

Kap 2 Resultat

2.1 Kvalitativa resultat

Här redovisas projektets arbete med det innehållsliga området Etik.

På samma sätt som inom Bedömningsstödet i samhällskunskap "Se den andre"¹ (PBB/Sh), så är proven av relativt öppen karaktär. Anledningarna är två:

1. Eftersom de nya kunskapskraven i princip alltid utgår ifrån den kvalitativa bestämningen "att kunna resonera" inom respektive område, så måste vi ge eleverna en chans att göra just detta.
2. Många lektioner inom ämnet religionskunskap har karaktären av öppna frågor och diskussioner. Eleverna tränar på detta och finner ofta just denna del meningsfull. Lärare upplever dock inte sällan att det är svårt att göra bedömningstillfällen, prov, till detta innehåll. Projektet vill stimulera till och visa på att det går att genomföra prov av denna typ, med vetenskaplig rimlighet. För ett heltäckande prov för att bedöma ämnets alla förmågor måste andra format också användas. Men reliabla stöd för öppna, resonerande frågeformat behövs alltid inom ämnen av religionskunskaps karaktär.

På samma sätt som inom PBB/Sh så menar deltagande lärare att eleverna skriver mycket på proven. Eleverna verkar anse att frågeställningarna är relevanta och intressanta.

Resultaten angående de olika proven har dock utfallit olika. I detta inledande kapitel 2.1 ser vi på de mer allmänna resultaten. Resultat av statistisk typ redovisas under kapitel 2:2, Kvantitativa resultat.

Proven inom etikområdet har utfallit **bra** såtillvida att lärarna menar att det är lätt att bedöma och att åtskilja i olika provbetygsnivåer och att eleverna inte kommenterat att de skulle vara svåra att förstå - snarare tvärtom.

2.1.1 Etikproven

Efter projektet så menar vi att proven Etik 1 och Etik2 tillsammans med bedömningsstödet i Etik fungerar väl att arbeta med inom ramen för de nya kunskapskraven och den nya betygsskalan.

Här beskrivs i vilka bilagor resterande information finns för etikområdet. ([I databasen är alla klickbara och finns i menyn till vänster eller ovanför](#))

¹ Skolverket 2009 "Prov- och bedömningsbank i samhällskunskap, Se den andre" (Dnr 2009:00222).
Nås även över Skolverkets hemsida: <http://www.skolverket.se/2.3684/2.1191/2.1195/2.2975> eller direkt till hemsidan: www.shbedomning.se

Kapitel 7 ytterligare information om Etik 1 och Etik 2

I kapitel 7.1.3 finns förslag på upplägg av undervisning, som lärarna i projektet gjorde.

I bilagorna finns

1.1 Prov Etik 1

1.2 Prov Etik 2

2.1 Bedömningsstödet område Etik

3.2 – 3.3 12 st. elevsvar till prov Etik 1 och 2. Två stycken per betygsnivå A – F.

OBS! De ligger i slumpmässig ordning, så man kan ta bedömningsstödet, bedöma själv, och sedan jämföra med förslag till bedömningar som ligger i bilaga 4.1 – 4.4. Bilaga 3.6 "Facit" här ligger de provbetyg som elevsvaren fått i omdömnings-testet i projektet, med bearbetade bedömningskommentarer och ett exemplar på bedömningsprotokoll.

2.2 Kvantitativa resultat

2.2.1 Inledning

[Eleverna skriver mycket på dessa prov](#)

[Tabell 1 Antal skrivna sidor i medeltal per prov](#)

Prov	Medeltal sidor	Antal elever
E1	2,7	252
E2	2,9	287
		539

I kapitel 1 redogörs för omdömnings genomförande. Eftersom alla kapitel skall kunna läsas fristående inom databasen så görs en kort sammanfattning här:

Från Kapitel 1, sid. 10:

När stöden fanns i sin nya form, så valdes 400 (100 elevsvar per prov) fram till testet. Efter att ha tagit hänsyn till vissa inledande kriterier, valdes elevsvaren ut slumpmässigt. De inledande kriterierna var en någorlunda jämn fördelning mellan flickor och pojkar och fler elevsvar på högre betygsnivåer så att vi som underlag hade fler att välja mellan. En del F-svar var i princip helt tomma prov, dessa valdes bort.

De återstående 400 svaren fördelades så att de fem lärare som själva arbetat med undervisning i etik, genomfört provet Etik 1, de omdömde elevsvaren till Etik 2-provet osv. Man jobbade således inom samma område som man själv gjort proven i P02, men med ett annat prov.

De utvalda 400 (av de 935) anonymiserades och försågs med en ny elevkod, för omdömnings. Varje elevsvar kopierades i tre exemplar, totalt 1200 kopior. Eftersom snittet av svaren var på 3,1 sidor (se resultatkapitlet) så utgjorde denna process ett

relativt omfattande arbete. Elevsvaren lottades slumpmässigt ut till lärarna inom respektive grupp. Varje lärare omdömde 40 elevsvar. De provansvariga omdömde 100 elevsvar. Provansvarig för Etik 2 bedömde 100 elevsvar från Etik 1 osv. Till varje elevsvar så skrev lärarna ett provbetyg, lämnade in ett bedömningsprotokoll, och skrev en kommentar om c:a 5 till 15 rader. Dessa kommentarer skrevs digitalt. De provansvariga skrev en kortare kommentar. Till varje elevsvar finns således 3 digitala kommentarer, förutom den som fanns till PO2. Alla dessa 1200 kommentarer fogades samman till var sitt 100-sidigt dokument per prov och distribuerades tillbaks till lärarna. På detta sätt fick de tre helt nya kommentarer till de bedömningar de själva gjort av sina elever under hösten. Varje lärare hade c:a 10-30 elever med i omdömningen. Kommentarer till att på detta sätt kunna "kalibrera" sin egen bedömning har varit mycket positiva.

En önskvärd information, svarens längd, missades. Två lärarstuderande har i efterhand gått igenom alla 935 elevsvar omräknat till rader per elevsvar. Eftersom alla elever skrev på en standardiserad provsida så kan det göras relativt goda jämförelser mellan svaren, trots skillnader i handstil.

Detta material om 1200 bedömningar på de utvalda 400 originalsvaren har sedan bearbetats statistiskt. Detta redovisas i resultatkapitlet.

Som underlag för de publicerade exemplen samlades alla lärare efter omdömningarna till en heldag där de utifrån de egna nya förslagen på "Intressanta elevsvar" tog fram exempel på sådana i varje grupp.

Proven skiljer sig åt, men framförallt skiljer sig kunskapsområdena. Man bör ha i minnet att alla dessa prov innehåller mycket öppna essäfrågor, typexemplet är väl Etik 2 där provfrågan lyder: "Vad är en god människa?". Ett av bedömningsbankens mål är att försöka visa att det går – med bibehållen reliabilitet – att bedöma öppna frågor av den typ som många lärare arbetar med under lektioner, men kanske är lite avvaktande till att använda vid ett provtillfälle.

Till varje prov har således 100 elevsvar valts ut till omdömning. Varje svar har bedömts tre gånger: Omdömare 1 och 2 genom ett "nästlat" förfarande och sedan har en provansvarig (PA) bedömt alla 100.

2.2.2. Resultat

2.2.2.1 Jämförelse nationella prov

Materialet vid omdömningen kan analyseras på flera sätt. En vanlig metod för att mäta överensstämmelse mellan större mängder elevsvar är att använda Pearsons produktmomentkorrelationskoefficient, "r". Samband 1 innebär fullständig matchning, ingen avvikelse alls i materialet och 0 ingen matchning. Eftersom vi inte kunde använda den ursprungliga bedömningen i PO2 så gjordes omdömningen som beskrivs ovan. Lärarna delades slumpmässigt in som bedömare 1 respektive 2. Erfarna och mindre erfarna bedömare blandades.

För att få perspektiv på resultaten så kan vi se på den studie Skolverket har låtit genomföra av bedömaröverensstämmelse av de nationella proven². Deras upplägg var något annorlunda. De har slumpmässigt dragit 100 elevsvar från de nationella proven och sedan låtit två eller tre "professionella" bedömare läsa och bedöma dessa 100 svar.

Som jämförelse ser vi på bedömaröverensstämmelsen mellan den skriftliga uppgiften i engelska, skrivuppgiften i svenska och problemlösningssuppgiften i matematik. Resultaten fördelar sig på följande sätt:

Tabell 2. Bedömarsamstämmighet i nationella prov

	<i>svenska skriv</i>	<i>engelska skriv</i>	<i>matematik problemlösning</i>
lägsta	0,34	0,86	0,72
högsta	0,47	0,93	0,95

De nationella proven har verkat under en lång följd av år, medan detta projekts bedömningsstöd är testat för första gången. Efter undersökningen så har Skolverket satt in åtgärder för att öka reliabiliteten vid proven i svenska.

2.2.2.2 Etikproven

Etikprov nummer 1 har relativt god samstämmighet, Etik 2 har något mindre.

Tabell 3. Bedömarsamstämmighet inom projektets etikprov

	<i>Etik 1</i>	<i>Etik 2</i>
lägsta	0,75	0,68
högsta	0,82	0,78

Projektets prov i religionskunskap har haft mycket öppna frågor. Bedömarsamstämmigheten är rimlig, något bättre än resultaten från projektet i samhällskunskap. Spridningen är liten, vilket i detta sammanhang är bra.

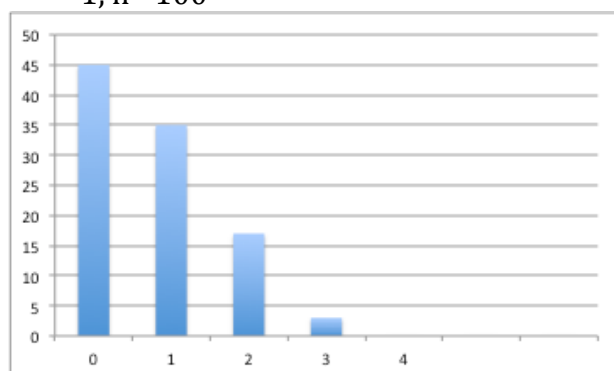
En synpunkt är att koefficienten "r" tenderar att bli högre när man jämför material med fler steg. Ur denna aspekt så skulle det vara svårare att i vårt material få högre "r" än i ett material med tio betygssteg. Detta är dock en randanmärkning, då så många andra faktorer givetvis påverkar mer.³

² Skolverket *Bedömaröverensstämmelse vis bedömning av nationella prov* Dnr 2008:00286 (2008). Observera texten sid 23, att upplägget av testen inte är samma i projektet, då projektet använder sig av nästlad undersökning.

³ Ference Marton visar detta i sin avhandling 1970 "Structural dynamics of learning (Stockholm 1970)

Ett annat sätt att se på överensstämmelse är att se på avvikelse i betygssteg. Vid Skolverkets studie så användes för engelska och svenska den så kallade 10-gradiga betygsskalan, dvs. från EUM (ej uppnått målen) till MVG med alla + och - på varje betygssteg (enligt Lpo94) inräknat. I vår studie användes den nya betygsskalan F - A utan +/- dvs. sju steg.

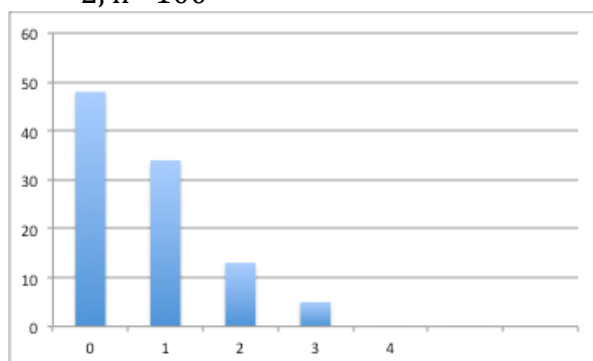
Diagram 1. Skillnad i betygssteg mellan bedömarna i ombedömningen av Etik 1, n= 100



I 80 % av elevsvaren så var skillnaden ingen eller som mest ett betygssteg mellan tre oberoende bedömare.

På Etik 2 så får man följande diagram:

Diagram 2. Skillnad i betygssteg mellan bedömarna i ombedömningen av Etik 2, n= 100



i 82 % av elevsvaren så var skillnaden ingen eller som mest ett betygssteg mellan tre oberoende bedömare.

Det bör återigen påpekas att provfrågorna är mycket öppna. Att så hög procentsats av fristående bedömningar ligger inom noll eller ett provbetygsstegs skillnad bör vara en god grund för fortsatt utveckling att arbeta med bedömningsstöd av beläggskaraktär.

2.2.2.3 Elevsvarens längd

Vi har också räknat antal rader på alla de avgivna 935 elevsvaren ([här ingår elevsvaren från den här ej redovisade livsfrågadelen](#)). Många av de deltagande lärarna vittnar om

att eleverna skriver mycket på alla våra *fyra* prov. Eftersom alla elevsvar är skrivna på en standardiserad sida (med 24 rader) så kan vi jämföra svaren. Nu skriver eleverna givetvis med lite olika stil och storlek, men med så här stort antal svar kan vi ändå göra vissa jämförelser.

Några resultat avseende elevernas skrivmängd

Vi har 935 elevsvar från PO2 på de fyra proven. Totalt har eleverna skrivit 2814 sidor omkring religiösa resonemang.

Flickor och pojkar.

Alla deltagande lärare fick innan PO2 i nov -10 skriva in bakgrundsdata för alla elever; kön, betyg i Sv/Sv2 och So/Re. Vi har "missing data" vad gäller kön i 31 fall. Av vår undersökta mängd ser det ut så här:

Tabell 5. Deltagande elevsvar i undersökningen av skrivmängd:

Pojkar	471
Flickor	433
	904

Den sedan flera år konstaterade skillnaden i grundskolan mellan flickor och pojkarnas skrivande syns också här:

Tabell 6. Skillnad pojkar/flickor i skrivmängd:

	Antal sidor	Snitt	%-uell skillnad	Kortaste	Längsta
Alla	2 814	3,1			
Pojkar	1 212	2,6		0	10,5
Flickor	1 511	3,5	+36%	0	12,8

Det längsta svaret är ett svar som bedöms ligga på provbetygsnivå E.

Våra deltagande elever skriver således något mer än 3 sidor i snitt till projektet.

Problemet att bedöma ämnesmässiga resonemang istället för språkmässiga

Kunskapskraven inom Lgr11 tar fasta på nyckelordet "resonera". För att få godkänt i ämnet måste man kunna resonera på olika nivåer. "Att kunna resonera" var den aspekt som "resoneraprojektet" arbetade med 2006⁴. Där kom vi fram till att det var möjligt att dekonstruera elevresonemang genom att titta på lärares inneboende kvalitetsuppfattning av "att kunna resonera" inom ramen för ämnet samhällskunskap i åk 9 i grundskolan. Dessa resultat ledde så småningom fram till en första variant av bedömningsstöd av beläggskaraktär. Problemet med att kunna "resonera" inom ramen för ett ämne är att vi måste kunna bedöma ett resonemang som är ämnesinriktat. "Att kunna resonera" skulle ju också kunna sägas vara en

⁴ Resoneraprojektet

allmängiltig förmåga som egentligen "enbart" mäter kunskaper inom modersmålet. Genom "resoneraprojektet" tyckte vi att vi fick metoder att knyta förmågan till ett ämne. Många av de deltagande lärarna vittnade om att de tyckte att bedömningsstödet hjälpte dem att dels se kvaliteter i ett svar som i en holistisk bedömning såg svaga ut (ofta kortare och språkligt bristfälliga svar), dels att de kunde "se igenom" svar som vid en holistisk bedömning såg starka ut (ofta längre, välskrivna svar). Denna slutsats är inte vetenskapligt prövad än, utan får stå för de deltagande lärarna.

Inom ramen för Prov- och bedömningsbanken i samhällskunskap "Se den andre" ¹ hade vi ett större statistiskt material i ett enda prov: 576 provsvar med elevernas betyg i samhällskunskap eller So och svenska eller svenska 2. En misstanke vid essäsvår i olika ämnen är att man inte mäter ämnesmässigt resonemang, utan kunskaper i svenska.

Vi gjorde då en statistisk jämförelse och skrev i den rapporten " **Bedömning inom ett ämne som samhällskunskap bygger ofta på kvalitativa överväganden**".

Korrelationer mellan betyg i Sv, betyg i So/Sh och provbetyget

Vi har sett på sambandet mellan provbetyget och höstbetygen i So/Sh respektive Sv/Sv2 för att se om det finns något starkare samband med något av höstbetygen. Det kunde finnas anledning att tro att en öppen fråga t.ex. kan gynna elever med mycket god skrivförmåga, kanske t.o.m. mer än sambandet med betyg inom So/Sh

Vid en regressionsanalys så visar det sig att r (se 5.5.2) = 0,60 för svenska och 0,58 för sambandet So (n=544).

Det finns sålunda ett relativt högt samband mellan provbetyget och båda höstbetygen. Det faktiska sambandet underskattas dock något eftersom höstbetyget bara finns på fyrgradig betygsskala medan provbetyget anges på tiogradig skala. Hypotesen att det skulle föreligga ett högre samband mellan provbetyget och det ena av höstbetygen får dock inget stöd, eftersom korrelationerna är praktiskt taget identiska.

Det finns ett relativt högt samband mellan höstbetygen i So/Sh resp. Sv/Sv2, vilket innebär att de i stor utsträckning mäter samma sak, men eftersom sambandet inte är perfekt finns också en unik komponent i varje betyg. För att tydligare framhäva den gemensamma variansen mellan betygen skapades en ny variabel som summan av höstbetygen, och för att framhäva kontrasten mellan de unika komponenterna bildades differensen av de två betygen. R för summan blev 0.63 och för differensen blev korrelationen nära 0. Denna analys visar sålunda att det är den gemensamma komponenten i höstbetygen som samvarierar med provbetyget, medan de unika komponenterna inte har något samband. Det måste dock understrykas att den fyrgradiga skalan för betygen underskattar sambanden, och

gör det svårt att upptäcka några skillnader i samband mellan provbetyget och de två höstbetygen⁵

Således påverkar båda, men några tydliga belägg för att betyget i svenska skulle ha avgjort större påverkan på dessa typer av frågor fanns inte. Detta är dock en svårare statistisk bedömning då det också givetvis finns en påverkan mellan betygen i sig, oavsett svaren på våra provfrågor. Men resultaten visar en klar indikation på att ämneskunskaper är mycket avgörande för en god förmåga att resonera i ämnet.

Jämförelse sidor/skrivmängd och provbetyg

Detta menar vi är en särskilt intressant jämförelse. Man skulle kunna anta att det lönar sig att skriva längre.

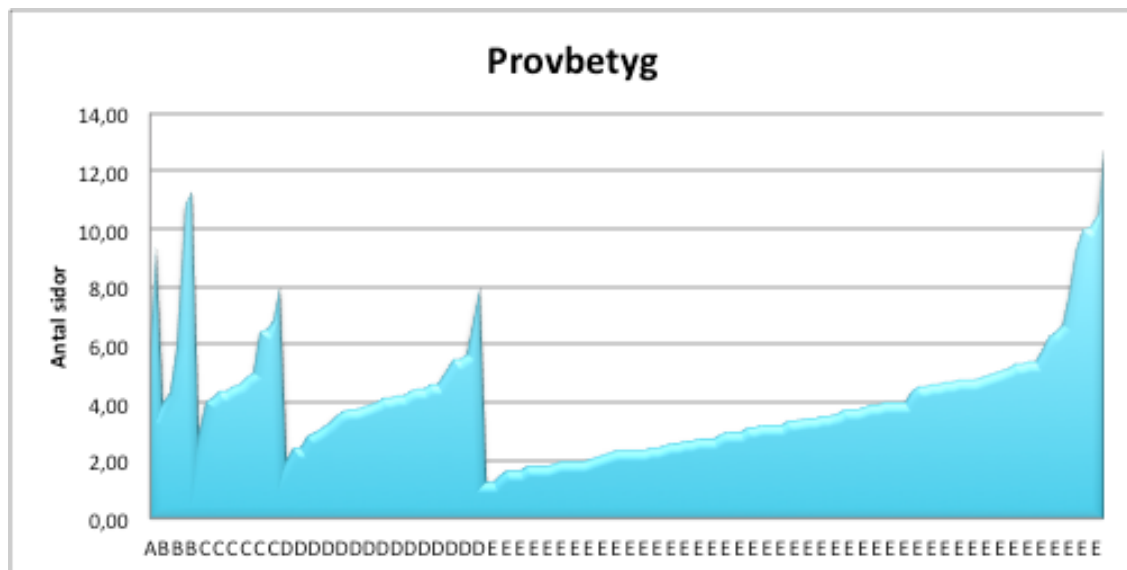
Vi övergår nu till att enbart se på de 400 elevsvaren, som varit med i omdömningen våren -11. Provbetygsunderlaget bör vara säkrare i den populationen än i den stora. Vi har 385 elevsvar där kön angetts. 148 av dem har i omdömningen noll som skillnad mellan de tre omdömnarna, 156 har ett betygssteg i skillnad, totalt 304 svar som har ingen eller max ett stegs betygsskillnad mellan de tre omdömnarna. För att få större undersökningsmängd så har vi, i de fall det funnits ett stegs skillnad mellan bedömnarna, tilldelat elevsvaren det högre provbetyget. Observera att eleverna inte är bedömda av sina egna lärare här!

Tabell 7. Skrivmängd inom de olika provbetygsstegen

	Kortaste	Längsta	Medel	n=
A	5,7	9,4	7,7	2
B	4	11,3	7,3	5
C	3	8	5,1	13
D	2,1	8	4,2	26
E	1,3	12,8	3,9	89

Diagram 5. Skillnad i skrivmängd/antal sidor inom de olika provbetygsstegen

⁵ PBB/Sh kap 5.4



Tolkning av tabell och diagram kan vara problematiskt, och göras på olika sätt. Några konstateranden dock:

- Högre betyg har i snitt längre svar.
- Dock är skillnaden mellan kortaste och längsta svaren *inom* varje provbetygssteg stort.

Vill man leta belägg för att korta svar med kvalitet kan ge högre betyg så kan man konstatera:

- Det kortaste A-svaret har ungefär samma längd som snittet av C-svaren
- Mer än 1/3 (36 %) av E-svaren har lika långa eller längre svar än det kortaste B-svaret
- Hälften av C-svaren är kortare än de 25 % längsta E-svaren

Man kan vrida siffrorna åt andra hållet, men med viss säkerhet ändå konstatera att det inom högre provbetygsnivåer finns ett antal elevsvar som är påtagligt kortare än elevsvaren på lägre.

De nya kunskapskraven är relativt kognitivt inriktade i alla ämnen. För att kunna göra bedömningar inom ämnet som inte enbart bedömer förmåga att uttrycka sig i allmänhet, så är behovet av bedömningsstöd där lärare kan få hjälp att se igenom och upptäcka ämnesmässiga kvaliteter stort.

Resultaten ger stöd för att vi med hjälp av bedömningsstöd av beläggskaraktär kan få hjälp med att bedöma svar på essäfrågor inom ämnet religionskunskap.