
Ämnesproven i grundskolans årskurs 6

Provrapport Samhällskunskap

Årskurs 6

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 6. Provet genomfördes 9 och 10 april 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.

Beskrivning av ämnesprovet och utprovningar

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 utprovningar av enstaka uppgifter har fyra större utprovningar genomförts. Under framtagandet av ämnesprovet har tusentals elever och hundratal lärare medverkar vid dessa provningar (3017 elever och 109 lärare från 198 skolor över hela landet för åk 6-provet). Se bilaga 1.

Ämnesprovet måste ge möjlighet för eleverna att visa kunskaper trots att det innehåll 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 till mycket hög grad försökt besvara uppgifterna (klart mer än 90 %). Vi har genomfört mindre underökningar 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 provning 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.

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

¹ Kunskapskraven i samhällskunskap åk 6 kan lätt ses som sex olika innehållsmässiga delar. Vi kallar detta delkunskapskrav.

Utseende

Siffrorna inom parentes nedan avser provet 2013.

Ämnesprovet i samhällskunskap i åk 6 bestod av två delprov, fördelat på fyra häften. Detta pga av en rast i varje delprov. Det fanns ingen principiell skillnad mellan delproven. Samhällskunskap åk 6 har relativt många uppgifter, 30 stycken. Eleverna har haft relativt många möjligheter att visa sina kunskaper. Detta för att få mätsäkerhet (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.

Np sh åk 6	Antal uppg./item	Upp till nivå			Antal strukturerade
		E	C	A	
2014	30	30	30	24	8
2013	29	30	29	20	12

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 åk 9 och den 6 dec för åk 6. Sedan förra året har de nationella proven inom So en digital inrapportering där lärare har rapporterat in ett urval av elever (födda den 10:e, 20:e och 30:e i varje månad). 1536 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 mycket betydelsefull för analys och framtagning av proven. I texten nedan används beteckningen ”urval” för dessa elever.

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

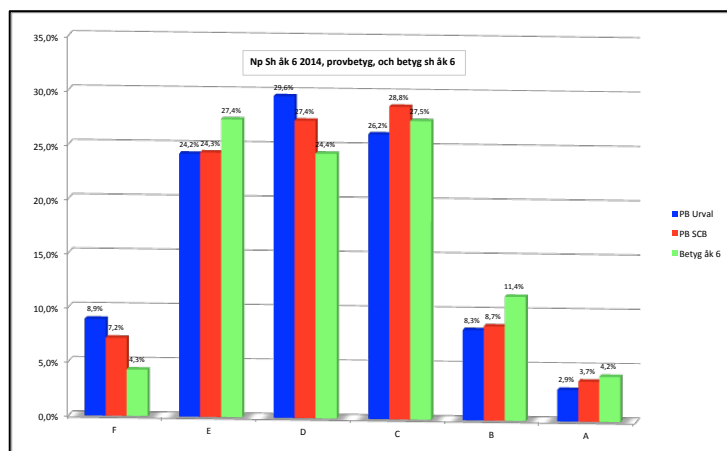


Diagram 1. Fördelning av provbetyg och slutbetyg samhällskunskap i ämnesproven i årskurs 6: /Urval provbetyg/ SCB provbetyg/ betyg åk 6 sh 2014.

Till skillnad mot åk9 så skiljer sig vårt urval och SCBs provbetyg (det egentliga provbetyget) något åt, främst på provbetyg F. Skillnaden finns inte i åk9. Vi fortsätter analysera detta.

De av lärarna satta betygen skiljer sig ju också åt, och skall så göra, de bygger på en mycket bredare bedömning. Skillnaderna är ändå inte så stora. 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

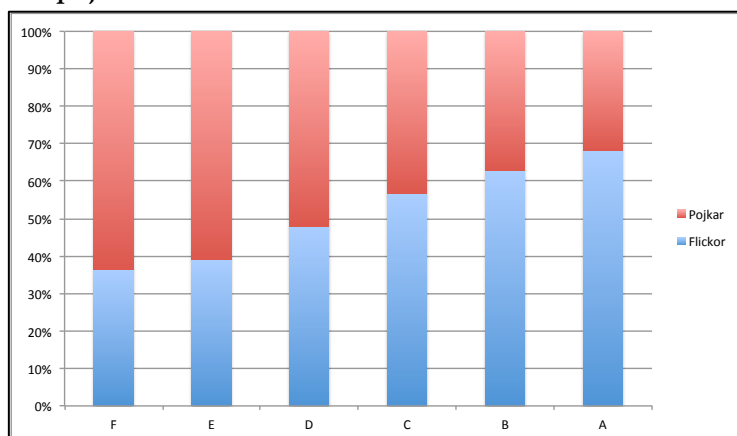


Diagram 2. Fördelning av provbetyg med avseende på kön/SCB. Åk6 2014

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”. Med Mantel-Haentzels DIF-analys (Differential item functioning) kan man göra 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. I detta fall uppdelad i steg, dvs. varje uppgift ges ett spann av varians där man anser att skillnaden mellan de studerade grupperna är: 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	Olika levnadsvillkor
2	Gynnar pojkar svagt	Sluten	Rättsliga begrepp
3	Gynnar pojkar svagt	Semi	Hur Sverige styrs
4	Gynnar pojkar svagt	Sluten	Påverka vem
5	Gynnar pojkar svagt	Sluten	Sh ekonomi
6	Gynnar pojkar svagt	Öppen	Uppgifter i rättsväsendet
7	Gynnar pojkar svagt	Sluten	Hur Sv styrs

Tabell 2. DIF-analys avseende kön, Np Sh åk 6 2014

Uppgifterna är ”anonymiserade”, dvs. tilldelat uppgiftsnummer är slumpmässigt. Fördelning pojkar/flickor är inte jämn, dvs. det finns uppgifter där båda grupperna gynnas, men pojkarna gynnas oftare (svagt). Ojämligheten är något större än förra året. Pojkarna gynnas av strukturerade svarsformat, vilket är känt sedan tidigare. De strukturerade är (till viss del) mer begreppsorienterade. En annan fundering kan vara att innehållet i uppgift A 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 behövs, så en slutsats är att provet således inte kan bestå av endast ett format.

Övriga 23 uppgifter gynnar/missgynnar inte avseende denna aspekt.

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 ”förmåga” i ämnet, på enklare svenska, de som är lika duktiga i ämnet (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	Hur kan man påverka

Tabell 3. DIF-analys avseende att följa kursplanen inom SvA, Np Sb åk 6 2014

Efterom kunskapskraven enligt Lgr11 inom samhällskunskap ofta bygger på att kunna redogöra för olika skeenden i samhället, så gäller det att provet prövar andra format än ren skrivförmåga som indikator på förmåga i samhällskunskap.

De intressanta resultaten blir nu att endast en (sex 2013) av uppgifterna gynnar någon av grupperna och då gynnar den SvA-gruppen (svagt).

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 A ovan är inom demokratiområdet.

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.

Standards²

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*”³. Ett nationellt provs bakomliggande ”*construct*” är läroplanens kunskapskrav inom samhällskunskap för åk 6.

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 delkunskapskrav⁴ 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 15-16 januari 2014, 18 lärare från 11 kommuner, från Skellefteå till Ystad 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

² Standards for Education and Psychological Testing (2014) är standardverket för amerikansk provtagning. Egentligen ett s.k. ramverk. Använd också mycket internationellt.

³ 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*”.

⁴ Kunskapskraven i samhällskunskap åk 6 kan lätt ses som sex olika innehållsmässiga delar. Vi kallar detta delkunskapskrav.

institutionen, Institutionen för sociologi och arbetsvetenskap, Institutionen för nationalekonomi, 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 varit till mycket stor hjälp här, 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 hjälp 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

Korrelation – cronbach alpha

Reliabilitet innebär att det valida provet ska mätas på rätt sätt. Ett test kan inte vara valit utan god reliabilitet. Reliabilitetsmått avser ibland s.k. inre konsekvens, om uppgifter och elevsvar ”samverkar” så att det finns konsekvens i resultatet. Np Sh6 2014 har ett reliabilitetsmått på **0,94** (0,92 2013), vilket måste anses som mycket 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 Sh6 2014:

Np Sh6 14			
På hela provet		0,94	
Per dkk/förmåga		Uppgifter	
A	nd strukturer	11	0,84
B	era påverka	6	0,78
C	mhällsfrågor	6	0,81
D	ternet källor	7	0,78

Tabell 4. Cronbach alpha Np Sh åk 6 2014

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. Utan att här utveckla bakomliggande teori så brukar man inom detta mått ange att provet skall vara över 2, över 3 är bra, över 4 mycket bra.

Separation index för NpSh6 2014:

Separation index	4,02
------------------	------

Tabell 5. Separation index Np Sh åk 6 2014

Korrelation-SEM

Genom att använda IRT vid utprovningar, ger det information om de olika uppgifternas svårighetsgrad i förhållande till elevernas olika förmåga 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 så kan vi se en del intressanta resultat:

Provets 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 Sh6 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.

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 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

NpSh6 14 har en lägsta cut score på 24 belägg och högsta på 72 belägg (gränserna mellan F/E och B/A respektive). Kurvan för SEM ser ut som följer:

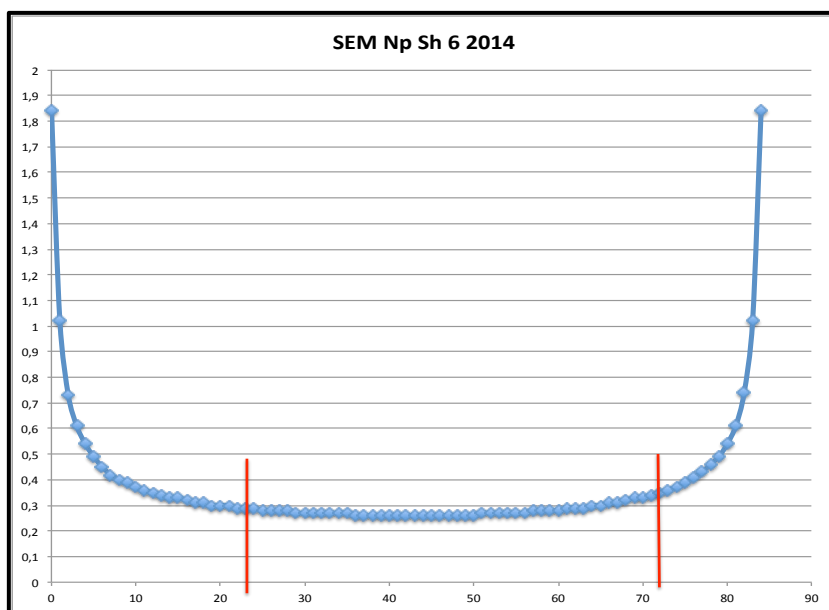


Diagram 4. SEM, NpSh 9, 2014

SEM sjunker under 0,3 vid 19 belägg och stiger över vid 62 belägg, och har 0,31 vid 66 belägg. Vid gränsen 72 belägg är SEM 0,35. Lägsta nivå är 0,26. Detta ett mycket 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.

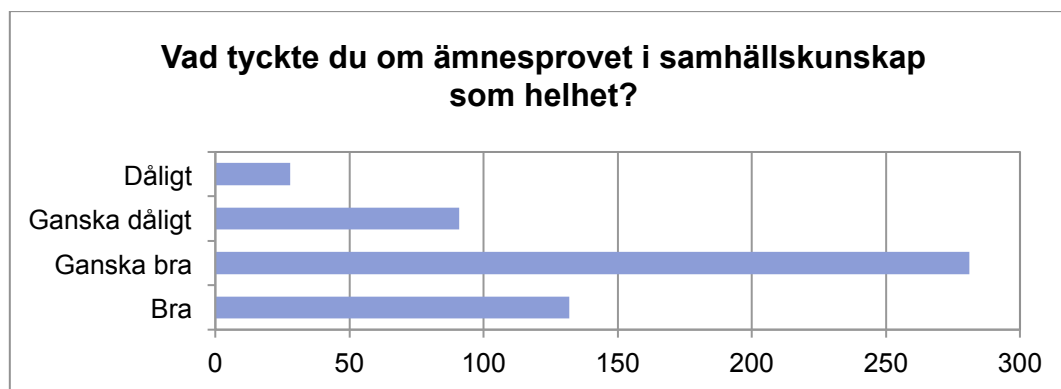
Många av dessa statistiska mått ovan kan verka väl tekniska för lärare som arbetar aktivt med proven. Men om inte är tillförlitligt både vad gäller validitet och mätsäkerhet, så blir provet med ett tydligare språk orättvist, det mäter inte det som var avsett att mäta. Och om det blir osäkert så drabbas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

Nationella provet i samhällskunskap åk 6 2014 var ett – i många stycken – tillförlitligt prov.

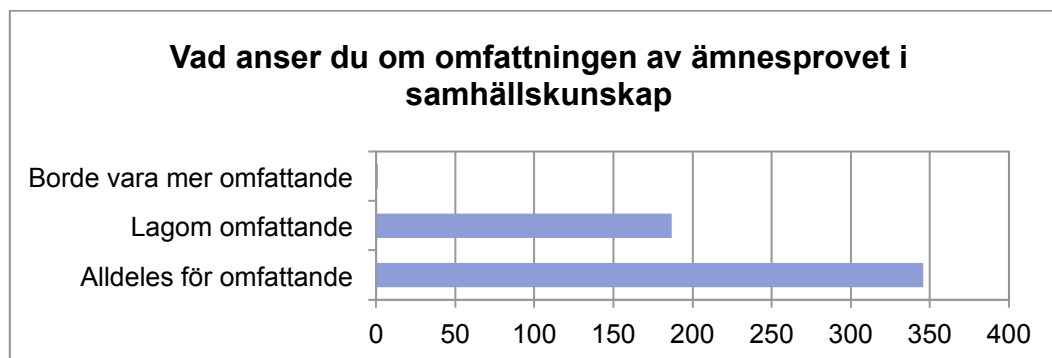
Lärarenkäten

De lärare (ca 1500) som genomförde nationella prov i samhällskunskap uppmanades att svara på en webb-enkät. De strukturerade frågorna besvarades av 539 lärare och där det fanns möjlighet att skriva kommentarer utnyttjades det av mellan 18 och 494 lärare. Svaren på de frågor med svarsalternativ ger på det stora taget en positiv bild av provet medan det i kommentarerna framkommer kritik kring mer specifika delar.

Det andra ämnesprovet i samhällskunskap har enligt enkäten tagits emot bra. 78 % av de svarande angav ”bra” eller ”ganska bra” i sitt svar. 5 % upplevde provet som dåligt.

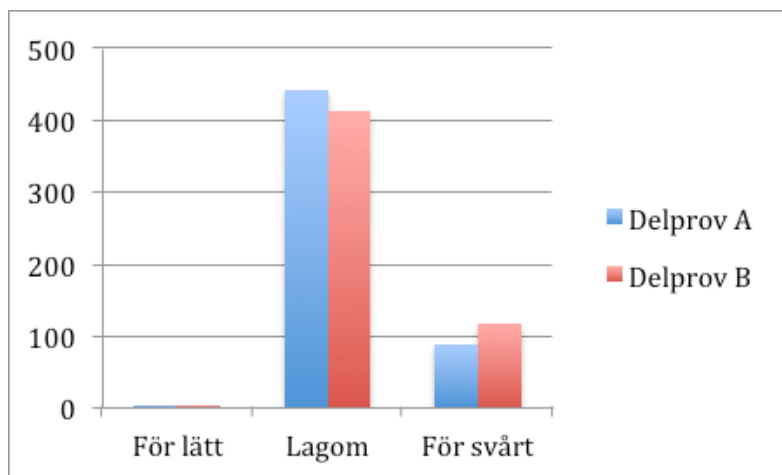


Det som riktas mest kritik kring i enkäten är omfattningen av provet och då framförallt antalet öppna uppgifter. Föregående år var det vanligaste svaret att provet var lagom omfattande. I årets enkät anser 65 % av de svarande lärarna att provet var för omfattande. I kommentarerna kan man läsa att uppgifterna som sådana är uppskattade men att det blev för mycket att skriva för eleverna. I enkäten framkommer också att bedömningsbördan för lärarna blev omfattande eftersom öppna svarsformat är mer krävande att bedöma än strukturerade.

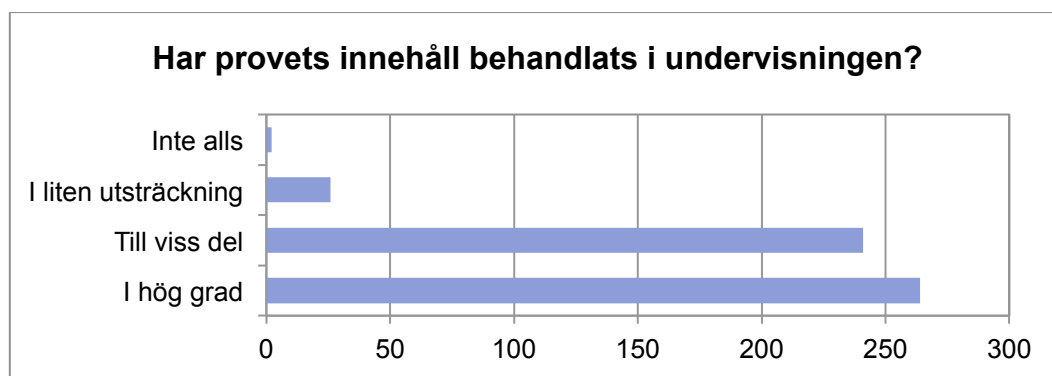


På frågan om svårighetsgraden i respektive delprov så anser lärarna i enkäten att provet hade lagom svårighetsgrad. Delprov B anses som något svårare och i kommentarerna kan man se att många lärare kopplar det till omfattningen och antalet öppna uppgifter. Andelen som uppfattade provet som helhet som för lätt var 1 %, lagom 80 % och för svårt 19 %.

Hur bedömer du svårighetsgraden i respektive delprov?

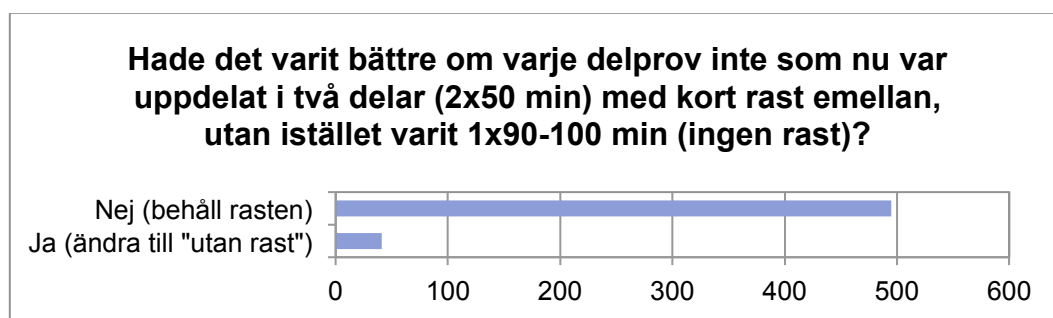


När det gäller frågan om innehållet på provet hade behandlats i undervisningen svarar 50 % ”I hög grad” samt 45 % ”Till viss del”.



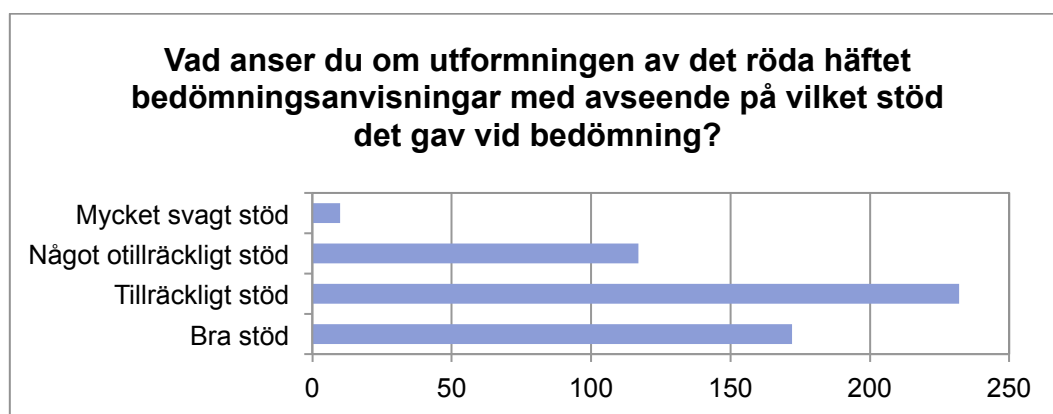
Det som framkommer i kommentarerna är att specifika innehåll ännu inte hunnits gå igenom, eller att det var länge sedan ett innehåll lyftes i undervisningen. Flera kommentarer berör också att läraren inte haft eleverna under så lång tid och därmed inte har full kontroll på vad de gått igenom.

Ämnesprovet i samhällskunskap har valt att lägga en rast i mitten av varje delprov. Enkäten visar att 92 % av lärarna tyckte att rasten var bra vilket är samma resultat som förra året.



Kommentarerna tar upp att rasten är viktig för eleverna både för att det underlättar att "få lite luft" men också för att delproven uppfattades som skrivtunga.

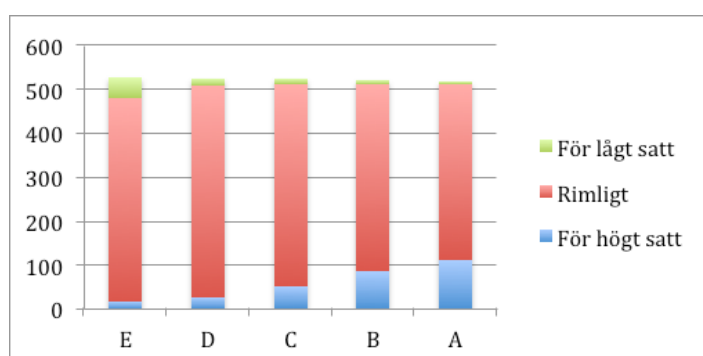
76 % av lärarna i enkäten ansåg att bedömningsanvisningarna gav ett *tillräckligt* eller *bra stöd* vid bedömningen. 2 % ansåg att det var ett mycket svagt stöd.



De kommentarer som berör bristerna tar upp svårigheten att bedöma öppna svar, som exempel ges att det är svårt att avgöra om konkretiseringarna eleven gör är tillräckliga eller inte och att elevernas svar ofta skiljer sig från de exempelsvar som återfinns. Det framkommer också kritik om att det inte i vissa uppgiftsformuleringar tillräckligt tydligt framgår vad som förväntas av eleverna på de olika nivåerna. När det gäller exemplen på elevsvar som återfanns i bedömningsanvisningarna anser 49 % av lärarna att exemplen var "till god hjälp" medan 3 % ansåg att de var "till ringa hjälp".

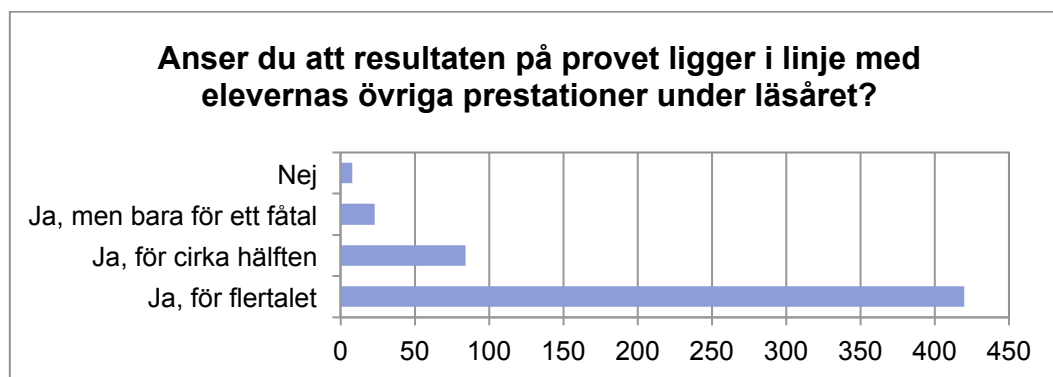
I diagrammet framgår det att det är nivån för A som flest anser är för högt satt.

Vad anser du om gränserna för respektive provbetyg?



Det finns kommentarer både kring att nivåerna var för höga och att de var för låga. Främst framkommer det att nivån för A var hög med tanke på att en del elever inte hade tid att svara i den utsträckning som de hade förmåga till på grund av provets omfattning.

På frågan om provresultatet ligger i linje med elevernas övriga prestationer under läsåret svarar 94% ”Ja, för flertalet” och ”Ja, för cirka hälften”.



Även på denna fråga framkommer att vissa elever inte haft tid att prestera på sin nivå på grund av tidsbrist men även mer positiva kommentarer återfinns, ”Ja det kändes bra att även de elever som inte ”visar sig” så mycket på lektionerna i form av muntliga prestationer presterade så bra på provet. Provet känns som ett bra stöd för min bedömning.”

Sammanfattningsvis visar enkäten att lärarna ansåg att provet var för omfattande med för många öppna uppgifter men att uppgifterna som sådana var uppskattade. I detta sammanhang visar kommentarerna att man i begreppet ”svårt” även lägger in omfattningen som en faktor i svårighetsgraden. Det framförs också önskemål om att förtydliga för eleverna vad som förväntas av dem i uppgifterna samt att bedömningsanvisningarna kan göras mer konkreta. Provets innehåll verkar uppfattas positivt av lärarna samt att nivåerna för de olika provbetygsgränserna låg bra.

Slutsatser

Ämnesprovet i samhällskunskap åk 6 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”. 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 måste denna ambition bibehållas.

Np Sh6 har haft fler uppgifter med öppna än strukturerade svarsalternativ. Under rubriken omfattning har lärare kommenterat att det upplevts som ”för omfattande” dubbelt så ofta som ”lagom”. Detta är en klar 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, tyder också på det. Däremot är provet nästan helt neutralt vad gäller undergrupperna SvA/Icke-SvA . Dock kommer ett ämnesprov i samhällskunskap alltid att behöva innehålla ett stort antal öppna frågor bl.a. beroende på kunskapskravens formuleringar.

I förhållande till betygen i åk6 så ligger provbetyget relativt nära.

Np Sh6 2014 ger många möjligheter att visa förmåga på nivå A.

I elevhäftena har den stödstrukturen ”ugglan Sam” även i år mottagits bra.

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

Utifrån de analyser vi gjort har provet 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. Reliabilitetsmått 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 de på ett reliabelt och valid 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.sobedomning.se/npsportal

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