
Ämnesproven i grundskolans årskurs 9 och specialskolans årskurs 10

Provrapport Samhällskunskap

Årskurs 9

Vårterminen 2016



Innehåll

Inledning.....	3
Konstruktionsprocess för provet	3
Bakgrund.....	3
Provets sammansättning.....	4
Bedömningsanvisningar.....	4
Provresultat med kommentarer.....	4
Resultat fördelat på faktorererna kön och undervisning i SvA.....	5
Validitet och reliabilitet	8
Validitet	8
Reliabilitet	9
Enkätresultat med kommentarer	12
Avslutning	17
Referenser	18
Bilaga 1: Utprövningar 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 de nationella proven i samhällskunskap för årskurs 9. Denna rapport behandlar ämnesprovet i årskurs 9. Provet genomfördes den 17 och 19 maj 2016.

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 cirka en fjärdedel av de möjliga lärarna 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.

Konstruktionsprocess för provet

Bakgrund

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

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 del av kunskapskraven¹. 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 tre större utprovningar genomförts inför proven 2016. Under framtagandet av ämnesprovet har tusentals elever och hundratals lärare medverkar vid dessa provningar (2620 elever och 90 lärare från 82 skolor över hela landet - från Haparanda till Veberöd - för årskurs 9-provet). Se bilaga 1.

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 "stödja" åt fel håll varför vi försökt hitta en lagom nivå. Även till provet 2016 har vi haft en s.k. "tankebubbla" till de öppna uppgifterna. Enligt lärarenkäten (se nedan) har detta fallit väl ut.

¹ Kunskapskraven i samhällskunskap årskurs 9 kan lätt ses som sex olika innehållsmässiga delar. Vi kallar detta "delar av kunskapskraven".

Provets sammansättning

Ämnesprovet i samhällskunskap i årskurs 9 bestod av två delprov. Det fanns ingen principiell skillnad mellan delproven. Det nationella provet i samhällskunskap för årskurs 9 har de senaste åren minskat antalet uppgifter totalt. Från 30 stycken 2013, till 26 stycken för 2016. Vi har också arbetat aktivt för att minska andelen helt öppna uppgifter, som ofta kräver ett längre svar. Problemet blir att bibehålla hög validitet och hög reliabilitet med färre antal uppgifter och fler av strukturerad natur. Mer om detta utvecklingsarbete nedan.

Nationella provet i samhällskunskap för 2016 har den hittills högsta siffran för reliabilitet på hela provet som något prov inom SO haft, 0,941 (Cronbach alpha), vilket indikerar att vårt utvecklingsarbete fungerar.²

Np sh åk 9	Antal uppg./item	Upp till nivå			Antal strukturerade	
		E	C	A		
2013	30	30	29	16	19	
2014	30	30	30	19	12	
2015	28	28	24	18	9	(av de 19 öppna var 5 s.k. Kortsvar)
2016	26	26	26	24	10	(av de 16 öppna var 4 s.k. Kortsvar)

Tabell 1. Fördelning av uppgifter, nivåer och format 2013-2016, ämnesprovet i samhällskunskap

Bedömningsanvisningar

Inga provbetyg ges direkt av elevens svar på uppgifterna. De måste först bedömas av en lärare med våra bedömningsanvisningar som stöd. 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 tidigare 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.

Provresultat med kommentarer

Resultaten från provbetygen återrapporterades av SCB den 30 november för årskurs 9. Under alla år proven genomförts har de nationella proven inom SO haft en digital inrapportering där lärare har rapporterat in ett urval av elever (födda den 6:e, 16:e och 26:e i varje månad). 1913 elever i detta ”samplade”³ urval 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 årskurs 9 2016 motsvarade en uppgift ett item.

³ Elever med födelsedatum 6, 16 och 26 i varje månad har rapporterats in.

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.

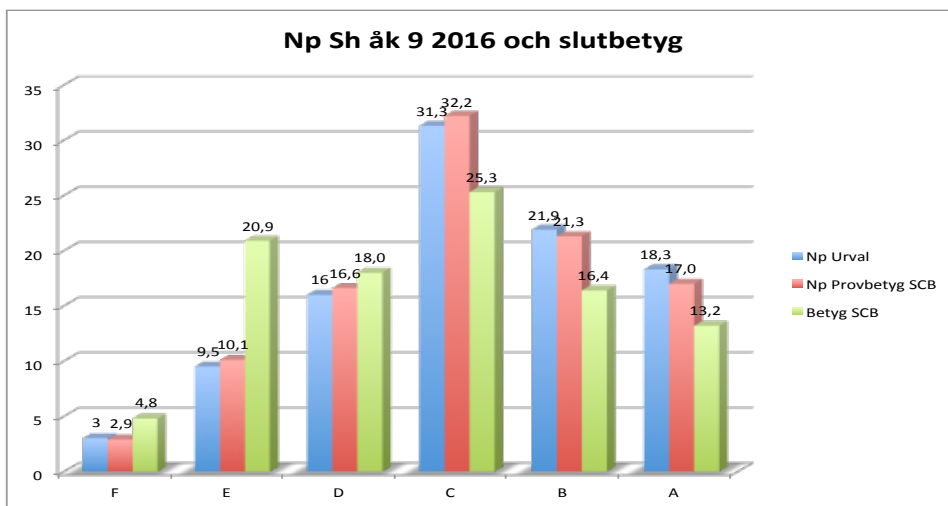


Diagram 1. Fördelning av provbetyg och slutbetyg samhällskunskap i ämnesproven i årskurs 9: /Urval provbetyg/SCB provbetyg/ slutbetyg 2016.

I enkäterna tycker de flesta lärarna att provbetygen är ”rimligt” satta, förutom att provbetygsnivån för F/E är för lågt satt. I år har en viss förskjutning av provbetygen skett, åt de högre nivåerna. Detta kommenteras nedan under punkten 3 d) ”Kravgränssättningsmöten”, i nästa kapitel Validitet och Reliabilitet

Resultat fördelat på faktorerna kön och de som följer kursplanen för SvA.

Skillnader mellan pojkar och flickor

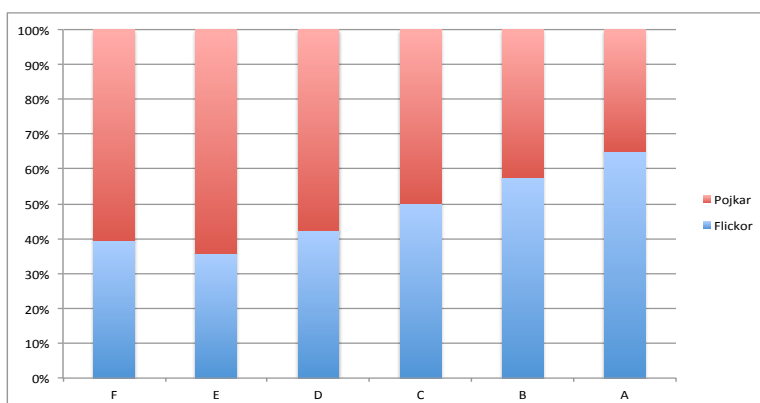


Diagram 2. Fördelning av provbetyg samhällskunskap med avseende på kön/SCB. Årskurs9 2016

Flickor presterar bättre än pojkar på provet. Vad gäller slutbetyget i årskurs 9 har flickor i många år haft bättre resultat i ämnet än pojkar.

Vi har liksom i förra årets provrapport analyserat om provet missgynnar eller gynnar något av könen 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 med syfte att se om enskilda uppgifter missgynnar/gynnar någon.

En vanligt förekommande teori internationellt är att alla uppgifter ska vara neutrala avseende undergrupper. Detta kan vara mycket svårt att uppnå inom prov av samhällskunskapsämnet eftersom uppgifterna ofta adresserar områden som kan vara olika 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. För att göra ett mycket enkelt exempel kan val av idrott i en fråga om ungdomars idrottande eventuellt vara olika engagerande för flickor respektive pojkar. Fler uppgifter detta år än förra året är ”neutrala”. För att se på skillnaden så används Mantel-Haentzels DIF-analys. Resultatet anges i steg: neutralt, gynnar ”svagt” eller gynnar ”starkt” (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 2016 ut som följer (provet är sekretessbelagt, men för analysen så ge vissa indikationer angående uppgiftens beskaffenhet):

"Uppgift"	DIF	Format	Innehåll
1	Gynnar flickor, svagt	Öppen	Internet
2	Gynnar flickor, svagt	Öppen	Internet kränkningar
3	Gynnar flickor, svagt	Semiöppen	Källkritik, internet
4	Gynnar pojkar svagt	Strukturerad	Ekonomi
5	Gynnar pojkar svagt	Strukturerad	Ideologier
6	Gynnar pojkar svagt	Strukturerad	Demokrati
7	Gynnar pojkar svagt	Strukturerad	Källkritik

Tabell 2. DIF-analys avseende kön, Np Sb årskurs 9 2016

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. Ingen av uppgifterna under 2016 gynnar någon grupp ”starkt”. Man kan fundera över om inte årets fördelning är så jämna man kan komma, få uppgifter – men inte alla – gynnar/missgynnar, grupperna är lika stora, ingen gynnar/missgynnar ”starkt”.

Detta visar att provet inte endast kan bestå av ett format.

Skillnader mellan de som läser och inte läser svenska som andrasprårskurs (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 i provresultatet. De elever som läser SvA har i snitt 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 andrasprårskurs, så kan vi analysera uppgifterna i provet.

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

NpSh9 2016

	"Uppgift"	DIF	Format	Innehåll
A13	A (6)	Missgynnar SvA, svagt	Strukturerad	Demokrati
B01	B (7)	Missgynnar SvA, svagt	Strukturerad	Källkritik
B04	C	Missgynnar SvA, svagt	Strukturerad	politisk struktur
A04	D (5)	Gynnar SvA, svagt	Strukturerad	Ideologier
A06	E (1)	Gynnar SvA, svagt	Öppen	Internet
B12	F	Gynnar SvA, svagt	Strukturerad	Demokrati

*Tabell 3. DIF-analys avseende att följa kursplanen inom SvA, Np Sb årskurs 9 2016
Siffror inom parentes hänvisar till att uppgiften också finns med i tabell 2, DIF för kön.*

Eftersom 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 kan inte vara öppna uppgifter som kräver mycket skrivande av eleverna. Skulle så vara fallet finns en risk att man i viss utsträckning istället prövar elevernas förmåga att uttrycka sig i skrift.

Ingen av uppgifterna i provet gynnar/missgynnar någon grupp ”starkt”.

Tidigare år har vi sett 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. I tabellen ovan stämmer detta för uppgifterna som gynnar SvA-gruppen. Men en av uppgifterna (A i tabell 3 = 6 i tabell 2) stämmer inte detta. Denna uppgift innehåller alternativ som är relativt språrkursligt utvecklade, vilket kan vara en nackdel för elever som följer kursplanen inom SvA.

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.

Standards⁴

Enligt citatet ovan så innebär validitet: tolkningen av provresultaten stödjer det som 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 inte har visat tillräckligt hög validitet *i sig*, det är ingen inneboende egenskap, utan de tolkningar vi gör är det vi skall se på.

Valid kommer från latinet och betyder från början ”stark”, och svensk betydelse idag snarast ”giltig”. Starkt giltig blir då en bra beskrivning av validitet.

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 konstrueras för att mäta kallas ofta internationellt för ett bakomliggande ”*construct*”⁵. Ett nationellt provs bakomliggande ”*construct*” (oftast idag ”konstrukt”) är kursplanen i ämnet, och bedömningen via kunskapskraven för oss de inom samhällskunskap för årskurs 9.

Vår valideringsprocess kan beskrivas i tidsordning på följande sätt:

- 1.. Första ledet i valideringen är att vi själva analyserar kursplanen och dess kunskapskrav ytterst noggrant och funderar över operationaliseringarna i form av uppgifter, och vilka effekter på validiteten som olika format kan ha.
2. Valideringen visavi avnämarna. Detta genomförs genom att vi till gruppen knyter 2 – 3 medarbetare ”från fältet”, aktiva lärare som arbetar i gruppen 1-2 dagar per vecka kontinuerligt över hela året. En eller flera av dessa byts varje år, för att vi skall få kontinuerlig ny ”input” från verksamheten.
3. För att få än större inflytande från verksamheten, så genomför vi ett antal åtgärder:

a) Alla elever och lärare som deltar i utprovningen av uppgifterna får ge sina synpunkter på dem i en enkät. Till detta prov totalt 2620 elever och 90 lärare från hela Sverige. Lärarna slumpas delvis och tas delvis från de som via enkäten till proven anmält intresse. Genom analys av alla enkäter får vi en bra uppfattning om hur frågorna tas emot.

b) Under året har vi ett antal större lärarsamlingar. En av dessa, kallad ”bedömarveckan” (BV, numera tre dagar, var tidigare en hel vecka), skedde v. 23, 1-3 juni 2016 inför detta prov. 15 lärare från 15 olika skolor i 14 kommuner medverkade. Till BV-dagarna brukar proportionen ”gamla/nya” vara 50/50, dvs hälften av lärarna har varit med någon gång förut för att leda arbetsprocessen i grupperna, hälften skall vara nya, dock har alla genomfört någon provning.

⁴ Standards for Education and Psychological Testing (2014) är standardverket för amerikansk provtagning. Egentligen ett s.k. ramverk. Använd också mycket internationellt.

⁵ 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*”.

Under veckan får lärarna genomföra bedömningar av alla öppna uppgifter som prövats med hjälp av bedömningsanvisningar framtagna av provgruppen. Därefter bearbetas både uppgift och bedömningsanvisning av lärarna där de kommer med förslag till förändringar och förbättringar

c) Under hösten, när provet går mot att vara klart, samlar vi en mindre grupp lärare som får arbeta med utformningen av bedömningsanvisningar (BA) till ett antal öppna uppgifter som vi bedömer som särskilt intressanta, Dessa två dagar kallas BA-dagar och skedde till detta prov den 24-25 oktober, 13 lärare, 6 ”gamla” och 6 ”nya”.

d) sista valideringen sker vid det s.k. kravgränssättningsmötet då lärare från hela landet och får där kommentera främst bedömningsanvisningarna (lärarna gör hela provet och huvuduppgiften är att ge underlag för betygsgränser). Mötet skedde 2-3 februari 2016, 15 lärare från 15 kommuner, från Piteå till Karlshamn medverkade.

I år fick vi klara indikationer vid prövningar och under detta kravgränssättningsmöte, att några uppgifter upplevdes som relativt ”svåra”, av både lärare och elever. Denna effekt justerades för i anpassningen av gränser för provbetyg. När eleverna genomförde provet så såg vi inte denna effekt alls hos dessa uppgifter, vilket innebar en viss förskjutning åt högre provbetyg.

4. Under den senare delen av provets utveckling så har vår expertgrupp avseende ämnesinnehåll gått igenom alla uppgifter, som ett led i validitetsprövningen. Expertgruppen består av personer från Statsvetenskapliga institutionen, Institutionen för sociologi och arbetsvetenskap och Institutionen för nationalekonomi, alla vid Göteborgs universitet. De är ämnesföreträdare, men har alla vid något tillfälle i sina karriärer undervisat på ämneslärarutbildningarna.

En sista del i vår kontinuerliga validering utgör också vårt arbete med våra statistiska metoder. 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 uppgiftsspecifik avseende sekretessbelagda uppgifter.

Arbetet med IRT har 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.

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 ofta sina prov 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. Äp Sh9 2016 har ett reliabilitetsmått på **0.941** (2015: 0.932, 2014: 0.925, 2013: 0.911), vilket är det högsta som något SO-prov hitills haft. Vi har detta år också analyserat reliabiliteten för de enskilda delarna av kunskapskraven som provet prövar, 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 delarna av kunskapskraven. Detta ger då en möjlighet för lärarna att bedöma utfallet av den egna undervisningen med avseende på de olika delarna av kunskapskraven.

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 grupp nivå så skall reliabiliteten ligga på minst 0,7. I följande tabell ges värdena för Np Sh9 2016:

Tabell 4. Cronbach alpha Np Sh årskurs 9 2016

NpSh9 2016			
På hela provet		0,9411	
Per dkk/förmåga			
1	<i>Samband-strukturer</i>	13 item	0,881
3	<i>Samhällsfrågor</i>	7 item	0,866
6	<i>Källor och trovärdighet</i>	6 item	0,754

I tabellen syns att värdena väl motsvarar gränsvärdena för individuell återrapportering för hela provet, och återrapportering av delarna av kunskapskraven på grupp nivå (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 delarna av kunskapskraven, 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 2016 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. 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 2016:

Separation index	3,88
------------------	------

Tabell 5. Separation index Np Sh årskurs 9 2016

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.

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, vilket innebär att 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 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 16 har en lägsta ”cut score” på 17 belägg och högsta på 58 belägg (gränserna mellan provbetyg F/E och B/A respektive). Kurvan för SEM ser ut som följer:

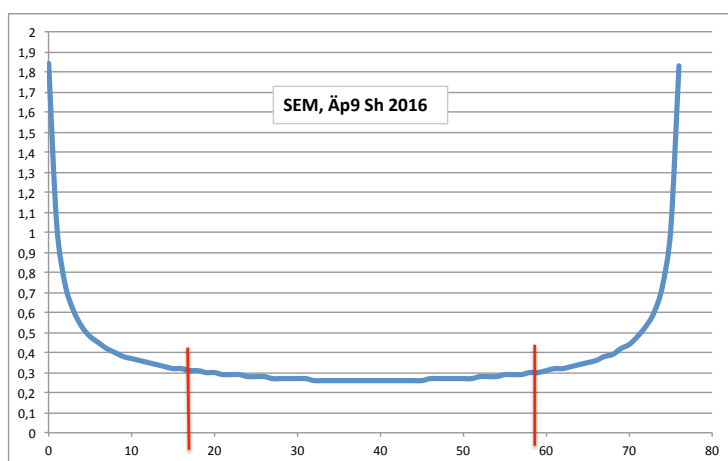


Diagram 4. SEM, NpSh 9, 2016

SEM sjunker under 0,3 vid 19 belägg och stiger över vid 62 belägg, och har 0,31 vid 60 belägg. Lägsta nivå är 0,26 (mellan 32 – 45 belägg). 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 språrskurs 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 årskurs 9 2016 var ett – i många stycken – tillförlitligt prov.

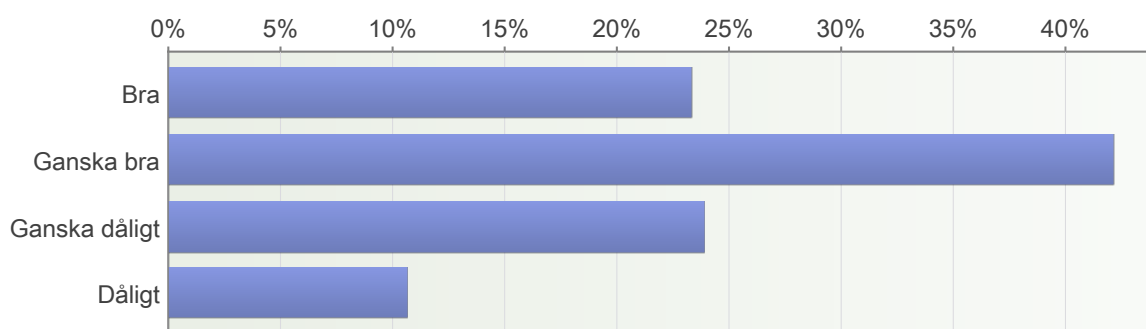
⁶ The Standards for Educational and Psychological Testing 2014

Enkätresultat med kommentarer

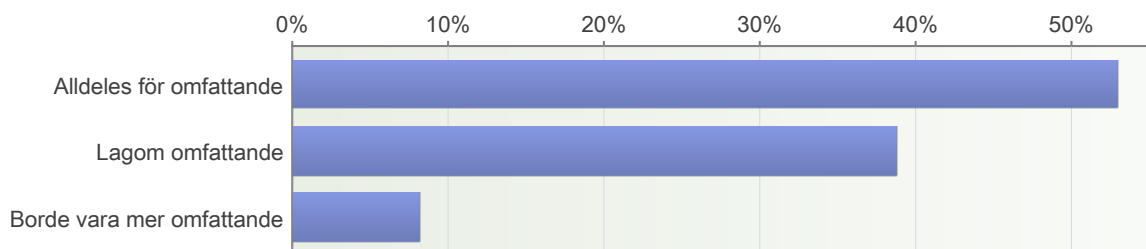
De strukturerade frågorna i enkäten besvarades av nästan 200 lärare och där det fanns möjlighet att skriva kommentarer utnyttjades det av i snitt 130 lärare. Detta är betydligt lägre än föregående år då ca 600 lärare besvarat enkäten. En eventuell förklaring till detta kan vara att en förändring av informationen kring enkäten i Lärarinformationen.

Det nationella provet i samhällskunskap togs 2016 emot mindre väl av lärarna jämfört med tidigare år, i kommentarerna kan man hitta förklaringar till detta – proven låg sent på terminen vilket innebar stor stress för lärarna vid bedömningen och provet upplevdes också som för omfattande. Av de lärare som har svarat på enkäten tycker merparten att provet är bra eller ganska bra, ca 65 %, ca 10 % anser att provet är dåligt.

Vad tyckte du om ämnesprovet i samhällskunskap som helhet?



Vad anser du om omfattningen av ämnesprovet i samhällskunskap?

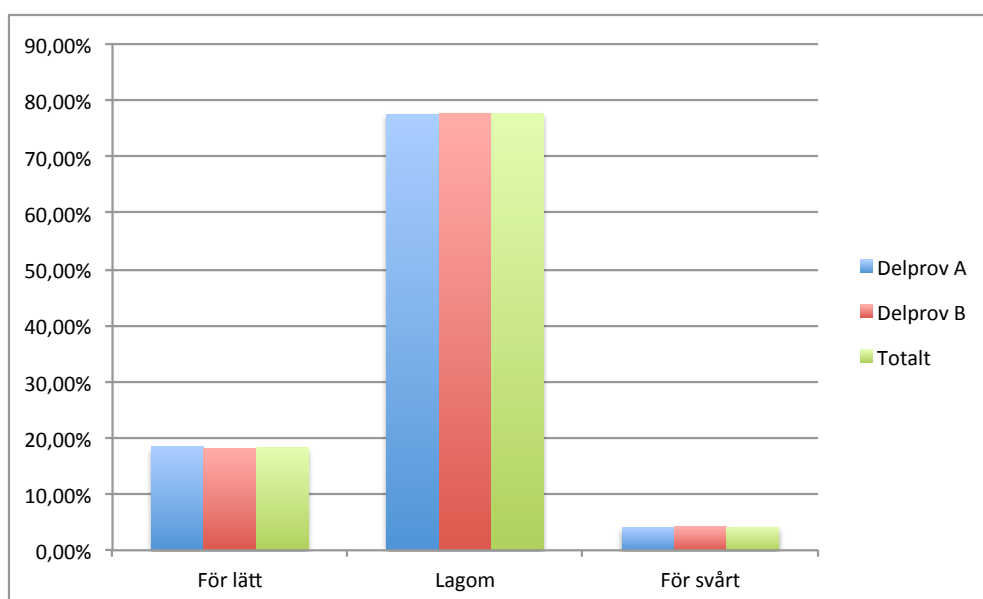


Jämfört med 2015 anser något färre lärare att provet är alldeles för omfattande, andelen lagom omfattande har också minskat. Något förvånande är att ca 6 % fler lärare jfr med 2015 anser att provet borde vara mer omfattande.

I kommentarerna till denna fråga skriver lärarna att provet innehåller för många öppna frågor (detta trots att vi minskat omfattningen av de öppna skrivfrågorna sedan 2015) och att en av de tre delarna av kunskapskraven som provas ges för stort utrymme i provet. Bland de positiva kommentarerna finns beröm för att provet verkligen är samhällskunskap och att uppgifterna och innehållet som eleverna möter är bra.

De allra flesta kommentarerna rör omfattningen både för eleverna att skriva och för lärarna att bedöma så sent på terminen.

Hur bedömer du svårighetsgraden i delproven?

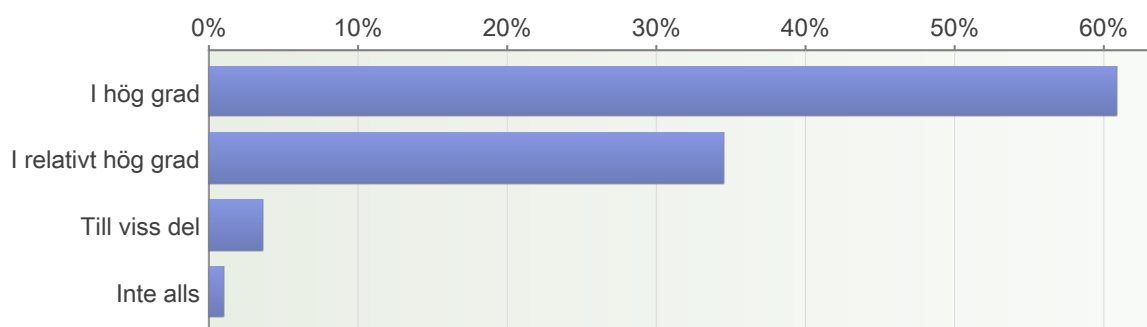


De allra flesta anser att provet är lagom 2016, men fler än tidigare år anser att provet är för lätt, ca 18 % per delprov.

Det finns både positiva och negativa kommentarer kring provets svårighetsgrad, bland de positiva överväger kommentarer kring provets blandning av olika format. Lärarna upplever svårighetsgraden på uppgifterna som bra, men flera menar att bedömningen sedan var för snäll och att det var för lätt att få betyget E på provet och på en del av uppgifterna. Bland kommentarena återfinns också synpunkten att det var för lätt att få E på en del av uppgifterna och alldeles för svårt att nå A på dem.

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. Vi har sedan 2014 ökat de strukturerade uppgifterna och minskat ner de öppna skrivfrågorna. Trots dessa ändringar anser lärarna att balansen mellan öppna och slutna uppgifter inte är lagom, tyvärr anser drygt 60 % av lärarna att balansen inte är bra.

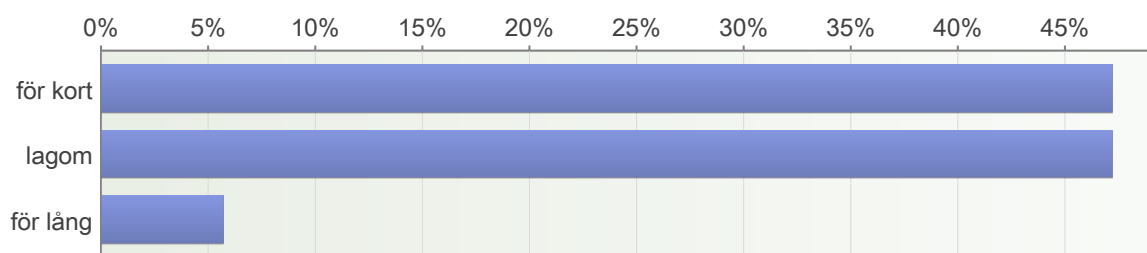
Har provets innehåll behandlats i undervisningen?



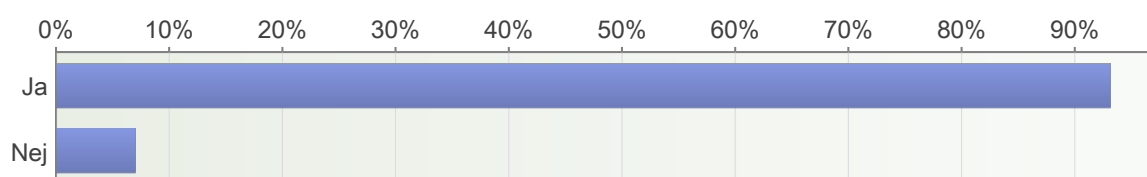
Kommentarerna signalerar att vi har träffat innehållet väl för elevernas del, en del av det är behandlat i årskurs 7 men eleverna och lärarna har känt igen sig.

Provet skrifttid var 120 minuter per delprov, detta upplevde majoriteten av lärarna som för kort och lagom.

Skrivtiden (120 min) var...

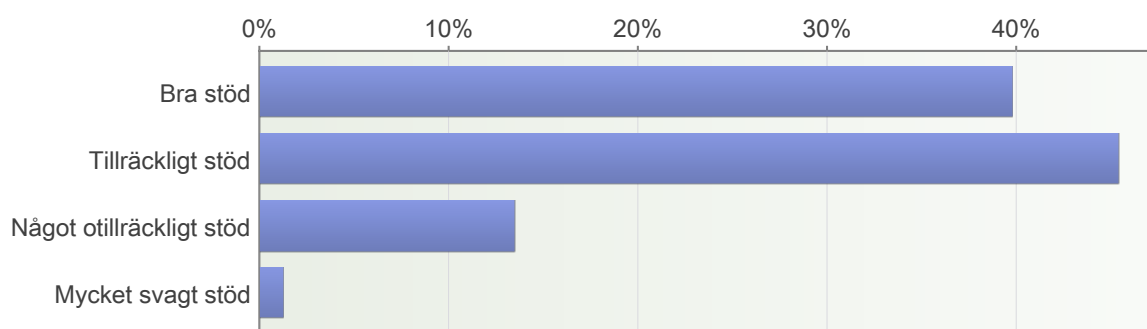


Tankebubblan – upplever du att eleverna hade hjälp av den?



Sedan provet 2014 har vi använt oss av en tankebubbla i de öppna uppgifterna för att signalera till eleverna vad som är viktigt att tänka på när de svarar på uppgifterna, men kan säga att bedömningsaspekterna är omgjorda till elevsprårskurs. Detta har fallit mycket väl ut och drygt 90 % av lärarna anser att den är till hjälp för eleverna. Vi har valt att inte (till eleverna) i varje fråga i detalj ange vad som krävs för de olika nivåerna för att minska risken för att det blir en avbockningslista att skriva efter. Vår tanke att eleverna då ges möjlighet att svara med högre kvalitet. Lärarna är nöjda med bubblan, några menar att texten skulle kunna bli än mer tydlig och att eleverna inte alltid har förstått hur viktig den är.

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



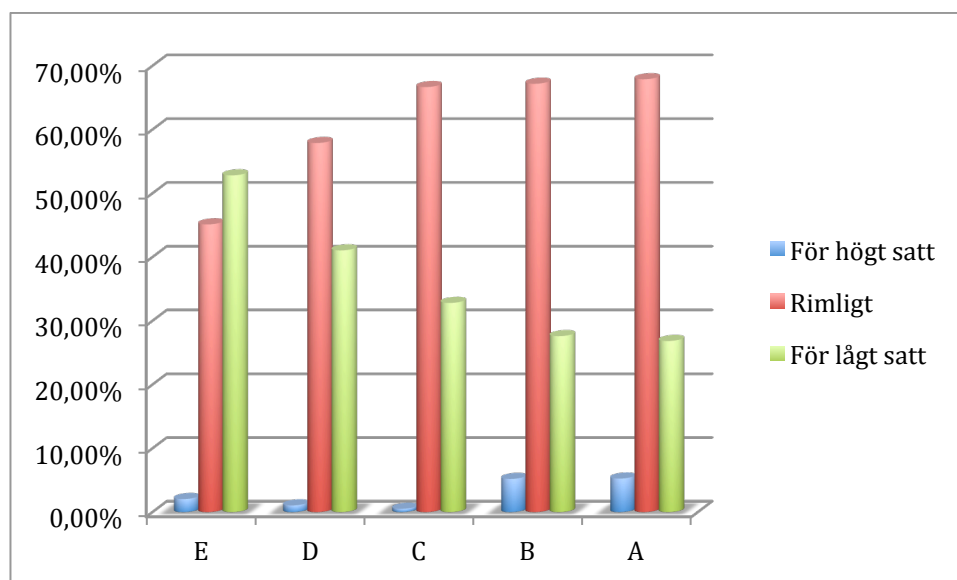
För att ett nationellt prov ska bli reliabelt behövs bedömningsanvisningar som lärarna upplever som stöd. De allra flesta lärarna upplever att bedömningsstödet gett ett bra eller tillräckligt stöd, ca 85 %. Endast några få anser att stödet är mycket svagt.

Bland kommentarerna finns att lärarna gärna hade velat se fler exempel på förtydliganden och konsekvenser för att få mer stöd i den egna bedömningen. Några efterfrågar också svar som inte godkänns samt att en del vill ha fler elevexempel.

När det gäller de autentiska elevsvaren upplever lärarna dessa som bra, men man hade gärna sett fler.

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

De flesta lärare i vår enkät anser att vi har tagit fram rimliga gränser för de olika provbetygen. Nytt för i år är dock att fler lärare än tidigare år anser att gränserna har varit för lågt satta.



Kommentarer:

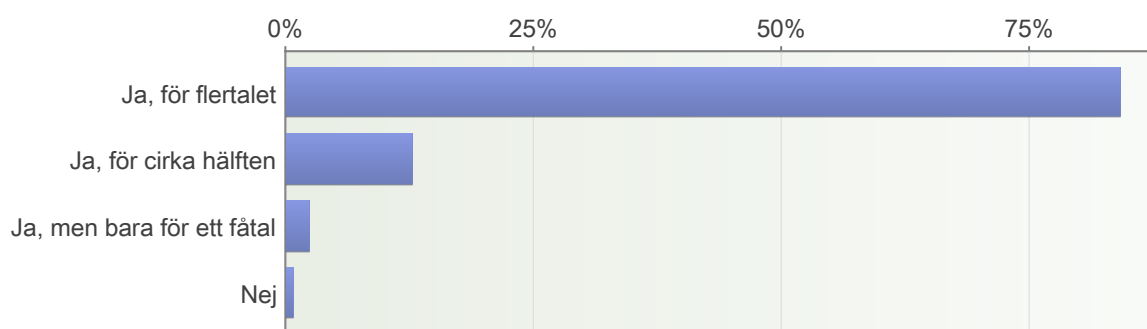
Hade förväntat mig lite högre krav eller så var mina elever väl förberedda och presterade på topp.

A-nivån stämde överens med min uppfattning.

Bedömningen kändes generös men det behövdes kanske när det var så omfattande.

Några lärare skriver att deras elever har kunnat kryssa sig till ett E, men kraven är så satta att det pga breddvillkoret inte är möjligt.

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

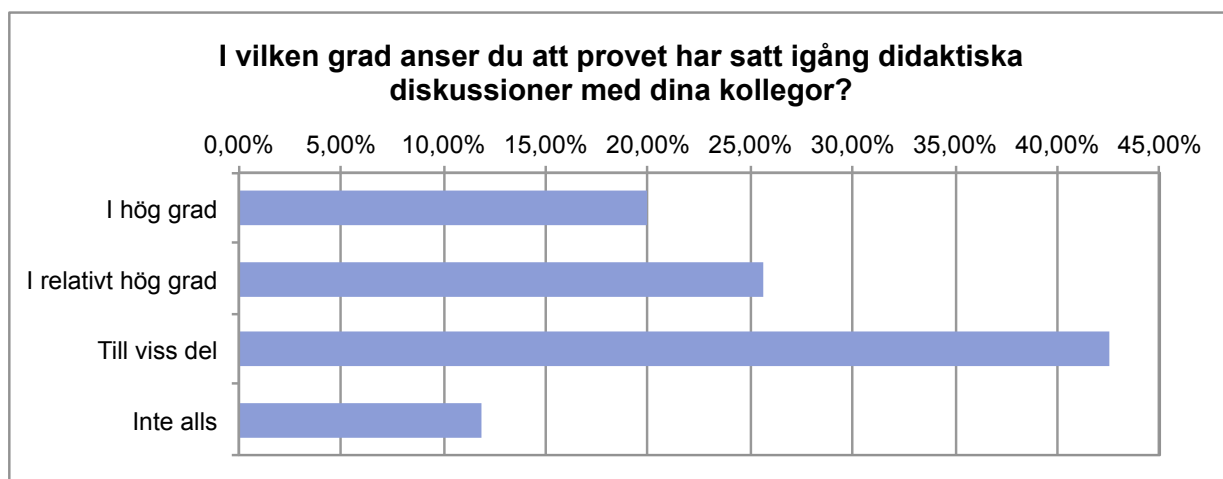
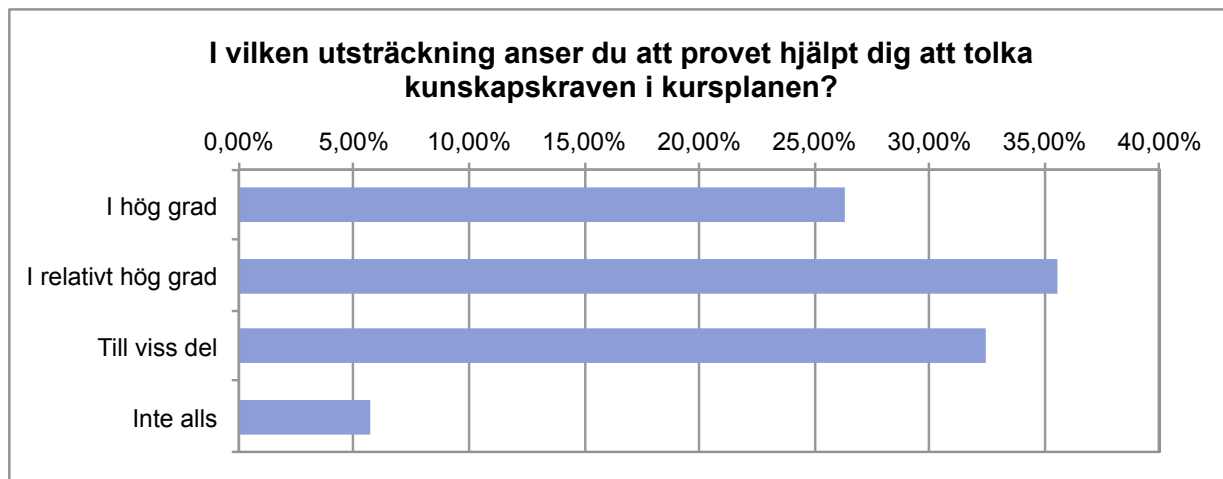


De allra flesta lärare anser att provresultatet ligger i linje med elevernas övriga prestationer, trots synpunkterna ovan. Bland kommentarerna finns sådant som att lärarna kanske har legat lågt i tidigare bedömning men också att de tycker att de har fler elever som når nivå A än vad nationella provet visar.

Det är alltid elever som av olika anledningar presterar bättre eller sämre på de nationella proven än annars. Vi har arbetat mycket med ämnet samhällskunskap under årskurs 8 och 9 och jag har därför ett bra

bedömningsunderlag. Mitt eget betygsunderlag stämmer ganska väl överens med utfallet på det nationella provet. Det diffar lite på enstaka elever.

Nya frågor för 2016 var :



Lärarna tycker att de har fått hjälp att tolka kunskapskraven men det har inte satt igång några större ämnesdidaktiska diskussioner, detta med hänvisning till att det inte getts tillräckligt med tid till detta från skolledningen. Även i denna fråga på enkäten framkommer det missnöje bland lärarna kring att provet är så sent inlagt på terminen.

Avslutning

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 till vida 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 2016 har haft en relativt jämn fördelning mellan uppgifter med öppna och strukturerade svarsalternativ. Under rubriken omfattning har lärare kommenterat att det upplevts som ”för omfattande” i något högre grad. Detta är en försämring mot förra året, som vi måste analysera vidare.

Np Sh9 2016 ger många möjligheter att visa kunskaper för nivå A.

I elevhäftena har den 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 - med god spridning - visa sin kunskapsnivå i ämnet. SEM ligger mycket bra till på hela antalet förekommande belägg inom betygsstegen.

Områden för fortsatt utveckling:

- Fortsatt arbete med att göra bedömningsanvisningarna än tydligare och precisa.
- 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.
- 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 behålla detta inslag i vår utveckling.
- Vi arbetar med en stor repertoar av statistiska mått och jämförelser, och vi avser att behålla och utveckla detta än mer.
- För andra året i rad har vi haft en frivillig inrapportering. Lärarna kunde där jämföra klassens resultat (på grupp nivå) vad avser enskilda delarna av kunskapskraven med rikets. Fler lärare än året innan genomförde detta. Mycket litet information om detta gavs i Lärarinformationen 2016. Vi avser att verka för att denna information återkommer.

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 9 2016

(Förklaring: 619 = inför 2016, I:a prövningen, årskurs 9)

Inför Äp 2016 har följande utprövningar gjorts:

1. 619

Totalt 22 häften konstruerades innehållande 4 uppgifter per häfte.

Årskurs 9

Antal elever: 1231

Antal lärare: 50

Antal skolor: 46

inne 11 nov 2014

2. 629

16 häften konstruerades innehållande 6-7 uppgifter per häfte.

Årskurs 9

Antal elever: 1057

Antal lärare: 34

Antal skolor: 32

inne 31 mars 2015

3. 639

Mindre prövning för några få strukturerade

(11) uppgifter, kortsvarsuppgifter, prövades ytterligare

Årskurs 9

Antal elever: 332

Antal lärare: 6

Antal skolor: 4

inne 15 september 2015

Totalt 2620 elever, 90 lärare på 82 skolor
från Haparanda till Veberöd.