
Ämnesproven i grundskolans årskurs 9 och specialskolans årskurs 10

Provrapport Samhällskunskap

Årskurs 9

Vårterminen 2018



GÖTEBORGS UNIVERSITET
INST FÖR PEDAGOGIK OCH SPECIALPEDAGOGIK

Innehåll

Inledning	3
Konstruktionsprocess för provet	3
Bakgrund	3
Provets sammansättning	4
Bedömningsanvisningar	4
Provresultat med kommentarer	4
Resultat fördelat på	
faktorerna kön och undervisning i SvA	5
Validitet och reliabilitet	8
Validitet	8
Reliabilitet	10
Enkätresultat med kommentarer	13
Avslutning	17
Referenser	18
Bilaga 1: Utprovningar 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 18 och 20 april 2018.

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 till provinstitutionen.

Denna rapport bygger på de inrapporterade provresultaten, de hela elevproven, lärarenkäterna och resultaten från SCB som kom 30 november 2018. Skillnaden mellan SCBs statistik och vårt urval redovisas också. Det är 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 också internationella studier, juridik, kriminologi osv. Det finns inga andra obligatoriska nationella prov i ämnet i Norden.

Vid provkonstruktion talar man ofta om god mätsäkerhet. Det kan låta tekniskt men det är samma sak som att eleven ska ges så goda möjligheter som möjligt att visa sina kunskaper och att provet speglar elevernas faktiska kunskaper. God mätsäkerhet kräver att elever prövas med fler frågor inom varje del av kunskapskraven¹. Om detta inte görs riskerar man att elevens resultat styrs för mycket av slump. Detta innebär också att vår verksamhet att utprova uppgifter blir stor och kvalitetskrävande.

Förutom mindre utprovningar av enstaka uppgifter har fyra större utprovningar genomförts inför proven 2018. Under framtagandet av ämnesprovet har tusentals elever och hundratals lärare medverkar vid dessa provningar (4127 elever, 108 lärare på ca 100 skolor från Sorsele till Karlshamn - 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, exempelvis genom att undersöka hur de olika uppgifterna fungerat i olika elevgrupper. Förutom denna ”hårda” provning så har vi under året genomfört flera varianter av ”mjuka” utvärderingar av uppgifter och prov:

- Varje elev som deltar i utprovningar svara på en enkät om hur uppgifterna upplevts
- Varje lärare som deltar får också besvara en enkät
- Vi har haft fyra större lärarsamlingar med referensgrupper av lärare från hela landet. Cirka hälften av dessa har varit helt nya i sammanhanget.
- I vår provgrupp ingår ett antal verksamma lärare som är med oss en dag i veckan.

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

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

Liksom tidigare år har vi arbetat med så kallade stödstrukturer i provet. Vi har genom åren också noterat att stödstrukturer kan ”stödja” åt fel håll varför vi försökt hitta en lagom nivå. I provet 2018 har vi en s.k. ”hjalpruta” till de öppna uppgifterna. Enligt lärarenkäten (se nedan) har detta fallit väl ut.

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 24 stycken för 2018. 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 nedan².

Np sh åk 9	Antal	Upp till nivå			Antal
	uppg./item	E	C	A	strukturerade
2013	30	30	29	16	19
2014	30	30	30	19	12
2015	28	28	24	18	9
2016	26	26	26	24	10
2017	24	24	24	22	11
2018	24	24	24	24	12

Tabell 1. Fördelning av uppgifter, nivåer och format 2013-2018, 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 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 2018 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). 1451 elever i detta ”samplade”³ urval har inrapporterats digitalt, vilket gjorde att vi tidigt hade preliminära resultat att arbeta med (30 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. Vi är varje år mycket tacksamma för att lärare gör denna inrapportering. När man gör den kan man också få olika sorters feedback på sitt resultat. Se vår egen (SO-ämnenas) portal för information om detta och mycket annat:

www.npsportal.se

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 2018 motsvarade en uppgift ett item.

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

Nedanstående diagram visar som jämförelse provbetyget för vårt urval i juni och SCBs redovisning av provbetyg 30 november. 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 görs.

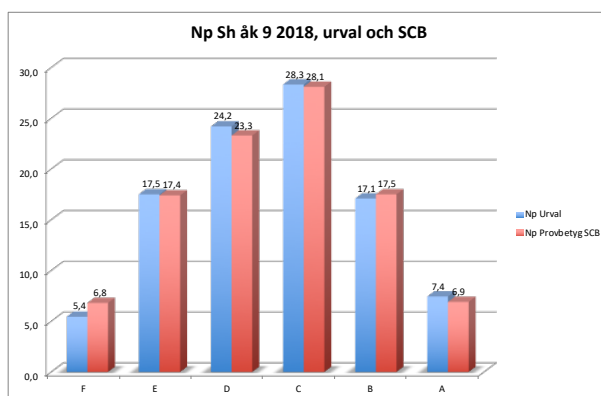


Diagram 1. Jämförelse av provbetyg vårt urval - SCB samhällskunskap i årskurs 9 2018.

I enkäterna tycker de flesta lärarna att provbetygen är ”rimligt” satta. På provbetyg E tycker en stor minoritet att det är för lågt satt (dvs provnivån för E är för ”lätt”) och en minoritet tycker provbetygsgränsen för A är för hög (dvs provnivån för A är för ”svår”).

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

Skillnader mellan pojkar och flickor

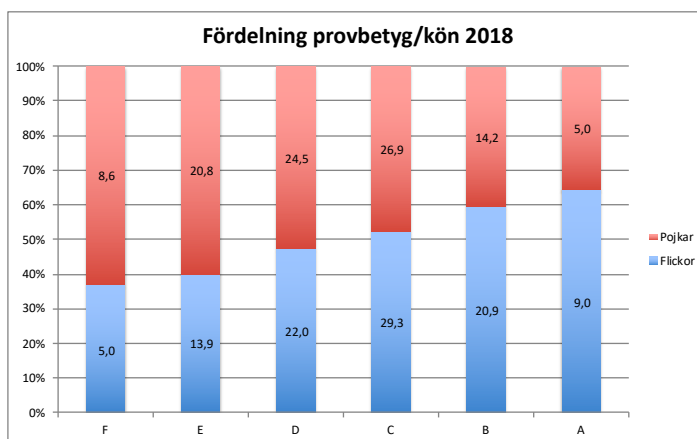


Diagram 2. Fördelning av provbetyg samhällskunskap med avseende på kön/SCB. Årskurs9 2018

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 de föregående årens 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 typ 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.

Antalet uppgifter detta år som är ”neutrala” är betydligt fler. Det är enbart två uppgifter som är inte är neural med avseende på DIF- kön, och en enda vad gäller DIF-SvA. En stor förändringa av resultatet mot 2017. 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 2018 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	sociala medier
2	Gynnar pojkar svagt	Strukturerad	Ekonomi

Tabell 2. DIF-analys avseende kön, Np Sh årskurs 9 2018

Uppgifterna är ”anonymiserade”, dvs. tilldelat uppgiftsnummer är slumpmässigt. Fördelning pojkar/flickor är jämn, dvs en uppgift vardera gynnar. Pojkarna gynnas av strukturerade svarsformat, vilket är känt sedan tidigare. De strukturerade är (till viss del) mer begreppsorienterade. Ingen av uppgifterna under 2018 gynnar någon grupp ”starkt”. Tabellen ovan visar också vikten av att det finns både öppna och strukturerade svarsalternativ

Skillnader mellan de som läser och inte läser svenska som andraspråkskurs (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. 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 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	Innehåll
3	Missgynnar SvA, svagt	strukturerad	Demokrati

Tabell 3. DIF-analys avseende att följa kursplanen inom SvA, Np Sh årskurs 9 2018

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.

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. Så inte i år. Men det kan också bero på att uppgiften testar ett speciellt område inom domänen "Demokrati".

Ämnesprovet i samhällskunskap 2018 hade således mycket liten skillnad i DIF-analys vad gäller dessa undergrupper.

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 latin 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 under året till gruppen har knutit fem 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 eller vartannat år, för att vi skall få ny ”input” från verksamheten. Detta förutom alla de större möten med referensgrupplärare som vi har under året.
3. För att få än större inflytande från verksamheten, så har vi under året genomfört 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 4127 elever och 108 lärare från hela Sverige. Lärarna slumpas delvis och tas delvis från de som via enkäten till proven anmält intresse. Genom analys av enkäterna får vi en bra uppfattning om hur frågorna tas emot.

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*”.

b) Under året har vi ett antal större lärarsamlingar. En av dessa, kallad ”bedömarveckan” (BV, numera tre dagar, var tidigare en hel vecka), skedde 3-5 april 2017 inför detta prov. 14 lärare från 14 olika skolor i 12 kommuner medverkade. Till BV-dagarna brukar proportionen ”gamla/nya” vara 50/50, d.v.s. hälften av lärarna har varit med någon gång förut för att leda arbetsprocessen i grupperna, hälften skall vara nya, dock har alla genomfört någon prövning.

Lärarna bedömer de öppna uppgifter som prövats, bedömningen sker med hjälp av bedömningsanvisningar som tagits fram i provgruppen. Efter bedömningen bearbetas både uppgift och bedömningsanvisning av lärarna där de kommer med förslag till förändringar och förbättringar. Under kalenderåret 2017, med arbetet för provet 2018, har dessa ”BV” ökat till två, och ligger också tidigare, då vi arbetat med att tidigarelägga prövningarna i förhållande till när provet genomförs i skolorna.

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 (2018) den 4-5 oktober 2017, Gruppen bestod då av 11 lärare, 5 av dessa hade jobbat med oss tidigare (en av dessa för åk 6 men nu med åk 9) och 6 var helt nya i sammanhanget. De helt nya lärarna fick vi kontakt med via lärarenkäten 2017, de hade där anmält sitt intresse till att delta på olika sätt i vårt arbete.

d) Den sista valideringen av provet sker vid det s.k. kravgränssättningsmötet då lärare från hela landet hjälper oss med att ta fram kravgränserna för de olika provbetygen. De får också kommentera främst bedömningsanvisningarna. Mötet skedde 17-18 januari 2018, 15 lärare från 15 kommuner, från Sorsele till Karlshamn medverkade. Här ska det vara fler ”nya”, detta år hade vi 10 ”nya”, 5 ”gamla”.

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 och Institutionen för nationalekonomi, alla vid Göteborgs universitet. De är ämnesföreträdare, men har alla vid något tillfälle i sina karriärer undervisat på ämneslärarutbildningarna. Vi använder oss också av en advokat och en jurist som granskar uppgifter med det specifika ämnesinnehållet.

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 urvalseleverna. Denna tekniska rapport publiceras inte, då den också är uppgiftsspecifik avseende sekretessbelagda uppgifter. Rapporten är relativt omfattande, 45 sidor, och tas fram i augusti för provet som gick i maj.

Arbetet med IRT har hjälpt oss med arbetet att skriva säkrare och mer stabila bedömningsanvisningar, i syfte 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 valda 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 2018 har ett reliabilitetsmått på 0,943, vilket är det högsta värde provet haft. (2017: 0,933, 2016: 0,941, 2015: 0,932, 2014: 0,925, 2013: 0,911). Reliabiliteten är beroende av antalet uppgifter. I år hade 24 uppgifter i förhållande till 2016 då vi hade 26. 2013 hade vi 30 uppgifter, så arbetet med att göra varje uppgift mer reliabel har också fungerat. Vi har detta år också analyserat reliabiliteten för de enskilda delarna av kunskapskraven som provet prövar. Vi har för tredje gången 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. Lärare till 606 elever har 2018 utnyttjat denna möjlighet.

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 2018:

NpSh9 2018			
På hela provet			0,9428
Per dkk/förmåga			
1	<i>Samband-str</i>	12 item	0,8864
3	<i>Samhällsfråg</i>	6 item	0,8303
5	<i>Demokrati et</i>	6 item	0,8246

Tabell 4. Cronbach alpha Np Sh årskurs 9 2018

I tabellen syns att värdena väl motsvarar gränsvärdena för individuell återrapportering för hela provet, och kraven för återrapportering av delarna av kunskapskraven på grupp-nivå. Item är beteckningen på ett ”bedömningsområde”, oftast är en uppgift ett item. För samhällskunskap 9 år 2018 motsvarar alltid en uppgift ett item. Men en uppgift kan bedömas utifrån två olika delar av kunskapskraven, och därmed ge två item (en metod provet i samhällskunskap inte använder). Eller så kan flera uppgifter, t.ex. ett antal kortsvar, bedömas ”i klump”, dvs. bilda 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 är mycket bra.

NpSh9 2018	
Separation index:	4,037

Tabell 5. Separation index Np Sh årskurs 9 2018

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 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.

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 att 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 Educational 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 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 ska ha det, och inte av slump någon annan gång skulle få några poäng färre eller fler.

NpSh9 18 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:

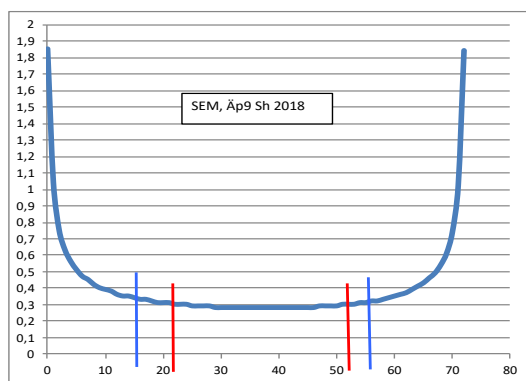


Diagram 4. SEM, NpSh 9, 2018

⁶ The Standards for Educational and Psychological Testing 2014

SEM sjunker under 0,3 vid 22 belägg och stiger över vid 53 belägg, och har 0,33 vid 58 belägg. Lägsta nivå är 0,28 (mellan 29 – 46 belägg). Detta ett bra resultat. Provet har en 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 provet är tillförlitligt både vad gäller validitet och mätsäkerhet, så blir provet 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. Vi behöver således data för att göra tillförlitliga prov.

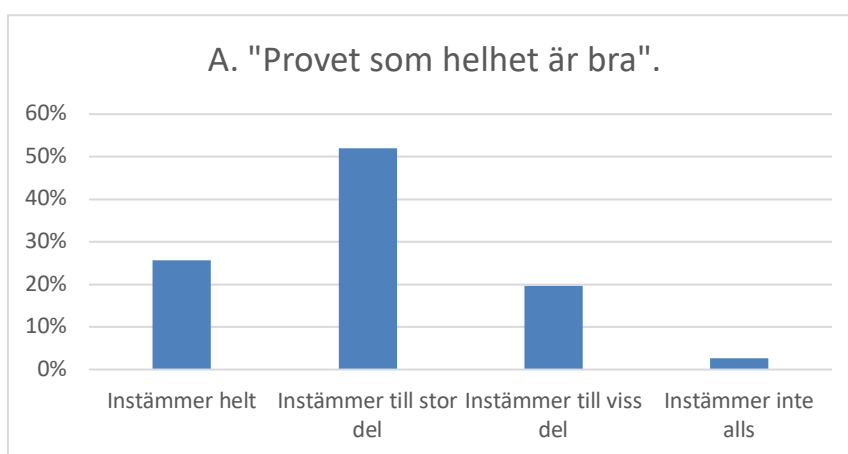
Nationella provet i samhällskunskap årskurs 9 2018 var ett – i många stycken – tillförlitligt prov.

Enkätresultat med kommentarer

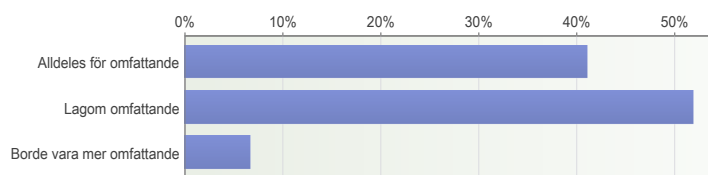
De strukturerade frågorna i enkäten besvarades av 149 lärare och i de frågor det fanns möjlighet att skriva kommentarer utnyttjades det av i snitt 50 lärare. Detta är lägre än föregående år, och betydligt lägre än för tre år sedan då 396 lärare svarade på enkäten. En eventuell förklaring till detta kan vara att en förändring av informationen kring enkäten i Lärarinformationen. 2017 har också enkäten omarbetats så flera frågor är gemensamma oavsett ämne.

Det nationella provet i samhällskunskap togs 2018 emot väl. Av de lärare som har svarat på enkäten har 78% svarat att de instämmer helt eller instämmer till stor del med påståendet ”provet är som helhet bra”. Detta är en ny fråga för lärarna att besvara, men vi kan jämföra denna fråga med ”Vad tyckte du om ämnesprovet som helhet?” och där har vi sedan starten 2013 kunnat avläsa att lärarna har tyckt bra eller mycket bra om provet i samhällskunskap.

Vad tyckte du om ämnesprovet i samhällskunskap som helhet?



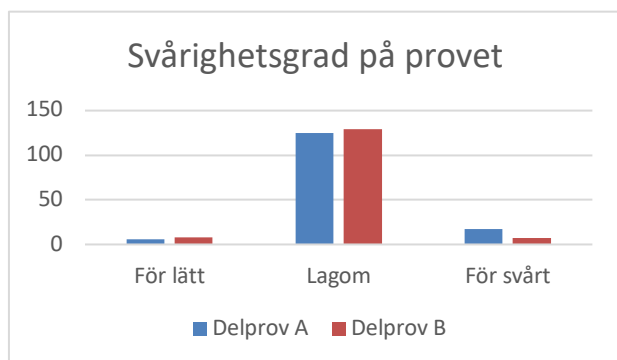
Vad anser du om omfattningen av ämnesprovet i samhällskunskap?



Jämfört med 2017 anser något fler lärare att provet är alldeles för omfattande, 41% 2018 och 37% 2017, andelen som uppfattar provet som lagom omfattande är i princip samma. Några fler, jämfört med 2017 anser att provet borde vara mer omfattande.

I frisvarsdelen till denna fråga finns många blandade kommentarer men det som nämns mest är att provet innehåller för mycket begrepp (7 kommentarer av 46) samt att vi har frågor som lärarna upplever liknar varandra (5 kommentarer). Vi har endast tre kommentarer på att det är för många öppna uppgifter för eleverna. Detta har sedan 2013 varit en återkommande synpunkt från flera lärare. Bland de positiva kommentarerna finns exempelvis att uppgifterna och innehållet som eleverna möter känns igen både av lärare och elever.

Hur bedömer du svårighetsgraden i delproven?

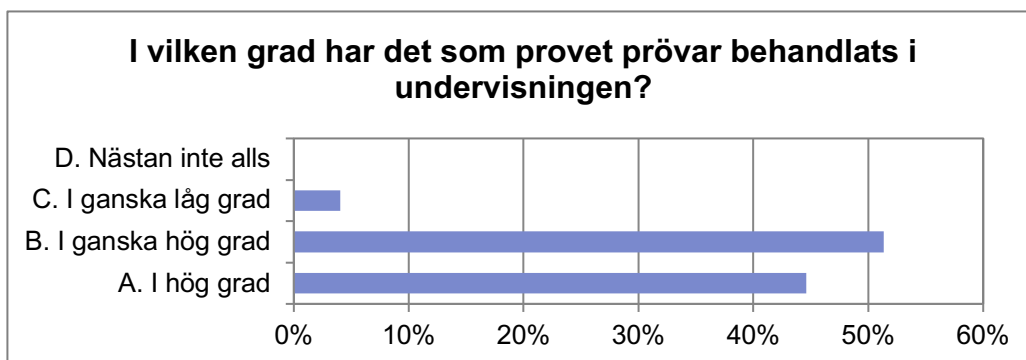


De allra flesta anser att provets svårighetsgrad är lagom.

Det finns både positiva och negativa kommentarer kring provets svårighetsgrad, bland de positiva överväger kommentarer kring att provets svårighetsgrad är just lagom och bland de negativa att det kändes lätt att få belägg för nivå E på en del av uppgifterna och svårt att få provbetyget A. Lärarna är också nöjda med att provet i år innehåller fler slutna uppgifter än tidigare som går snabbare att bedöma. Samtidigt finns det en viss tveksamhet till att de slutna uppgifterna verkligen kan vara lika väl fungerande som de öppna.

När vi utformar provet är det viktigt att det innehåller uppgifter av varierande format, alltså både uppgifter som kräver en del skrivinsats och uppgifter som är slutna och inte kräver långa svar. 2018 svarade 72 % av lärarna att de är nöjda med provets balans mellan öppna och slutna svarsformat. Detta är samma siffror som för 2017.

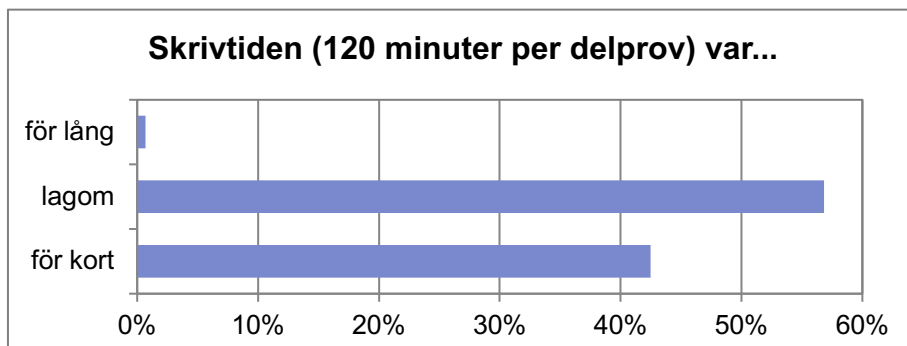
Har provets innehåll behandlats i undervisningen?



Detta är en mycket viktig fråga för oss, vi vill vara säkra på att det vi ställer frågor om också har behandlats i undervisningen. En del av kommentarerna rör att innehållet behandlats i årskurs 7 och så har det varit sedan proven började genomföras 2013. Men inget verkar vara helt obehandlat. Vi är mycket glada över dessa siffror då detta är något vi lägger mycket tid och kraft på i utformandet av våra frågor. Om eleverna känner igen innehållet blir de också mer positivt inställda till att genomföra provet.

Provet skrivtid var 120 minuter per delprov, detta upplevde majoriteten av lärarna som lagom. Andelen som upplever tiden för kort har tyvärr ökat sedan 2017 och detta måste vi se över till 2018, provet innehöll 2018 fler slutna uppgifter vilket tidigare indikerat till oss att det kräver kortare tid men i år har något förändrats avseende detta.

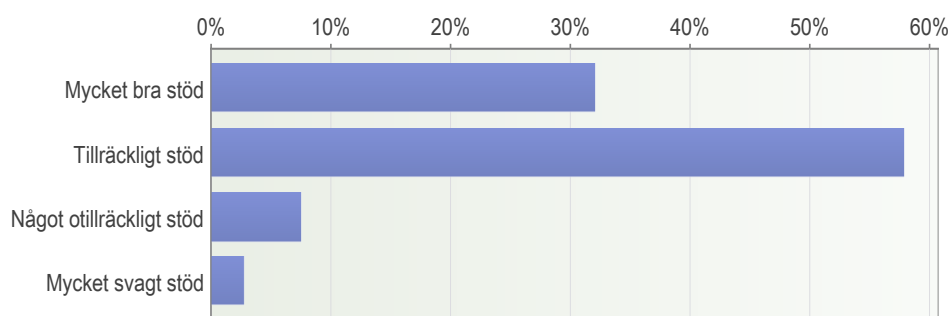
Skrivtiden (120 min) var...



Hjälprutan – upplever du att eleverna hade hjälp av den?

Sedan provet 2014 har vi använt oss av en hjälpruta (tidigare tankebubbla) i de öppna uppgifterna för att signalera till eleverna vad som är viktigt att tänka på när de svarar på uppgifterna, man kan säga att bedömningsaspekterna är omgjorda till elevspråk. Detta har fallit mycket väl ut och drygt 80% 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. Nytt är att vi har signalerat antal i rutan, det kan exempelvis stå ”försök att ange 2–4 orsaker”. Eleven har då i vårt informationsblad fått reda på att den lägre siffran anger kravet för belägg på nivå E och den högre är kravet för belägg för nivå A.

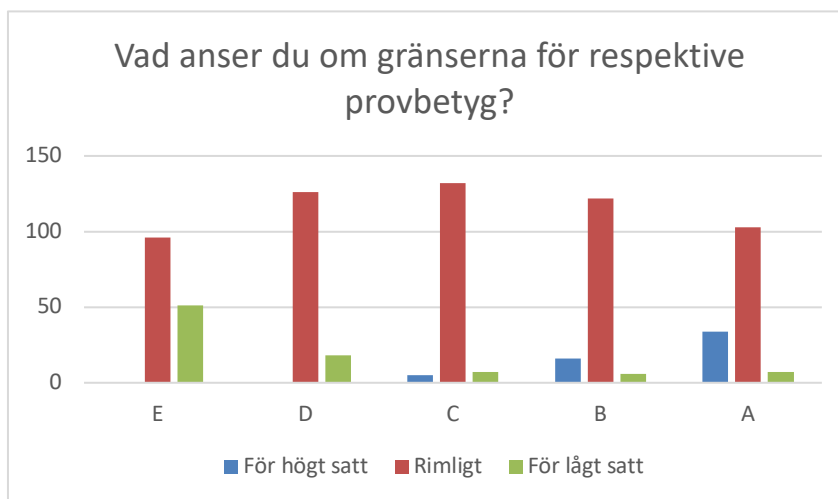
Vad anser du om utformningen av häftet bedömningsanvisningar med avseende på vilket stöd det gav vid bedömning?



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, nästan 90 %. Endast några få anser att stödet är mycket svagt.

Bland kommentarerna finns att lärarna gärna hade velat se fler exempel på elevsvar, utförligare exempellistor och svar som inte når nivå E efterfrågas också. Vi får också positiv respons där vi bland annat kan läsa att lärarna upplever bedömningsanvisningarna som mycket bättre än tidigare år. Samt att vår layout är enkel att greppa och det är tydligt hur bedömningen ska göras.

Vad anser du om gränserna för respektive provbetyg?

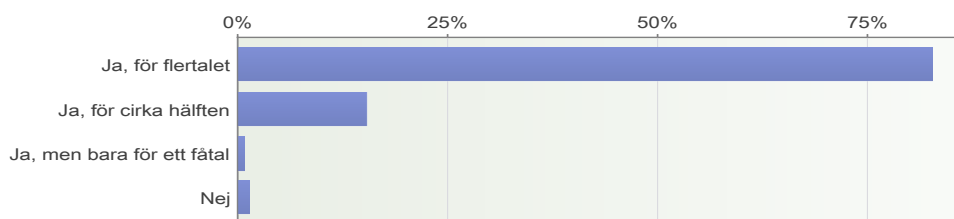


Liksom tidigare år tycker en del lärare att det är för enkelt att få E och för svårt att få A. Vi har också i år ett antal kommentarer om det borttagna breddvillkoret, där en del lärare ställer sig frågande till detta beslut.

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.

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 bland annat att lärarna känner sig stärkta i hur de tidigare har bedömt sina elever samt att överensstämmelsen är mycket god.

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.

Np Sh9 2018 har haft en jämn fördelning mellan uppgifter med öppna och strukturerade svarsalternativ. Detta återspeglas till viss del i enkäten. Svårighetsgraden på uppgifterna upplevs som ”lagom” av nästan alla svarande. Vi har inte heller någon uppgift som ”sticker ut” i den mening att det är många elever som har haft svårt att förstå den.

Np Sh9 2018 ger många möjligheter att visa kunskaper för nivå A. I elevhäftena har den stödstrukturen ”tankebubblan” fortsatt mottagits bra. Bedömningsanvisningarna och den av oss valda layouten verkar också fortsatt 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 och bibehållande av goda värden:

- 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. Detta är en pågående process.
- Vi har redan idag ett stort utbyte med lärarna ute i verksamheten. Vi avser att behålla, och än mer utöka, detta inslag i utveckling av proven. För provet i samhällskunskap har denna kontakt varit helt avgörande, inte minst för att göra innehåll och utformning av uppgifterna så tydligt som möjligt för eleverna. Genom vårt pågående arbete att lägga utprövningar tidigare så ges nu möjlighet att på än bättre sätt pröva tidiga varianter av uppgifter tillsammans med lärare och elever ute i skolorna.
- Vi arbetar med en stor repertoar av statistiska mått och jämförelser, och vi avser att behålla och utveckla detta än mer. Programvaror för detta utvecklas ständigt.
- För tredje året i rad har vi haft en frivillig inrapportering. Lärarna kunde där jämföra klassens resultat (på gruppnivå) vad avser enskilda delarna av kunskapskraven med rikets. Fler lärare än året innan genomförde detta. Mycket litet information om detta gavs i Lärarinformationen 2018. Vi avser att verka för att denna information återkommer. Utvecklandet av So-ämnenas provkonstruktörers egen hemsida, npsportal.se, skall fortsätta för kontinuerlig information till lärarkåren.
- Vi kommer att arbeta för att informationen skall bli tydligare om betydelsen av lärarenkäten för utvecklandet av proven. Troligen har den förändringen av informationen inneburit en sänkning av antalet svar.

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 2018

(Förklaring: 819 = 8 inför 2018, 1:a prövningen, årskurs 9) Inför Äp 2018 har följande utprövningar gjorts:

1. 819

Årskurs 9

Antal elever: 1099

Antal lärare: 30

2. 829

Årskurs 9

Antal elever: 1338

Antal lärare: 38

3. 839

Årskurs 9

Antal elever: 1136

Antal lärare: 28

4. 849

Årskurs 9 samt gymnasiet åk 1 (endast strukturerade uppgifter)

Antal elever: 554

Antal lärare: 12

Totalt 4127 elever, 108 lärare på ca 100 skolor från Sorsele till Malmö.