.

Samma typ av tabell som ovan, ”anonymiserade” uppgiftsnummer, men med viss indikation av uppgiftens beskaffenhet:

"Uppgift"	DIF	Format	CI
A	Gynnar SvA svagt	Öppen	Internationellt
B	Gynnar SvA svagt	Strukturerad	Ekonomi
C	Missgynnar Sva starkt	Strukturerad	Politisk struktur/begrepp
D	Missgynnar Sva svagt	Kort öppen	Källkritik

Tabell 3. DIF-analys avseende at följa kursplanen inom SvA, Np Sb åk 9 2014

Efterom kunskapskraven enligt Lgr11 inom samhällskunskap ofta bygger på att kunna redogöra för olika skeenden i samhället. Så det gäller att proven prövar elevernas kunskaper med hjälp av en mix av olika uppgiftstyper. Alla uppgifter får inte vara öppna uppgifter som kräver mycket skrivande av eleverna. Skulle så vara fallet finns en risk att man i viss utsträckning pröva elevernas förmåga att uttrycka sig som elevernas kunskaper i samhällskunskap.

De intressanta resultaten blir nu att endast fyra (sex 2013) av uppgifterna gynnar någon av grupperna, två gynnar SvA-gruppen och två de som inte läser SvA. En av uppgifterna missgynnar dock en grupp ”starkt”.

Redan förra året så såg vi att det innehållsmässigt fanns en intressant trend, de som läser SvA ser ut att klara sig bättre på uppgifter som frågar efter demokratins grundvalar och även på uppgifter som diskuterar ekonomi. Uppgift B ovan är en svårare uppgift som inte adresserar enbart begreppslig förståelse utan även förmågan att se samband i komplicerade ekonomiska frågor.

I år, liksom förra året, finns en indikation på att komplicerade strukturerade uppgifter kan missgynna SvA-eleverna. Dessa uppgifter tenderar att kräva en hel del läsförståelse, t.ex. uppgift C. Men uppgift B är också av den typen, och den gynnar SvA (svagt). Kanske innehållet som tar över, istället för format?

Validitet och reliabilitet

Validitet

Validity refers to the degree to which evidence and theory support the interpretations of test score for proposed uses of tests. Validity is, therefore, the most fundamental consideration in developing tests and evaluating tests. The process of validation involves accumulating relevant evidence to provide sound scientific basis for the proposed score interpretations.

Standards³

Enligt citatet ovan så innebär validitet: tolkningen av provresultaten stödjer det testet var avsett för (att mäta). All information som kan inhämtas bör således göras för att stödja arbetet med att se om proven är valida. Ett förenklat uttryck är att man mäter det man avser att mäta. Men citatet ovan går delvis bortom detta och fokuserar på vilka tolkningar vi gör, provet är inte valid *i sig*, det är ingen inneboende egenskap, utan de tolkningar vi gör är det vi skall se på.

Ett av problemen är just vilka bevis som används för att mäta korrelationen mellan den teoretiska definitionen och den operationella definitionen. I detta fall är den teoretiska definitionen kunskapskravens innehåll.

Det ett prov/test skall mäta kallas ofta internationellt för ett bakomliggande ”construct”⁴. Ett nationellt provs bakomliggande ”construct” är läroplanens kunskapskrav inom samhällskunskap för år 9.

För att pröva att provet adresserar vårt *construct* så analyserar vi noggrant dessa korta meningar i Lgr11. Inom konstruktionsgruppen har vi också alltid med medarbetare ”från fältet”, aktiva lärare. En eller flera av dessa byts varje år, för att vi skall få kontinuerlig ”input” från verksamheten. Under året så har vi också i större lärarsamlingar arbetat med att våra uppgifter skall ligga så nära våra delkunskapskrav⁵ som möjligt. Vi samlar också större lärargrupper vid ett antal tillfällen för att avgöra om uppgifterna ”ligger rätt” i förhållande till målgruppen. En av de större samlingarna kring detta är den s.k. bedömarveckan, som skedde v.40 2013 (skiftet sep/okt) inför detta prov. 16 lärare från 16 olika skolor medverkade, och från olika kommuner: Alingsås, Enskede, Falkenberg, Göteborg, Helsingborg, Hässelby, Nynäshamn, Oxelösund, Pajala, Skärholmen, Stockholm, Trosa, Ulricehamn, Vadstena, Västervik, Växjö, Ängelholm och Örnsköldsvik. Av de 16 lärarna var 6 män och 10 kvinnor.

Under veckan får lärare kommentera uppgifter och bedömningsanvisningar. Även vid det s.k. gränssättningsmötet medverkar lärare från hela landet och får där kommentera främst bedömningsanvisningarna (lärarna gör hela provet och huvuduppgiften är att ge underlag för betygsgränser). Mötet skedde 5 – 6 februari 2014, 17 lärare från 15 kommuner, från Hammerdal till Helsingborg medverkade.

Den mängdmässigt största valideringen görs dock vid prövningarna. Alla elever och medverkande lärare får fylla i enkäter i samband med prövningarna. Det blir således flera tusen enkäter över ett år. Inför provet 2015 kommer vi att utöka detta med ett lärarmöte till, enbart med fokus bedömningsanvisningar.

Under den senare delen av provets utveckling så har vår expertgrupp avseende ämnesinnehåll gått igenom alla uppgifter, som ett led i validitesprövningen. Expertgruppen består av personer från Statsvetenskapliga institutionen, Institutionen för sociologi och arbetsvetenskap, Institutionen för nationalekonomi,

³ Standards for Education and Psychological Testing (2014) är standardverket för amerikansk provtagning. Egentligen ett s.k. ramverk. Använd också mycket internationellt.

⁴ Vid en enkel översättningssökning på nätet så får man som förslag på *construct* t.ex. ”hypotetisk konstruktion” som i detta sammanhang torde väl svara mot ovanstående. Även ”tankeskapelse” är beskrivande. Ordet *construct* har nu blivit så etablerat att man också får det ”svenska” ordet ”konstrukt (akademiskt begrepp)”. Wikipedia anger ”*a hypothetical construct is an explanatory variable which is not directly observable*”.

⁵ Kunskapskraven i samhällskunskap år 9 kan lätt ses som sex olika innehållsmässiga delar. Vi kallar detta delkunskapskrav.

institutionen för kulturgeografi och institutionen för journalistik, media och kommunikation, alla vid Göteborgs universitet.

Under arbetet med bedömningsanvisningarna har vi kunnat använda flera statistiska analysmetoder för att se hur anvisningarna fungerat. Som ovan nämns har IRT (Item Response Theory) varit till mycket stor hjälp, och vi har under året haft möjlighet att utveckla dessa analysmetoder. Bland annat så bygger denna provrapport på en teknisk rapport baserat på urvaleleverna. Denna tekniska rapport publiceras inte, då den också är uppgiftspecifik avseende sekretessbelagda uppgifter.

Arbetet med IRT har inte minst hjälpt oss med arbetet att skriva säkrare bedömningsanvisningar, för att göra dessa valida i förhållande till uppgifter och elevsvar. Validitet innebär också att alla steg i kedjan Kunskapskrav-uppgift-prövning-bedömningsanvisning-tolkning av dessa genom lärare osv. skall fungera.

En gammal regel är också att ett prov aldrig kan vara valit om det inte är reliabelt.

Reliabilitet

Reliabilitet innebär att det valida provet ska mätas på rätt sätt. Ett test kan inte vara valit utan god reliabilitet. Provkonstruktörer världen över analyserar sina prov ofta med flera olika korrelationsmått för att kunna belysa provets för- och nackdelar på olika sätt. Vi har valt att analysera detta prov med hjälp av tre vanliga korrelationsmått.

Korrelation – cronbach alpha

Reliabilitetsmått avser ibland s.k. inre konsekvens, om uppgifter och elevsvar ”samverkar” så att det finns konsekvens i resultatet. Np Sh9 2014 har ett reliabilitetsmått på **0,93** (0,91 2013), vilket måste anses som bra eftersom det visar att ämnesprovets delar mäter samma sak. Vi har detta år också analysera reliabiliteten för de enskilda delkunskapskraven, då vi detta år gett möjlighet för lärare att frivilligt inrapportera sina elever i en databas (samma förfarande som när man rapporterar in de elever som medverkar i ”urvalet”). Vid den frivilliga inrapporteringen så får lärarna reda på ett medeltal för riket vad det gäller de enskilda delkunskapskraven. Detta ger då en möjlighet för lärarna att bedöma utfallet av den egna undervisningen med avseende på de olika delkunskapskraven.

Vad gäller reliabilitetsmått, så finns en internationell ”tumregel” som säger att om man skall återrapportera resultat till enskilda provtagare så skall reliabiliteten ligga på minst 0,9, och skall man återrapportera delar av prov på gruppnivå så skall reliabiliteten ligga på minst 0,7. I följande tabell ges värdena för Np Sh9 2014:

SH 9 14		
På hela provet		0,93
Per dkk/förmåga		
A	16 item	0,84
B	7 item	0,85
C	7 item	0,77

Tabell 4. Cronbach alpha Np Sh åk 9 2014

I ser i tabellen av provet i åk 9 2014 väl motsvarar gränsvärdena för individuell återrapportering för hela provet, och återrapportering av delkunskapskraven på gruppnivå (som alltså görs i den frivilliga inrapporteringen). Item är beteckningen på ett ”bedömningsområde”. Oftast är en uppgift ett item. Men en uppgift kan bedömas utifrån två olika delkunskapskrav, och därmed ge två item. Eller så kan flera uppgifter, t.ex. ett antal kortsvar, bedömas ”i klump”, dvs. bilda ett item. För samhällskunskap 9 år 2014 motsvarar alltid en uppgift ett item.

Korrelation-separation index

Ett annat mått på tillförlitlighet på provet är det s.k. Separation index. Detta anger i vilken grad som provet förmår skilja, separera, eleverna åt. Det gäller ju för ett prov att inte ”klumpa” eleverna för mycket utan man så träffsäkert som möjligt kan analysera deras förmåga i ämnet och separera eleverna åt. Vi utvecklar inte här bakomliggande teori men man brukar inom detta mått ange att provet skall vara över 2, över 3 är bra, över 4 mycket bra.

Separation index för NpSh9 2014:

Separation index	3,56
------------------	------

Tabell 5. Separation index Np Sh åk 9 2014

Korrelation-SEM

Genom att använda IRT vid utprovningar, ger det information om de olika uppgifternas svårighetsgrad i förhållande till elevernas olika förmåga/kunskap inom ämnet (eller egentligen visavi det som provet testar). Detta ger möjlighet att tillse att elever med olika förmåga alltid har uppgifter att arbeta med i provet. Både de elever som är ”duktiga” i ämnet och ”inte så duktiga” i ämnet måste ha uppgifter att arbeta med, annars förmår provet inte skilja eleverna åt på ett bra sätt. Det måste således finnas uppgifter för alla elever om provet skall vara ett bra prov.

För att testa om provet är sådant, kan man mäta med SEM, standard measurement error och TIF, test information function. Arbetet med att använda dessa informationskällor för att göra ett bra prov kommer att utvecklas under de närmaste åren.

Provet skall ju kunna ge underlag för bedömning av alla elever. Detta inte minst ur rättviseskäl. Dessa analyser kan bli relativt komplexa, betyder mycket för oss som provutvecklare, men går bortom denna rapport.

Men med hjälp av detta mått så kan vi här, även om vi inte går in i detaljer, se en del intressanta resultat:

Provet uppgifter har varit sådana att elever med både ”starkare” och ”svagare” förmåga i ämnet samhällskunskap har uppgifter att arbeta med, det finns uppgifter i Np Sh9 för alla elever. Det gäller dock att uppgifterna inom hela detta spann är av god kvalitet och säkerhet. Man kan säga:

Informationen om alla elever som tar provet, oberoende av resultat, eller i varje fall inom högsta och lägsta gräns för provbetyg, bör vara så god och säker så att resultatet för enskild elev kan anses med tillräcklig säkerhet givet.

Ett reliabilitetsmått som används för att mäta detta är SEM, Standard Error of Measurement. Det anges ibland som mer användbart än t.ex. cronbach alpha.

Standards for Education and Psychological Testing skriver:

The SEM is an indicator of a lack of consistency in the scores generated by the testing procedure for some population. A relatively large SEM indicates relatively low reliability/precision.⁶

SEM är således inte bara ett sammanfattande värde, utan ger värden över hela den skala man avser att mäta. En vanligt förekommande gräns för mycket god precision är ett SEM-värde under 0,3 på hela ”poängskalan” där man har sina *cut scores* (gränser för olika provbetyg). Detta mått är relativt ”hårt” men ger

⁶ The Standards for Educational and Psychological Testing 2014

en tydlig fingervisning om att den elev som har fått x poäng verkligen har det, och inte av slump någon annan gång skulle få några poäng färre eller fler.

NpSh9 14 har en lägsta cut score på 21 belägg och högsta på 64 belägg (gränserna mellan F/E och B/A respektive). Kurvan för SEM ser ut som följer:

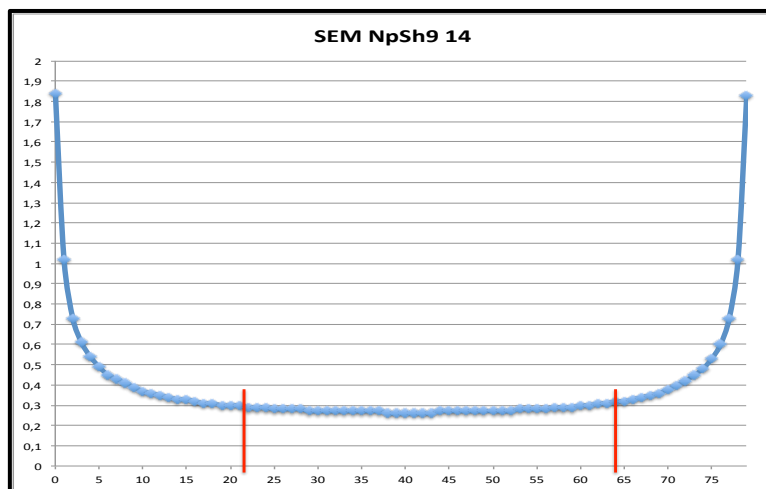


Diagram 4. SEM, NpSh 9, 2014

SEM sjunker under 0,3 vid 19 belägg och stiger över vid 62 belägg, och har 0,31 vid 64 belägg. Lägsta nivå är 0,26. Detta ett bra resultat. Provet har en mycket god precision vilket gör att vi kan uttala oss om elevernas antal belägg verkligen stämmer inom de gränser som provet har.

En mycket bidragande del i detta är att bedömningsanvisningarna särskiljer de olika nivåerna inom varje uppgift på ett bra sätt. SEM kan alltså användas för att utvärdera noggrannheten i bedömningsanvisningarnas texter.

Tillförlitligheten (reliabiliteten) på ämnesprovet är god. Detta är resultatet av relativt stora prövningar som har analyserats ur flera olika synvinklar, och med hjälp av flera olika metoder.

Många av dessa statistiska mått ovan kan verka väl tekniska för lärare som arbetar aktivt med proven. Men om inte är tillförlitligt både vad gäller validitet och mätsäkerhet, så blir provet med ett tydligare språk orättvist, det mäter inte det som var avsett att mäta. Och om det blir osäkert så drabbas någon, tyvärr vet vi då inte ens vem.

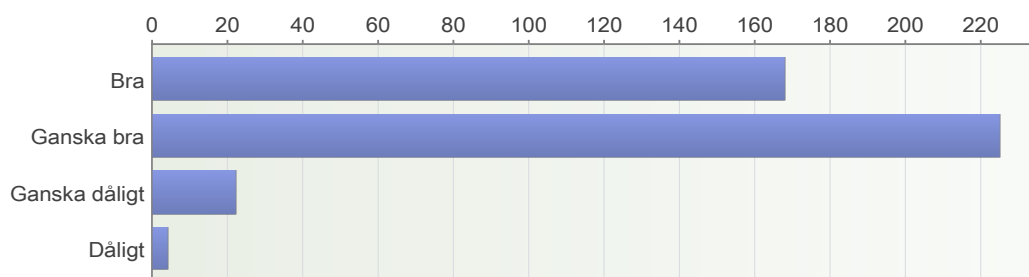
Nationella provet i samhällskunskap åk 9 2014 var ett – i många stycken – tillförlitligt prov.

Lärarenkäten

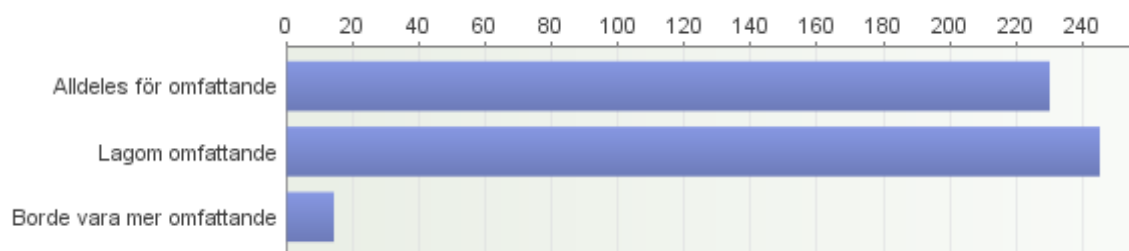
De strukturerade frågorna besvarades av nästan 500 lärare och där det fanns möjlighet att skriva kommentarer utnyttjades det av i snitt 200 lärare.

Det nationella provet i samhällskunskap har tagits emot väl av lärarna. Av de lärare som har svarat på enkäten tycker merparten att provet är bra eller ganska bra, 80 %. Knappt 2 % anser att provet är dåligt.

Vad tyckte du om ämnesprovet i samhällskunskap som helhet?



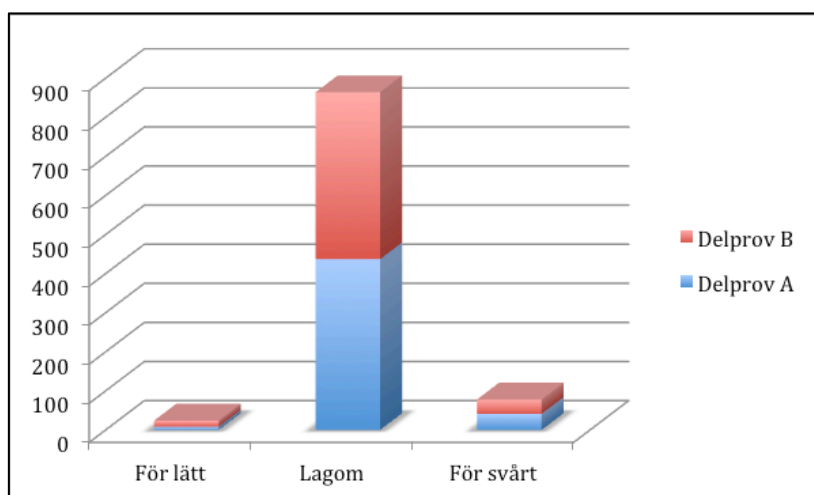
Fler lärare än förra året tyckte att provet var alldeles för omfattande, men det är också många som tycker att det är lagom omfattande.



De negativa kommentarerna handlar om att provet är för omfattande, att eleverna får mycket att skriva, det är mycket att bedöma och tillräcklig tid till detta har inte funnits. Det finns också kritik mot de öppna uppgifterna som anses för skrivkrävande för eleverna.

Provets svårighetsgrad upplevs som lagom av lärarna, den stora majoriteten har valt detta alternativ i enkäten.

Hur bedömer du svårighetsgraden i delproven?



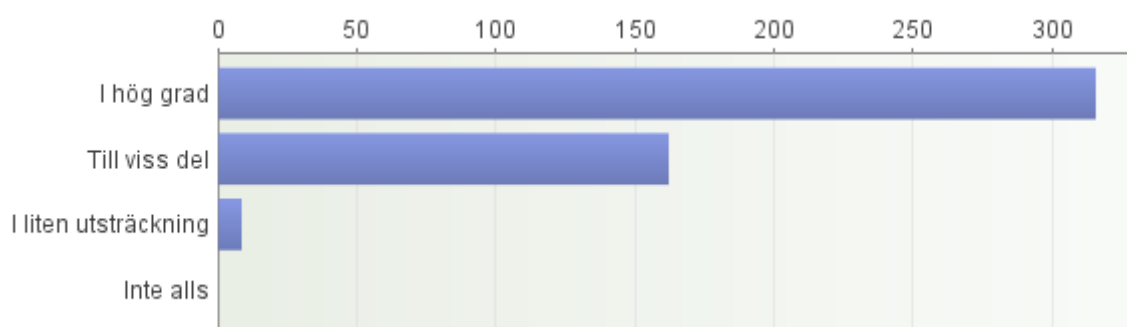
Det finns både positiva och negativa kommentarer kring provets svårighetsgrad, bland de positiva överväger kommentarer kring provets blandning av olika format ex. flervalsuppgifter och resonerande uppgifter *Jag tycker att det var aktuella och bra frågor. Allt från "basic" / allmänbildning till lite svårare och Mycket bra nivå på proven över lag!*

Bland de lärare som formulerar kritik så förekommer kommentarer att eleverna tydligare behöver få information om vad som efterfrågas samt att A-nivå på provet känns hög och E-nivå låg.

Det förekommer också kritik mot att språket är svårt för svaga läsare och för SvA, samt att ett av delkunskapskraven som vi testar får för stort utrymme.

När vi utformade provet var det viktigt för oss att uppgifterna innehöll varierade format och att det skulle finnas uppgifter som krävde skrivinsats och även uppgifter som inte krävde så mycket skrivinsats. Detta har lärarna uppfattat som positivt, 62 % av lärarna uppfattar balansen i provet mellan de olika formaten som bra.

Har provets innehåll behandlats i undervisningen?

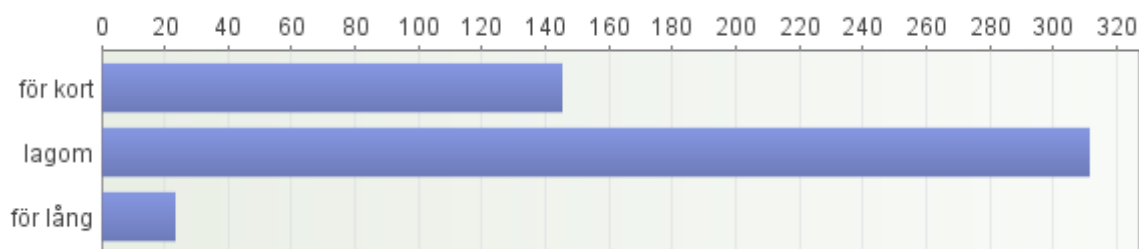


Bland kommentarer finns sådant som att *Innehållet har behandlats i undervisningen under de tre åren 7-9 och frågorna är ställda så att det verkligen är förmågorna som visas.*

De lärare som uttrycker kritik menar att provet i vissa fall berör sådant som eleverna har läst vissa områden i åk 7 vilket kan ha gjort att provet upplevs som svårare: *Men det spänner över tre år och alla kan inte minnas allt.*

Provet skrivtid var 120 minuter per delprov, detta upplevde majoriteten av lärarna som lagom. Lärarna kommenterar denna fråga i enkäten med att de flesta hann klart på avsatt tid men att provet var omfattande.

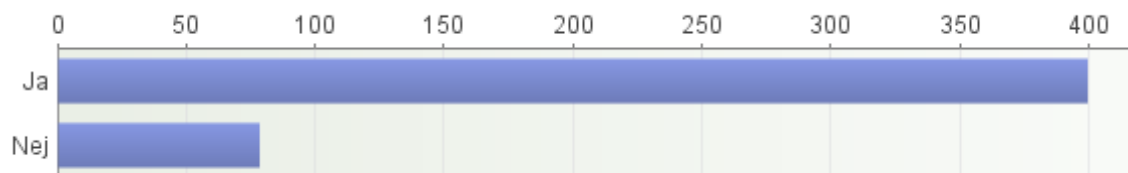
Skrivtiden (120 min) var...



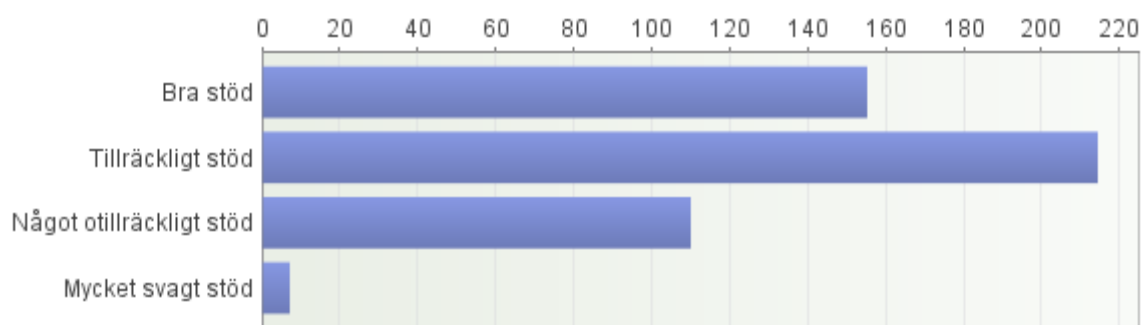
Eftersom ÄP samhällskunskap åk 6 har en rast inlagd i sitt prov, på varje delprov, så har vi i enkäten ställt frågan om lärarna tror att det skulle vara bra med rast för eleverna. Majoriteten tror inte detta eftersom de tror att eleverna då kan bli stressade och behöva göra avbrott i sina tankegångar. Minoriteten (de som ville ha rast i åk 9) var dock relativt stor, över 30%.

Tankebubblan – upplever du att eleverna hade hjälp av den?

Nytt för 2014 var att vi hade en tankebubbla med text i sig, i anslutning till de öppna frågorna. I denna text fanns information om vad som var viktigt för eleverna att tänka på när de besvarade frågan. Lärarnas uppfattning är att denna tankebubbla har varit till stöd för eleverna, vilket gör att vi kommer att fortsätta att använda denna form av stödstruktur.



Vad anser du om utformningen av häftet bedömningsanvisningar med avseende på vilket stöd det gav vid bedömning?



För att ett nationellt prov ska bli reliabelt behövs bedömningsanvisningar som lärarna upplever som stöd. De allra flesta lärarna upplever att bedömningsstödet gett ett bra eller tillräckligt stöd, 77 %. Endast 1 % anser att stödet är mycket svagt.

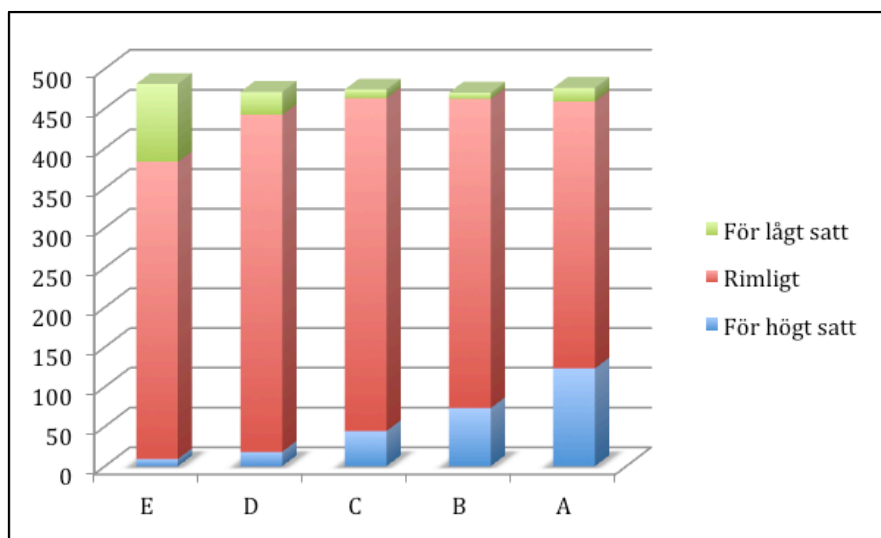
Bland kommentarerna finns att vi behöver förtydliga än mer vad vi menar i de olika bedömningsaspekterna – speciellt när eleverna ska konkretisera eller problematisera samt att det var små skillnader mellan nivå C och nivå A.

Exempel på lärare som upplevt bedömningsanvisningen som ett bra stöd: *Med exempeltexterna var de riktigt bra.*

Exempel på lärare som upplevt särigheter med konkretiseringar och problematiseringar: *Svårt att hålla isär konkretiseringar och problematiseringar, det gavs exempel som ibland tydde på att k och p var samma sak. Förvirrande*

Gränserna för de olika provbetygen är framtagna på gränssättningsmöten där aktiva lärare är med och tar fram gränserna.

De absolut flesta lärare i vår enkät anser att vi har tagit fram rimliga gränser för de olika provbetygen. Där man är mest tveksam är E-nivån som ganska många anser är för lågt satt. Fler än förra året anser också att vi har satt gränsen för A högt.



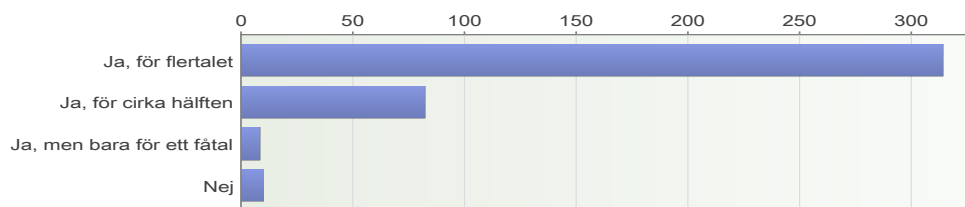
Kommentarer:

"Bra nivåer rakt igenom."

"Höj kraven för E."

"Ev för högt satt för betyg A. Det är lätt att missa genom att missförstå någon fråga".

Anser du att resultaten på provet ligger i linje med elevernas övriga prestationer under läsåret?



De allra flesta lärare anser att provresultatet ligger i linje med elevernas övriga prestationer. Bland kommentarerna finns sådant som att lärarna kanske har legat lågt i tidigare bedömning men också att de tycker att de har fler A-elever än vad nationella provet visar. Andra anser att eleverna nu verkligen fått tillfälle att visa vad de faktiskt går för:

"Mina elever presterade äntligen på sin nivå! De höjde sig och levererade det jag väntat på. Så därför, bättre prestation än tidigare, men inte pga att provet var för lätt, utan för att de tog det på allvar och visade de förmågor de besitter!"

Slutsatser

Ämnesprovet i samhällskunskap åk 9 genomfördes för första gången som betygsstödjande prov under vårterminen 2014.

Utifrån enkäterna har provet tagits emot bra och återspeglar lärarnas syn på ämnet och deras syn på elevernas kunskapsnivå. Man kan säga att provet har ”igenkänningsvaliditet” (face validity). Utifrån våra analyser har uppgifterna haft låg tröskel, så tillvida att eleverna ofta försökt lösa uppgifterna. Eftersom det varit ett uttalat mål för oss, är resultatet gott. Vid utvecklandet av uppgifter till kommande prov så måste denna ambition bibehållas.

Np Sh9 har haft en relativt jämn fördelning mellan uppgifter med öppna och strukturerade svarsalternativ. Under rubriken omfattning har lärare kommenterat att det upplevts som ”lagom” och ”för omfattande” i samma grad. Detta är en försämring mot förra året, som vi måste analysera vidare. Eftersom vi vill behålla antalet uppgifter för att ge eleverna goda möjligheter att visa sina kunskaper, måste arbetet med att utveckla strukturerade uppgiftsalternativ fortsätta. Analyserna för grupperna flickor/pojkar, SvA/Icke-SvA tyder också på det. Dock kommer ett ämnesprov i samhällskunskap alltid att behöva innehålla ett stort antal öppna frågor bl.a. beroende på kunskapskravens formuleringar.

I förhållande till slutbetygen i åk9 så ligger provbetyget mycket nära resultatet för betyget F.

Np Sh9 2014 ger många möjligheter att visa förmåga på nivå A.

I elevhäftena har den nya stödstrukturen ”tankebubblan” mottagits bra.

Bedömningsanvisningarna och den av oss valda layouten verkar också ha fallit väl ut.

Utifrån de analyser vi gjort har provet relativt höga värden för reliabilitet, bedömningsanvisningarna har kunnat åtskilja de olika nivåerna i elevsvaren, och provet som helhet ger eleverna möjlighet att hitta uppgifter för sin kunskapsförmåga på alla nivåer. Reliabilitetsmåttet har ökat mot förra årets ”försöksprov”. SEM ligger mycket bra till på (nästan) hela antalet förekommande belägg inom betygsstegen.

Områden för fortsatt utveckling:

- Ytterligare mer precisa och tydligare bedömningsanvisningar till de enskilda uppgifterna
- Ytterligare analys och utveckling av de strukturerade svarsalternativen, så att de på ett reliabelt och valit sätt mäter det som de är avsedda att mäta.
- Fortsatt arbete vad gäller validitet, med speciellt fokus på att uppgifterna ger eleverna möjligheter att uttrycka, och bedömningsanvisningarna att fånga kursplanens intentioner vad gäller svar på högre nivåer.
- Analys av ”omfattningen” av provet. Vad kan göras, med bibehållen mätsäkerhet, för att provet skall upplevas som inte fullt lika omfattande.
- Vi har redan idag ett stort utbyte med lärarna på fältet. Vi avser att öka det än mer, inte minst vid utformningen av bedömningsanvisningarna.
- Vi avser att fortsätta vårt fokus på interbedömarstudier, för att vinna säkerhet i bedömningsanvisningarna.
- Vi avser att än mer utöka vår repertoar av statistiska mått och jämförelser, så att vi på en än högre nivå kan analysera våra prov.
- I år inleddes försöken med frivillig inrapportering. Lärarna kunde där jämföra klassens resultat (på grupp-nivå) vad avser enskilda delkunskapskrav med rikets. Ett mindre antal lärare gjorde detta. Mycket litet information om detta gavs i Lärarinformationen 2014. Det avser vi att utöka.

Denna rapport kommer att publiceras på Skolverkets hemsida, So-ämnenas egen hemsida och provgruppens egen hemsida:

www.npsportal.se

www.ips.gu.se/forskning/forskningsprojekt/nationella-prov-samhallskunskap/

Referenser

Bond, T och Fox, C (2007) *Applying the Rasch Model*, Routledge 2007

Gustavsson, J-E., C.Cliffordsson C., Erickson G., *Likvärdig kunskapsbedömning i och av den svenska skolan – problem och möjligheter* SNS Förlag Stockholm 2013

Löfstedt, A (2015) *De första nationella proven i samhällskunskap – en studie i bedömersamstämmighet* Göteborgs universitet, Institutionen för pedagogik och specialpedagogik

Nering, M, och Ostini, R (red.) *Handbook of Polytomous Item Response Theory Models* , Routledge 2010

Standards for Educational and Psychological Testing (2014) American Educational Research Association (AERA), American Psychological Association (APA), och National Council on Measurement in Education (NCME)
Wilson, M, *Constructing Measures*. Psychology Press 2005

Bilaga 1. Utprövningar för Sh ÄP 6 och Sh ÄP 9 inför provet 2014

(Förklaring: 419 = inför 2014, I:a prövningen, åk 9)

1. 416 och 419

Totalt 17 häften (9 för åk 6 och 8 för åk 9) konstruerades innehållande 6-8 uppgifter per häfte.

Åk 6

Antal elever: 1440

Antal lärare: 53

Antal skolor: 45

Åk 9

Antal elever: 1284

Antal lärare: 40

Antal skolor: 36

2. 426 och 429

18 häften konstruerades (10 för åk 6 och 8 för åk 9) innehållande 5-6 uppgifter per häfte.

Åk 6

Antal elever: 1018

Antal lärare: 30

Antal skolor: 28

Åk 9

Antal elever: 1079

Antal lärare: 37

Antal skolor: 35

3. 436/516 och 439

11 häften konstruerades (7 för åk 6 och 4 för åk 9) innehållande 4-7 uppgifter per häfte.

Åk 6

Antal elever: 599

Antal lärare: 26

Antal skolor: 25

4 häften konstruerades

Åk 9

Antal elever: 466

Antal lärare: 15

Antal skolor: 10

4. 449/519

7 häften konstruerades innehållande 6-7 uppgifter per häfte.

Åk 9

Antal elever: 760

Antal lärare: 25

Antal skolor: 19