

Ombedömning av nationella prov 2014: ”Processerna spelar roll”

Redovisning av regeringsuppdrag Dnr U2013/7499/GV

Innehåll

FÖRORD	4
SAMMANFATTNING	5
UPPDRAGET	8
BAKGRUND OCH SYFTE	8
DENNA RAPPORT	8
RESULTAT	10
OMBEDÖMDA DELPROV	10
RESULTATENS REPRESENTATIVITET	10
HUR MYCKET SKILJER SIG BEDÖMNINGARNA AV SAMMA ELEV LöSNING?	10
<i>Andel avvikelser, prov i svenska och engelska</i>	10
<i>Andel avvikelser, prov i NO, SO och matematik</i>	11
<i>Variationen mellan skolor</i>	13
<i>Spelar skillnaden roll?</i>	14
BEDÖMNINGSPROCESSEN: HUR BEDÖMS PROVEN VID SKOLORNA?	15
<i>Vem gör bedömningen?</i>	15
<i>Är elevens identitet känd?</i>	16
<i>Tillämpas sambedömning?</i>	17
<i>Bedömningsprocessen påverkar utfallet i bedömningarna</i>	19
<i>Effekten av sambedömning och extern bedömning på skillnader mellan skolor</i>	22
SKOLAN, LÄRAREN ELLER PROVET – VAD FÖRKLARAR AVVIKELSEGRADEN?	25
REKOMMENDATIONER UTIFRÅN RESULTATEN	31
AVSEENDE PROVEN OCH BEDÖMNINGSPROCESSERNA	31
AVSEENDE KOMMANDE OMBEDÖMNINGAR	31
BILAGA 1: GENOMFÖRANDET AV OMBEDÖMNINGEN	33
URVAL AV PROV OCH SKOLOR	33
<i>Ämnen, delprov och årskurser</i>	33
URVALETS SAMMANSÄTTNING	34
INSAMLING AV PROV OCH FÖRDELNING MELLAN OMBEDÖMARE	34
OMFATTNING OCH BORTFALL	35
<i>Resonemang kring bortfallet</i>	37
<i>Frågor om bedömningen</i>	38
BILAGA 2: ETT HARMONISERAT MÅTT FÖR ATT MÄTA SKILLNADER I BEDÖMNING	40
BERÄKNING AV DET HARMONISERADE RESULTATMÅTTET	40

BILAGA 3: METODOLOGISK NOT AVSEENDE BERÄKNING AV VARIATIONEN MELLAN SKOLOR.....	42
BILAGA 4: BEDÖMARÖVERENSSTÄMMELSE MELLAN OLIKA OMBEDÖMARE	44
SKILLNADER MELLAN OLIKA OMBEDÖMARE	44
SKILLNADER I GENOMSNITTSBEDÖMNING SKOLA – OMBEDÖMARE – INSTITUTION	45
BILAGA 5: ANVÄNDNING AV RESULTATEN I TILLSYNNEN	47
BETYDANDE AVVIKELSER PÅ SKOLNIVÅ	47
PUBLICERING AV RESULTAT PÅ SKOL- OCH HUVUDMANNANIVÅ.....	48

Förord

Statens Skolinspektion har sedan 2009 haft i uppdrag av regeringen att varje år genomföra viss central rättning av nationella prov. Syftet med ombedömningen är att stödja en likvärdig bedömning och betygssättning av proven över landet.

Skolinspektionen redovisar här till regeringen resultatet av sjätte årets centrala rättning. Denna avser nationella prov i grundskolan och gymnasiet som genomfördes under 2014, och nationella prov inom den kommunala vuxenutbildningen på gymnasienivå som genomfördes under 2013.

Vid skolinspektionen har följande personer arbetat med att genomföra uppdraget: Björn Ljung (projektledare), Jonna Andersson (enhetschef), utredarna Klara Klingspor, Emmie Chau och Carl Command.

Sammanfattning

Den här rapporten redovisar resultaten från 2014 års omdömning av nationella prov, vilket är den sjätte i ordningen sedan 2009. Syftet med omdömningen är att stödja en likvärdig bedömning och betygssättning av proven över landet. Tidigare års omdömningar har bland annat visat att:

- Det är vanligt att omdömaren sätter ett annat betyg än läraren vid skolan. För vissa delprov skiljer sig bedömningarna åt för mer än varannan elevlösning.
- De största och mest frekventa avvikelserna återfinns för de delprov där eleven ska avge sitt svar i form av en längre text.
- Det är vanligare att omdömaren sätter ett *lägre* betyg än läraren vid skolan, än att omdömaren sätter ett *högre* betyg.

2014 års omdömning bekräftar samtliga dessa slutsatser. Årets rapport fördjupar och kompletterar dessutom analysen inom ett antal områden.

Tidigare rapporter har satt fokus på andelen avvikelser mellan skolornas bedömningar och Skolinspektionens omdömning. Fokus i årets analyser är att utöver detta även skatta skillnaderna i genomsnittsbedömning mellan olika skolor.¹ Det vill säga, sätter vissa skolor *i genomsnitt* högre provbetyg på likvärdiga elevlösningar än andra skolor? Hur stora är i så fall dessa skillnader, och vad skulle kunna förklara dem? De slutsatser som dras är i korthet följande:

Resultaten indikerar att skillnaderna i bedömning mellan olika skolor kan uppgå till mer än ett betygssteg i snitt

För de delprov som omdömts indikerar resultaten att en elev på en skola som sätter relativt låga betyg typiskt sett får ett delprovsbetyg som ligger cirka ett halvt betygssteg lägre på skalan A-F än om han eller hon lämnat in samma lösning på en skola som sätter relativt höga betyg. Sett till de tio procent mest utstickande skolorna växer skillnaden till mellan ett och ett och ett halvt betygsstegs skillnad i snitt i bedömningen av likvärdiga elevlösningar.

Resultaten indikerar alltså att i ett hypotetiskt scenario där samma klass får lämna in sina prov för bedömning vid två olika skolor, varav den ena skolan hör till de mest generösa i bedömningen av proven och den andra hör till de mest strikta, så skulle elevernas genomsnittsbetyg för det aktuella delprovet

¹ Utifrån en statistisk metod där resultaten från omdömningen används som en gemensam referenspunkt för att indirekt skatta skillnader mellan olika skolor.

ligga drygt ett betygssteg högre på skalan A-F på den första skolan än vid den andra skolan.

Eftersom resultatet på de nationella proven för många lärare spelar en viktig roll vid betygssättningen av elevens kunskaper, är det rimligt att anta att den variation i bedömningen av likvärdiga elevlösningar som ombedömningen visar även kan slå igenom på elevens ämnes- och kursbetyg. Hur ofta så är fallet går dock inte att uttala sig om utifrån ombedömningsresultaten.

Skillnaderna i bedömning beror på både läraren, skolan och proven i sig

Beror skillnaderna i bedömning på enskilda lärare, på skolövergripande faktorer, eller på något annat? Analysen tyder på att flera olika faktorer bidrar till variationen. Det verkar förekomma systematiska effekter på både lärarnivå och skolnivå, det vill säga vissa lärare sätter generellt sett högre betyg än andra lärare för motsvarande elevprestation, och vissa skolor sätter generellt sett högre betyg än andra skolor.

En stor del av variationen i bedömningarna kan dock inte förklaras av detta. Exempelvis kan två lärare med samma allmänna tendens för att sätta höga eller låga betyg ändå göra olika bedömning av en viss elevlösning. Resultaten indikerar därför att en betydande del av bedömaravvikelseerna bottnar i "slumpvisa" skillnader i uttolkningen av bedömningsanvisningarna, snarare än i faktorer utanför proven såsom skolornas olika arbetssätt eller olika lärares relativa generositet i betygssättningen, även om dessa faktorer också bidrar till den samlade variationen.

Bedömningsprocesserna påverkar provbetygen

Enligt de svar som samlats in från skolorna så tillämpar olika skolor olika processer vid bedömningen av proven. För ungefär vartannat insänt prov tillämpade skolan en organiserad process för sambedömning, det vill säga en process där flera lärare på något sätt samverkat kring bedömningen av provet. Extern bedömning, det vill säga att någon annan än den ordinarie ämnesläraren har huvudansvaret för betygssättningen av provet, är mer ovanligt och tillämpades för runt vart femte insänt prov. Kan skillnaderna i valet av process bidra till skillnaderna i bedömning mellan skolor?

Tidigare ombedömningar har konstaterat att det är vanligare att Skolinspektionens ombedömare sätter ett *lägre* betyg än läraren vid skolan, än att ombedömare sätter ett *högre* betyg. Årets analys visar att det bara är fallet för de prov där ursprungsbedömningen gjorts av elevens ordinarie lärare utan inslag av sambedömning. För prov som sambedömts på skolan, eller bedömts av någon

annan än elevens ordinarie lärare, finns för de flesta ombedömda delprov ingen generell tendens till att skolans bedömning är mer generös än ombedömarnas.

Både sambedömning och extern bedömning verkar alltså ha en viss återhållande effekt på betygsättningen av delprovet, det vill säga leder till ett något lägre betyg för en given elevlösning i snitt än om elevens ordinarie lärare bedömer provet ensam. Däremot kan resultaten inte ge något entydigt svar på om eller i vilken grad sambedömning och extern bedömning också minskar skillnaderna mellan olika skolor.

Ökad likvärdighet kräver insatser på flera nivåer

Sammanfattningsvis understryker resultaten att en bred palett av åtgärder är viktigt om likvärdigheten i bedömningarna ska kunna öka samtidigt som proven behåller möjligheten för elever att visa sådana kunskaper som inte låter sig prövas i frågor med fasta svarsalternativ. Såväl enskilda lärare som rektorer, huvudmän och provutvecklare har viktiga uppgifter i detta, men ingendera parten kan ensam lösa problemet.

Skolinspektionen välkomnar vidare det arbete som för närvarande bedrivs för att samordna struktur och terminologi mellan de olika ämnesproven. Långsiktigt torde det ge bättre förutsättningar för systematiska utvärderings- och utvecklingsarbete på systemnivå än vad som idag är fallet.

Uppdraget

Bakgrund och syfte

Sedan 2009 har Skolinspektionen årligen fått i uppdrag av regeringen att genomföra viss central rättning av nationella prov. Syftet med den centrala rättningen är att stödja en likvärdig bedömning och betygssättning av proven över landet. Uppdraget genomförs genom att Skolinspektionen varje år samlar in kopior av genomförda prov från ett urval av skolor. Proven avidentifieras och bedöms sedan på nytt av erfarna lärare som rekryterats av Skolinspektionen.

Bakgrunden till uppdraget är risken för skillnader i hur olika lärare och skolor bedömer proven, framförallt avseende de delprov där elevens svar ges i form av en eller ett mindre antal längre texter som inte kan bedömas utifrån ett entydigt facit. Tidigare års ombedömningar har visat att det i dessa fall inte är ovanligt att det skiljer två betygssteg eller mer mellan två olika lärares bedömning av samma elevsvar. Betygen på nationella prov påverkar i nästa led betyget på ämnet eller kursen. Med tanke på betygens många gånger avgörande roll för urval till högre studier eller vid ansökan till jobb kan det få stora konsekvenser för enskilda elever om till exempel prestationer som anses motivera ett A på en skola endast ges betyget C på en annan.

De prov som ombedöms har inga fasta svarsalternativ, och därför heller inget "facit". Att Skolinspektionens ombedömare gör en annan bedömning av en elevlösning än skolan betyder inte att skolan gjort en felaktig bedömning. En stor variation i bedömningen av likvärdiga elevlösningar mellan *olika skolor* är dock problematiskt, eftersom ett av provens huvudsyften är stödja en likvärdig och rättvis bedömning och betygssättning. Därför är det av stor vikt att skolorna får ett underlag som ger dem möjlighet att jämföra sina bedömningar av de nationella proven med andra skolors, och att Skolinspektionen följer upp dessa indikatorer i sin tillsynsverksamhet.

Denna rapport

Skolinspektionen redovisar med denna rapport resultatet av sjätte årets centrala rättning. Rapporten presenterar både deskriptiva resultat och ett antal statistiska analyser med bäring på frågan om likvärdig bedömning av proven. Dessutom redogörs i korthet för myndighetens användning av resultaten vid tillsyn av skolor och huvudmän.

Den detaljerade redogörelsen för projektets genomförande, urvalet av skolor och delprov samt beräkningstekniska detaljer har samlats i bilagor. Den läsare

SKOLINSPEKTIONEN

2015-11-16
9 (48)

som vill skaffa sig en djupare förståelse av resultaten uppmanas att läsa dessa bilagor.

Resultat

Ombedömda delprov

2014 års ombedömning innefattade proven i svenska, engelska och matematik i årskurserna 6 och 9 samt gymnasieskolans grundläggande kurser. För årskurs 6 och 9 omfattades även proven i kemi, biologi och fysik (NO) samt samhällskunskap, religionskunskap, geografi och historia (SO). Utöver detta ombedömdes prov i svenska, engelska och matematik inom den kommunala vuxenutbildningen på gymnasienivå.

Från varje prov valdes endast *ett delprov* ut för ombedömning, nämligen det där risken för skillnader i bedömning mellan olika lärare bedömts som störst. Mer om urvalet av prov och delprov för ombedömning finns att läsa i Bilaga 1.

Resultatens representativitet

Urvalet av skolor för ombedömning följer från och med 2014 Skolinspektionens tillsynscykel. Urvalet är alltså inte gjort för att vara representativt på nationell nivå, till exempel avseende region eller huvudmannatyp. Alla resultat som presenteras i denna rapport baseras på det faktiska urvalet av ombedömda prov, och kan inte med säkerhet generaliseras till riket i stort. Skolinspektionen bedömer dock att urvalets sammansättning och storlek är sådant att det ändå är möjligt och meningsfullt att använda som en indikation på den allmänna situationen i riket.

Det totala bortfallet på provnivå är 16 procent, men varierar kraftigt mellan olika delprov. Framförallt är det högt för vuxenutbildningen samt för vissa av NO- och SO-proven i årskurserna 6 och 9. Resultaten för dessa prov får därför anses mindre stabila än för språk- och matematikproven. Den kompletta bortfallstabellen redovisas i Bilaga 1.

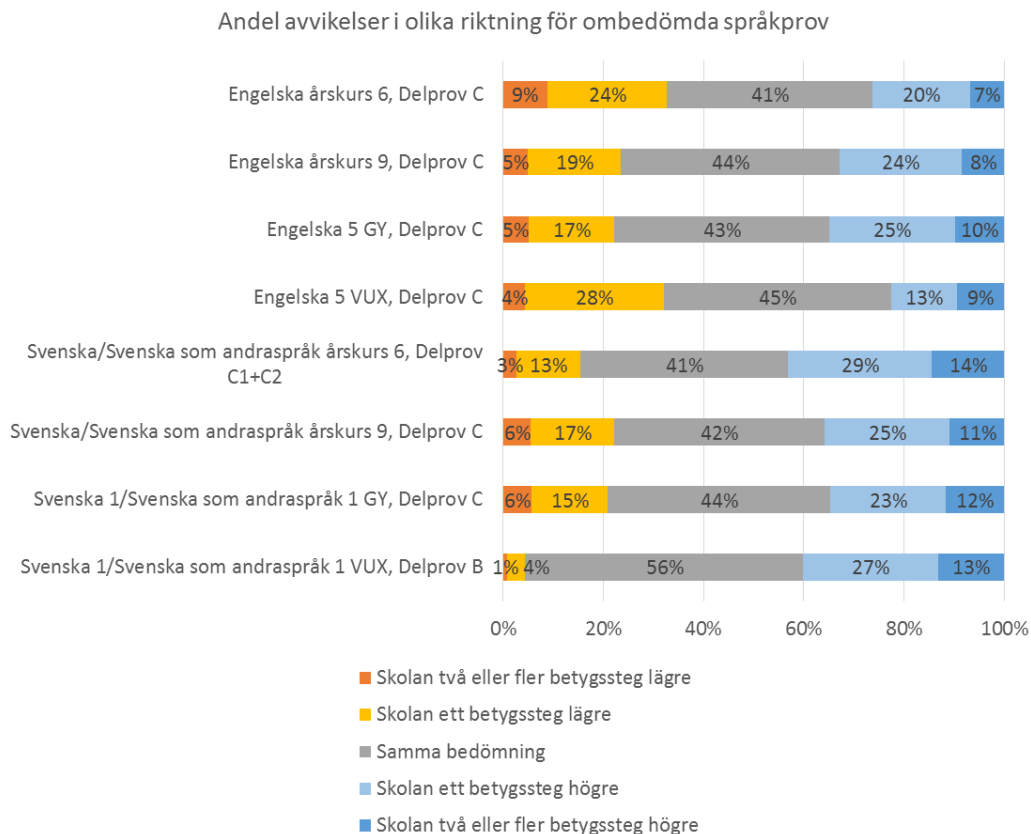
Hur mycket skiljer sig bedömningarna av samma elevlösning?

Andel avvikelser, prov i svenska och engelska

För de ombedömda språkproven, som alla är uppsatsprov bestående av ett eller ett mindre antal omfattande fritextsvar, är det bara i knappt 45 procent av fallen som skolan och ombedömaren sätter samma betyg på skalan A-F. I ungefär 35 procent av fallen sätter ombedömaren ett betyg som är ett eller flera steg lägre än skolans bedömning, och i ungefär 20 procent av fallen sätter ombedö-

maren ett betyg som är ett flera betygssteg lägre än skolans bedömning. Andelen avvikelser mellan skolans bedömning och ombedömningen är i linje med resultaten från tidigare år.²

Andelen avvikelser är mycket likartad för proven i årkurs 6, årkurs 9 och på gymnasiet.



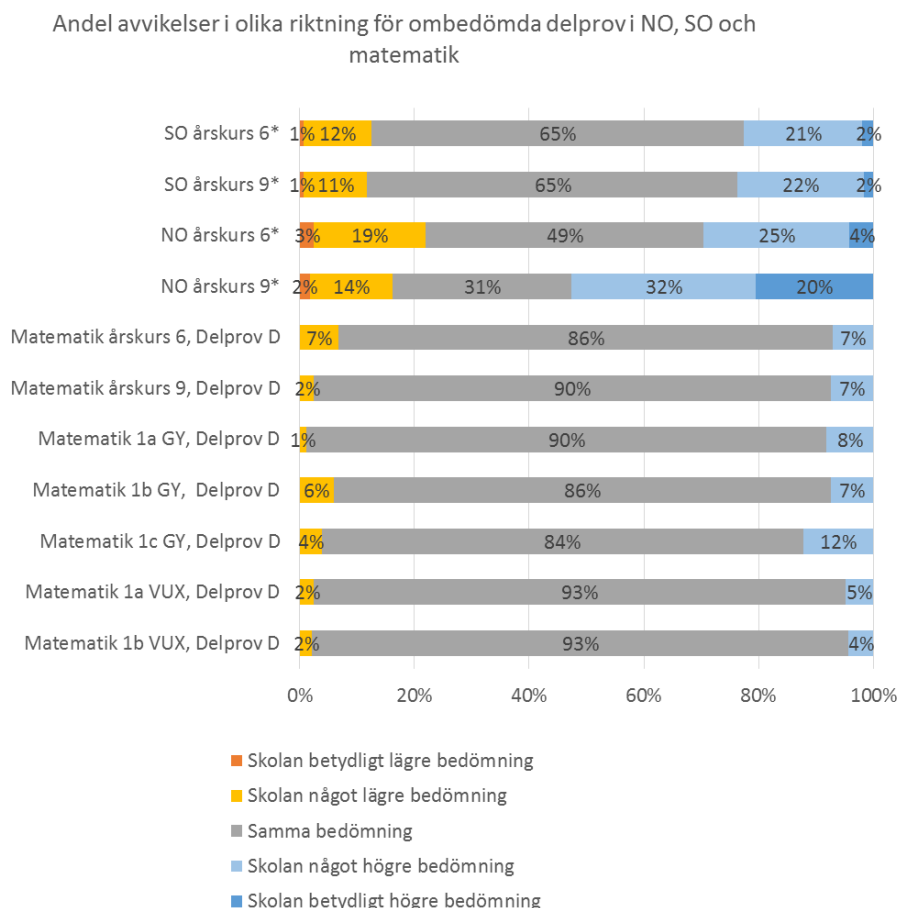
Figur 1: Andel avvikelser mellan skolans bedömning och ombedömningen för ombedömda delprov i svenska och engelska.

Andel avvikelser, prov i NO, SO och matematik

För NO-, SO- och matematikproven sätts inga delprovsbetyg på skalan A-F, utan endast ett antal *poäng*. För NO och SO är termen *belägg*. Antal belägg som maximalt kan uppnås varierar mellan olika delprov. För att kunna ge en översiktlig samlad bild över andelen avvikelser för de delprov där bokstavs- eller poängbetyg

² Där resultaten är jämförbara. Under de år ombedömningen pågått har såväl provens utformning som de betygsskalor som används i olika årskurser förändrats utifrån reformer vid flera tillfällen, vilket försvårar en analys av resultaten över tid.

inte sätts har beläggen för varje delprov räknats om till en gemensam skala 0-20. I Bilaga 2 redovisas i detalj hur resultatmättet beräknats.



Figur 2: Andel avvikelser mellan skolans bedömning och ombedömningen för ombedömda delprov i NO-ämnen (Kemi, Biologi, Fysik), SO-ämnen (Samhällskunskap, Religionskunskap, Historia, Geografi) och Matematik. I diagrammet redovisas prov där det skiljer mindre än 2 på den standardiserade skalan 0-20 mellan de två bedömningarna som "samma bedömning". En skillnad på 2 motsvarar approximativt 10 procent skillnad i antal utdelade belägg mellan de två bedömningarna. Skillnader på mellan 2-6 (det vill säga cirka 10-30 procent skillnad i antal utdelade belägg) redovisas som något lägre/högre bedömning och skillnader på mer än 6 (mer än 30 procent skillnad i antal belägg) redovisas som betydligt högre/lägre bedömning.

*Sammanslagning av resultat från flera ämnesprov, se Bilaga 1 för en redogörelse över antal prov och skolor per ämne/delprov.

De olika ombedömda delproven i NO, SO och matematik har mycket olika konstruktion och prövar olika typer av färdigheter. Vissa består till exempel endast av en lång skrivuppgift, medan andra består av flera kortare uppgifter som lämnar mindre utrymme för subjektivitet i bedömningen. Vissa prövar övervä-

gande faktakunskaper, medan andra prövar kritiskt resonerande och argumentation. Det är därför naturligt att bedömaröverensstämmelsen är olika hög för de olika proven. Vad som kan konstateras är dock att flera av proven uppvisar en betydande variation sett till hur två olika bedömare bedömer samma elevlösning.

Variationen mellan skolor

På många sätt är det mer intressant att titta på spridningen i bedömningar *mellan olika skolor*, än mellan en viss skolas resultat och omdömningen.

Det går inte att utifrån omdömningen avgöra *andelen* avvikelser mellan olika skolor, eftersom varje skola bara bedömer sina egna prov. Resultaten från omdömningen kan däremot användas som en gemensam referenspunkt för att undersöka om olika skolor *i genomsnitt* sätter lika höga eller låga betyg för likvärdiga elevlösningar.³ I Bilaga 3 beskrivs tillvägagångssättet i beräkningarna närmare.

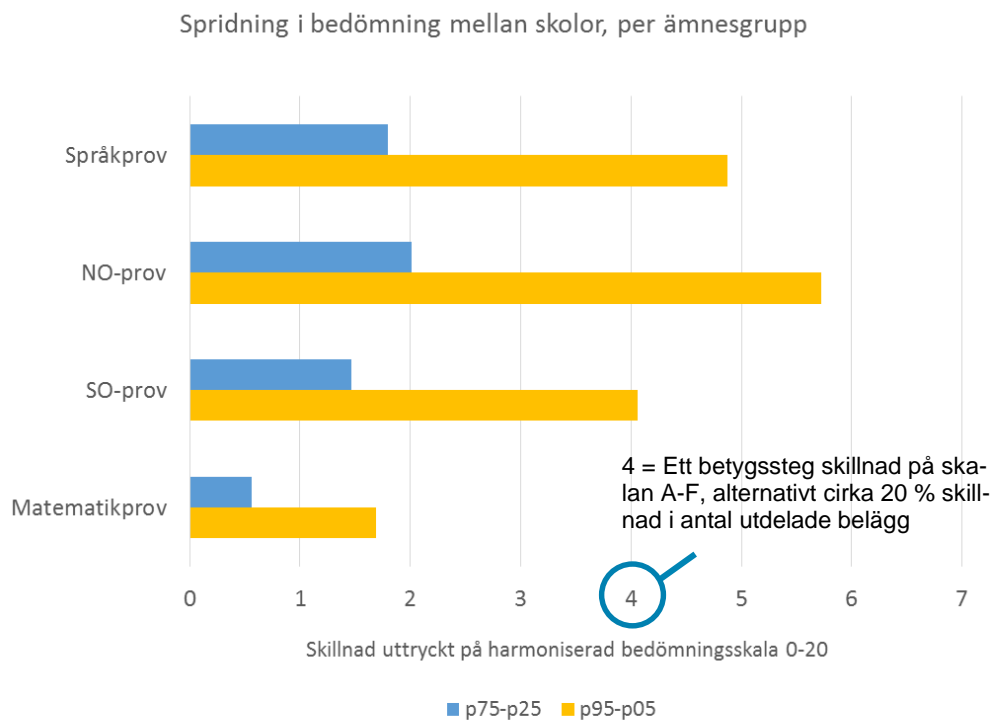
I figur 3 nedan redovisas den beräknade genomsnittliga skillnaden mellan olika skolor i bedömningen av likvärdiga elevlösningar, redovisat per ämnesgrupp. Resultaten visar att en elev på en skola som sätter relativt låga betyg kan förvänta sig att få ett betyg som ligger cirka ett halvt betygssteg lägre än om han eller hon lämnat in samma svar på en skola som sätter relativt höga betyg.

Ser vi till de allra mest utstickande skolorna så växer skillnaderna. En elev som går på en av de fem procent skolor i landet som har den striktaste betygssättningen av proven kommer typiskt sett att få ett betyg som ligger ett till ett och ett halvt betygssteg lägre på skalan A-F, än om han eller hon lämnat in samma svar på en av de skolor som hör till de fem procent mest "generösa" i betygssättningen av proven.

Spannet mellan de skolor som är relativt strikta respektive relativt generösa i betygssättningen är ungefär lika stort för de omdömda delproven i NO- och SO-ämnena som för språkämnen.⁴ Den lägsta spridningen mellan skolor har de omdömda proven i matematik.

³ "Likvärdiga elevlösningar" syftar här på en hypotetisk situation där olika skolor bedömt samma elevlösningar.

⁴ Uttryckt på den harmoniserade bedömningsskalan 0-20. För delproven i NO och SO ges inga bokstavsbedömningar.



Figur 3: Estimat av spridning i bedömning av likvärdiga elevlösningar mellan olika skolor. p75-p25 betyder skillnaden mellan den 75:e percentilens och den 25:e percentilens skolor, med avseende på genomsnittlig avvikelse från omedömningen. För språkproven motsvarar 4 på skalan ett betygsstegs skillnad på skalan A-F. För övriga prov motsvarar 4 på skalan en skillnad på ca 20 procent i antal utdelade belägg.

Spelar skillnaden roll?

Eftersom endast ett delprov per prov omedöms är det inte givet att en uppmätt skillnad för ett enskilt delprov får genomslag på det slutliga provbetyget. Ofta räcker det dock med ett betygsstegs skillnad för ett delprov för att det totala provbetyget ska påverkas. För till exempel proven i svenska i årskurs 9 är det fallet för knappt en tredjedel av de möjliga kombinationerna av delprovsbetyg. Ett annat exempel är NO-proven i åk 9, där en skillnad på 2-3 belägg, vilket är vanligt förekommande för de omedömda delproven, inte sällan har en effekt på det totala provbetyget.

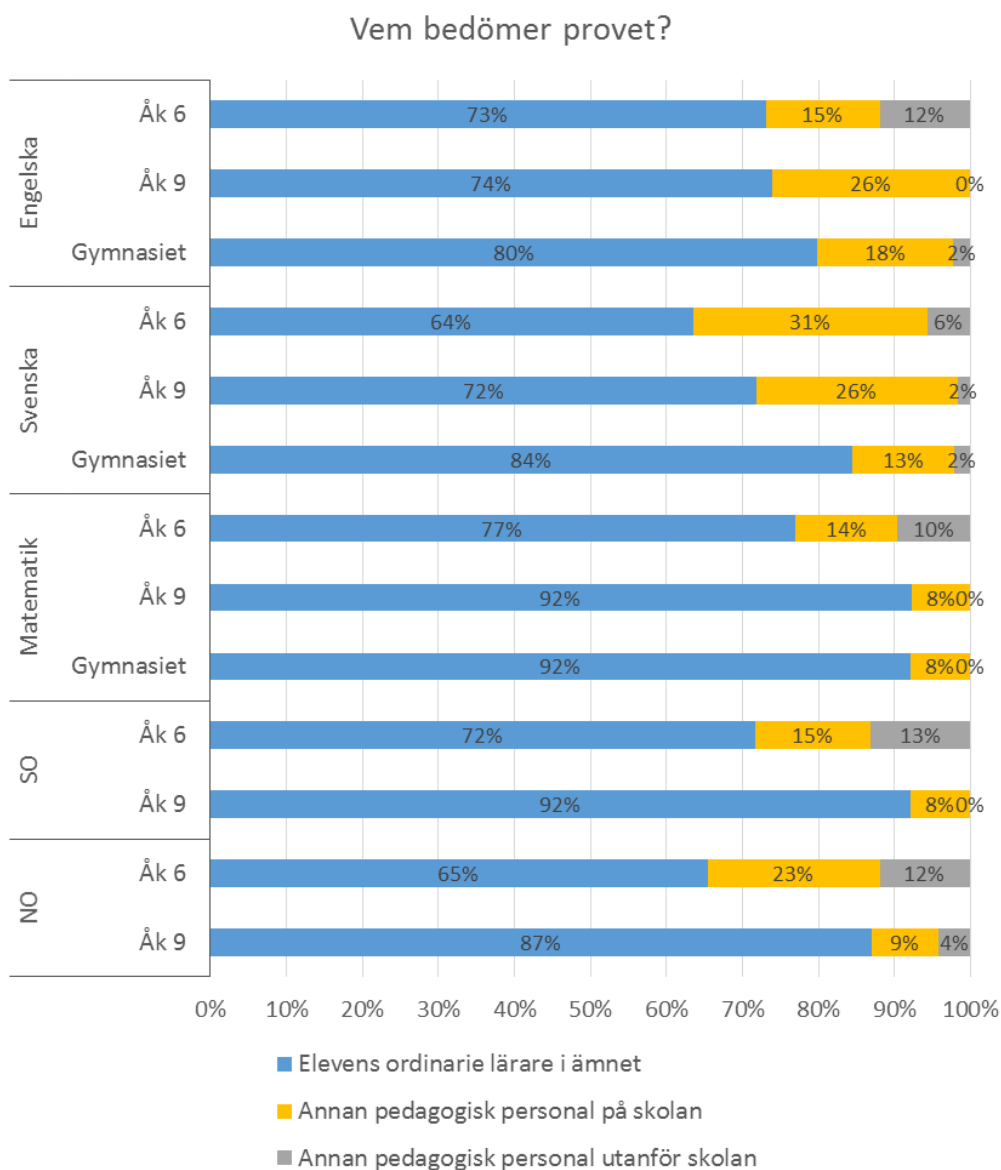
Det står alltså klart att de avvikelser som uppmäts i omedömningen på delprovsnivå är av en storleksordning som *kan* leda till att det totala provbetyget påverkas, även om det långt ifrån alltid är fallet.

Bedömningsprocessen: hur bedöms proven vid skolorna?

I samband med att proven för omdömning samlas in får skolorna också besvara tre frågor kring hur bedömningen av provet gått till. Svaren från de skolor som medgett att svaren analyseras redovisas nedan. SO- och NO-ämnena redovisas även här sammanslaget per ämnesgrupp.

Vem gör bedömningen?

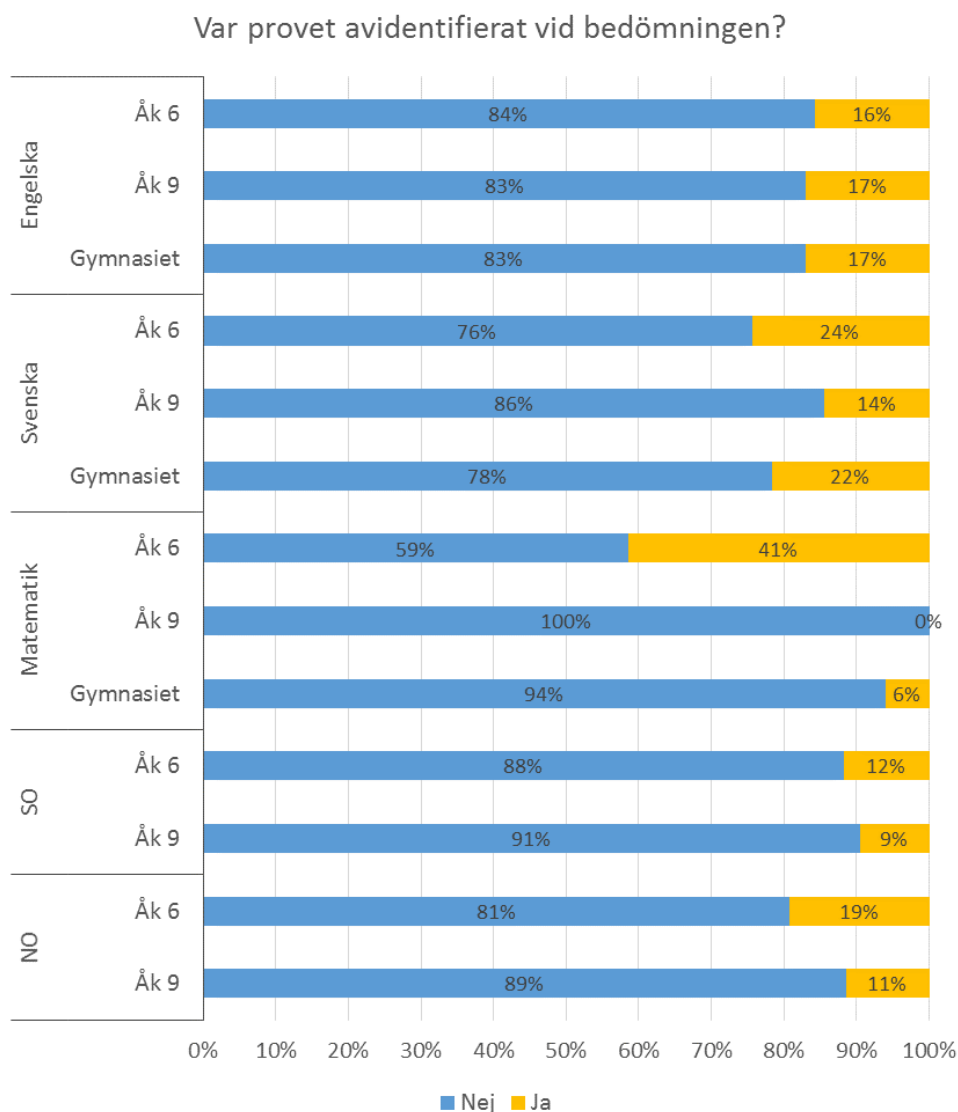
Resultaten visar att det klart vanligaste är att elevens ordinarie ämneslärare har huvudansvaret för bedömningen av elevens prov. Det är fallet för fyra av fem omdömda prov. Att någon annan än den ordinarie ämnesläraren står för bedömningen är dock vanligare i de lägre årskurserna. I årskurs 6 bedöms ungefär var tredje elevlösning av någon annan än den ordinarie ämnesläraren.



Figur 4: Skolornas svar på frågan om vem som bedömt den insända provet. Frågan besvaras för varje enskild elevlösning, men är frivillig för skolan att besvara. Se Bilaga 1 för antal svar per delprov.

Är elevens identitet känd?

Provet är typiskt sett inte avidentifierat vid bedömningen: endast för 15 procent av de inrapporterade proven var detta fallet. Högst var andelen avidentifierade prov i matematik årskurs 6, men då dessa resultat endast bygger på information om 365 prov från 15 skolor går det inte att säga att resultaten återspeglar den allmänna situationen i landet.



Figur 5: Skolornas svar på frågan om det insända provet var avidentifierat vid bedömningen. Frågan besvaras för varje enskild elevlösning, men är frivillig för skolan att besvara. Se Bilaga 1 för antal svar per delprov.

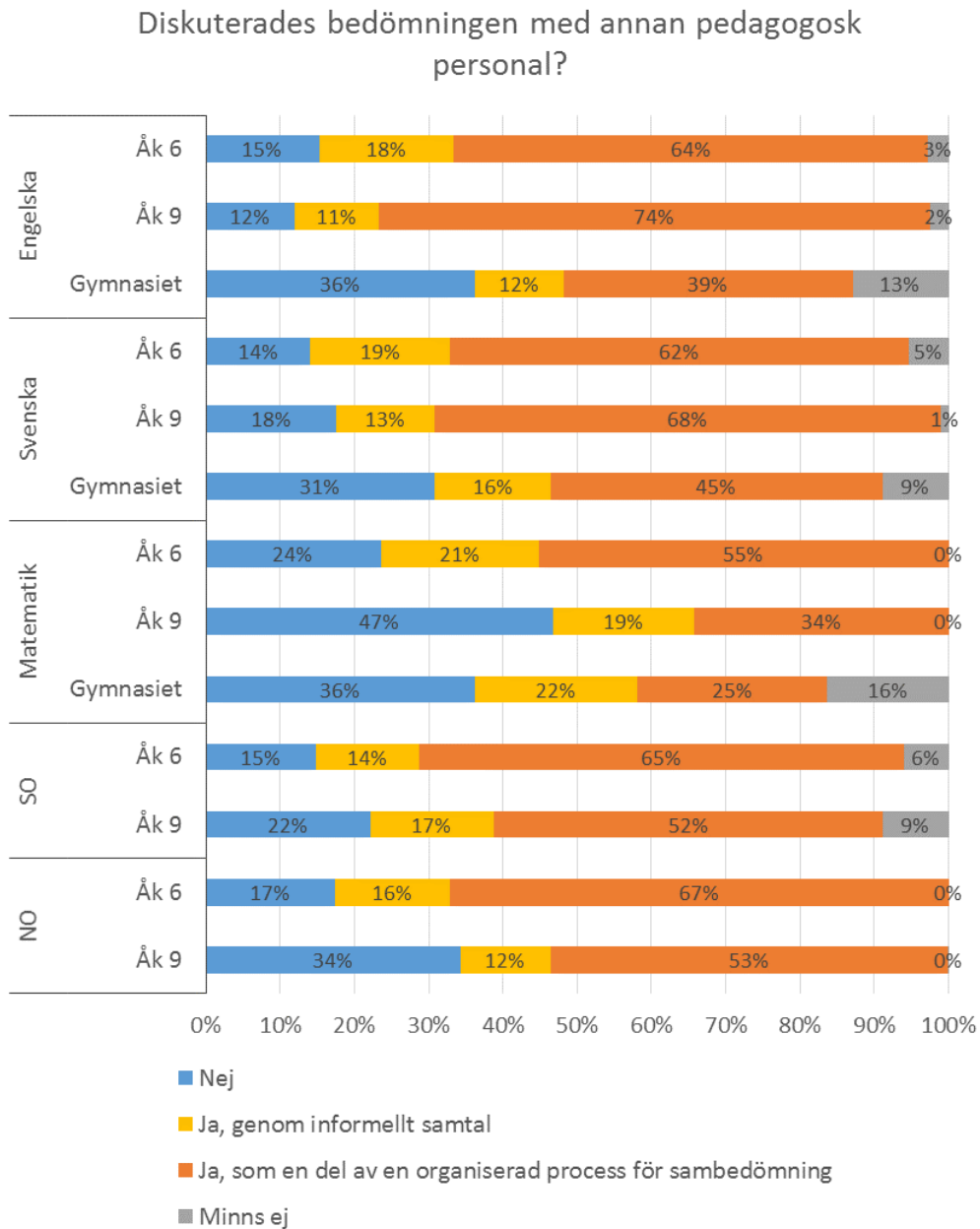
Tillämpas sambedömning?

För en knapp majoritet (51 procent) av de inrapporterade proven har bedömningen skett inom ramen för en organiserad process för sambedömning. Drygt ett av fyra prov bedömdes helt utan sambedömning, och i ett av sex fall var inslaget av sambedömning av mer informell karaktär. För 7 procent av proven minns inte eller vet inte den som inrapporterat provet om sambedömning skett.

Avseende förekomsten av organiserad sambedömning finns en tydlig skillnad mellan grundskolan, där ungefär två av tre insända prov bedömts i organiserad sambedömning, och gymnasieskolan där drygt ett av tre insända prov bedöms på detta sätt.

Det finns även en tydlig skillnad mellan skolor med kommunal och enskild huvudman, där organiserad sambedömning är betydligt vanligare på kommunala skolor. Organiserad sambedömning skedde till exempel för 65 procent av de omdömda proven i årskurs 6 från kommunala skolor, men endast för 34 procent av proven från skolor med enskild huvudman. För gymnasiet var motsvarande andelar 49 procent respektive 25 procent.

Eftersom urvalet i omdömningen på gymnasienivå innehåller en överrepresentation av skolor med enskild huvudman (Se Bilaga 1) så är sannolikt graden av sambedömning på gymnasiet något högre sett till hela riket än vad som redovisas i figur 6.



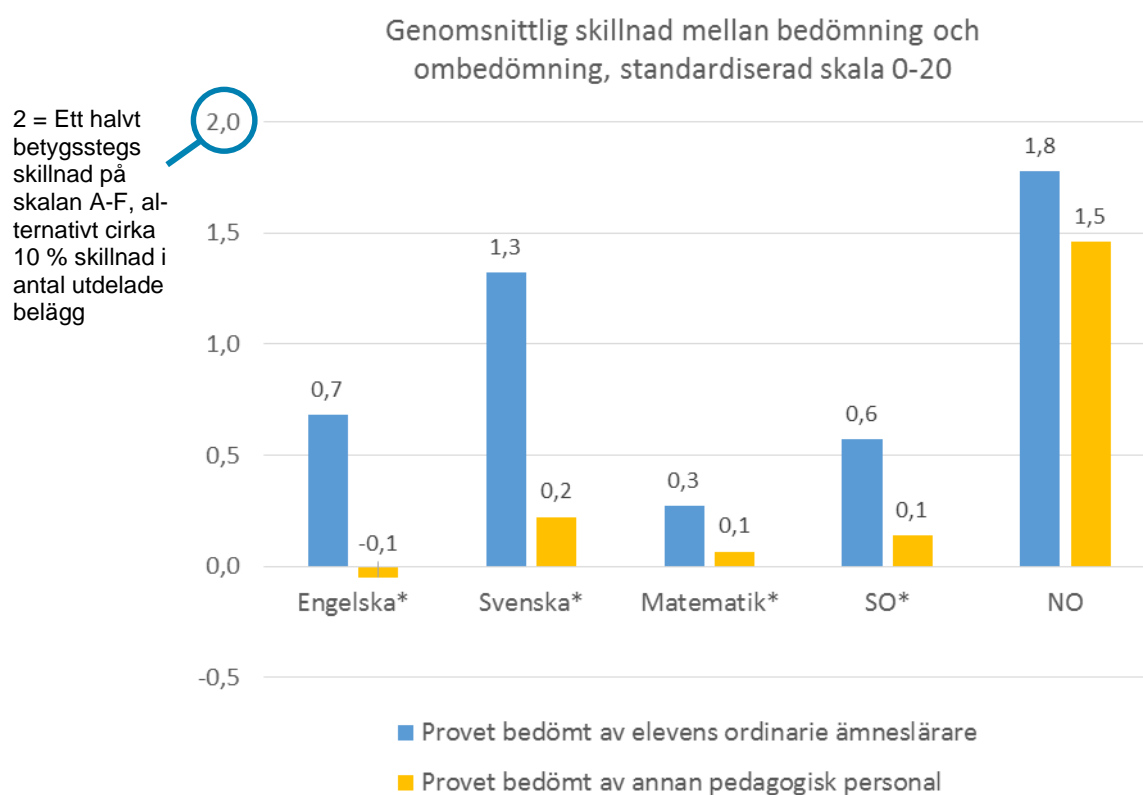
Figur 6: Skolornas svar på frågan om sambedömning skett vid bedömningen av det insända provet. Frågan besvaras för varje enskild elevlösning, men är frivillig för skolan att besvara. Se Bilaga 1 för antal svar per delprov.

Bedömningsprocessen påverkar utfallet i bedömningarna

Som figur 7 nedan visar så är den genomsnittliga skillnaden mellan skolans bedömning och ombedömningen signifikant mindre för prov som bedömts av någon annan än elevens ordinarie lärare, för fyra av fem ämnesgrupper. Den ge-

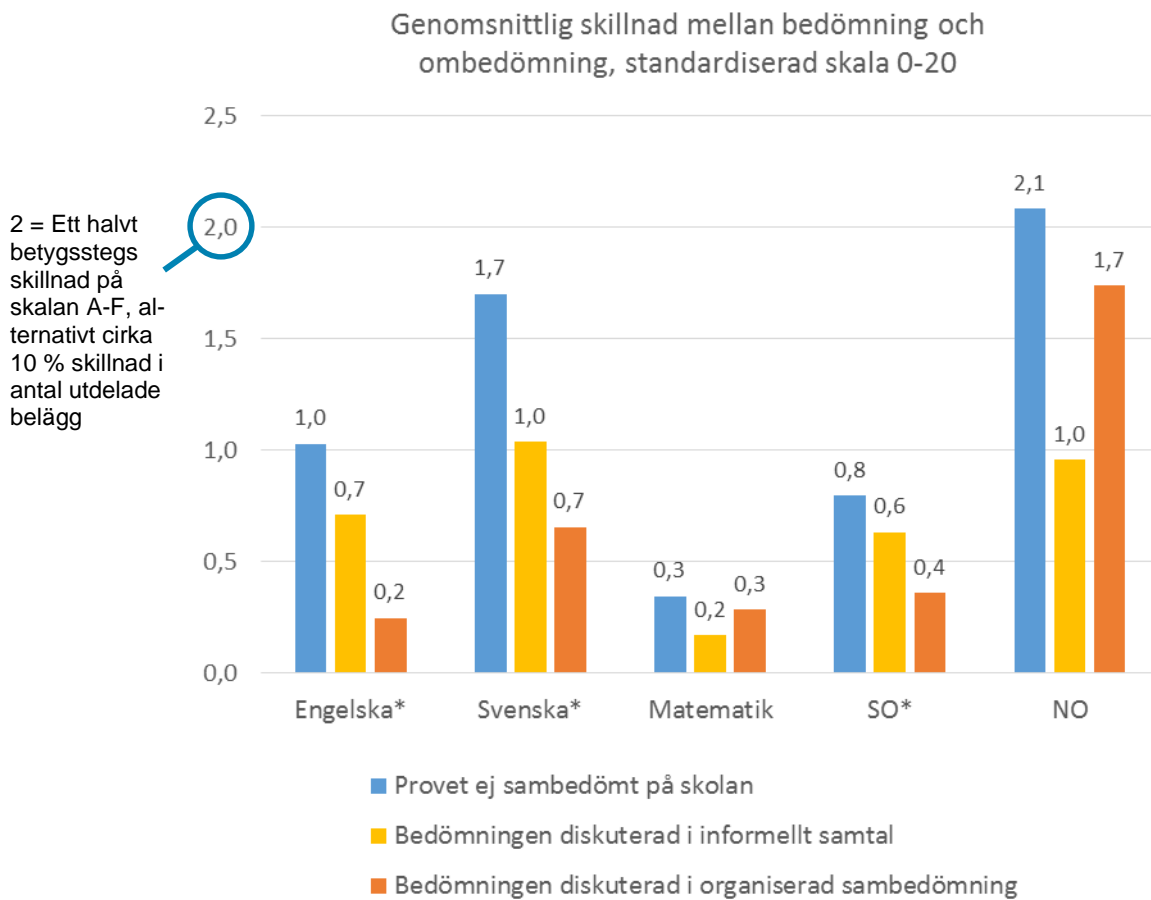
nerella tendensen till att skolan gör en något mer generös bedömning än ombedömaren minskar kraftigt när någon annan än den ordinarie läraren står för bedömningen. I flera fall försvinner den helt.

Tydligast är resultaten för språk- och matematikproven. För SO- och NO-proven ses olika resultat för olika prov, och i vissa fall signifikanta skillnader i *motsatt* riktning. Antalet skolor som resultaten bygger på är dock litet för de enskilda SO- och NO-ämnena. För NO-proven ses heller ingen signifikant effekt av extern bedömning på aggregerad nivå.



Figur 7: Genomsnittlig skillnad mellan bedömning vid skolan och ombedömningen för prov som på skolan bedömts av elevens respektive av annan pedagogisk personal på eller utanför skolan. *Signifikant lägre genomsnittsavvikelse för externt bedömda prov

Även sambedömning verkar, framförallt för språkämnen, ha en tillbakahållande effekt på betygsättningen. För tre av fem ämnesgrupper är skillnaden mellan skolans och ombedömarens bedömningar signifikant mindre för prov som bedömts i en organiserad process för sambedömning, än för prov där ingen sambedömning skett. För språkproven är det också tydligt att organiserad sambedömning har en större återhållande effekt på betyget än mer informell sådan.



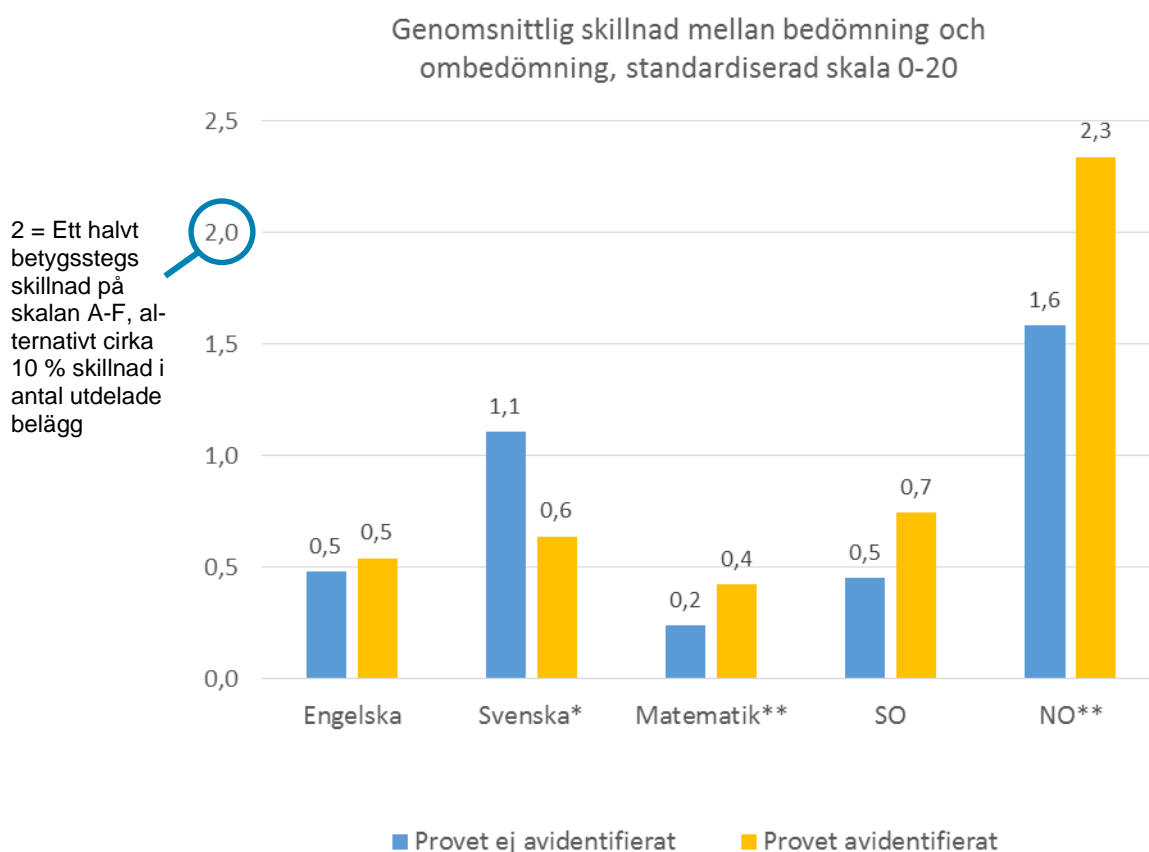
Figur 8: Genomsnittlig skillnad mellan bedömning vid skolan och ombedömningen för prov där bedömaren inte diskuterat bedömningen med någon annan, för prov där bedömningen diskuteras genom informellt samtal, respektive prov som bedömts i organiserad sambedömning.
*Signifikant lägre genomsnittsavvikelse för sambedömda prov

Avidentifiering av proven verkar intressant nog inte ge samma resultat. Endast för en ämnesgrupp av fem ses ett signifikant resultat i den "väntade" riktningen. För två ämnesgrupper ses en signifikant skillnad i motsatt riktning: avidentifierade prov bedöms mer generöst än prov där bedömaren känner till elevens identitet.

Om man för svenska och engelska på gymnasiet⁵ skiljer ut effekten av vem som gjort bedömningen (ämnesläraren eller någon annan) från om provet var avidentifierat eller inte, så visar det sig att bedömareffekten är den helt domine-

⁵ De prov som har de högsta bastalen och därför medger en sådan analys

rande, medan effekten av avidentifiering är liten eller icke existerande. Möjligt tyder detta på att den ordinarie ämnesläraren i många fall kan avgöra vilken elev som skrivit provet även om det "avidentifierats", till exempel via handstil eller uttrycksätt.



Figur 9: Genomsnittlig skillnad mellan bedömning vid skolan och ombedömningen för prov som var avidentifierade vid bedömningen respektive där elevens identitet var känd.

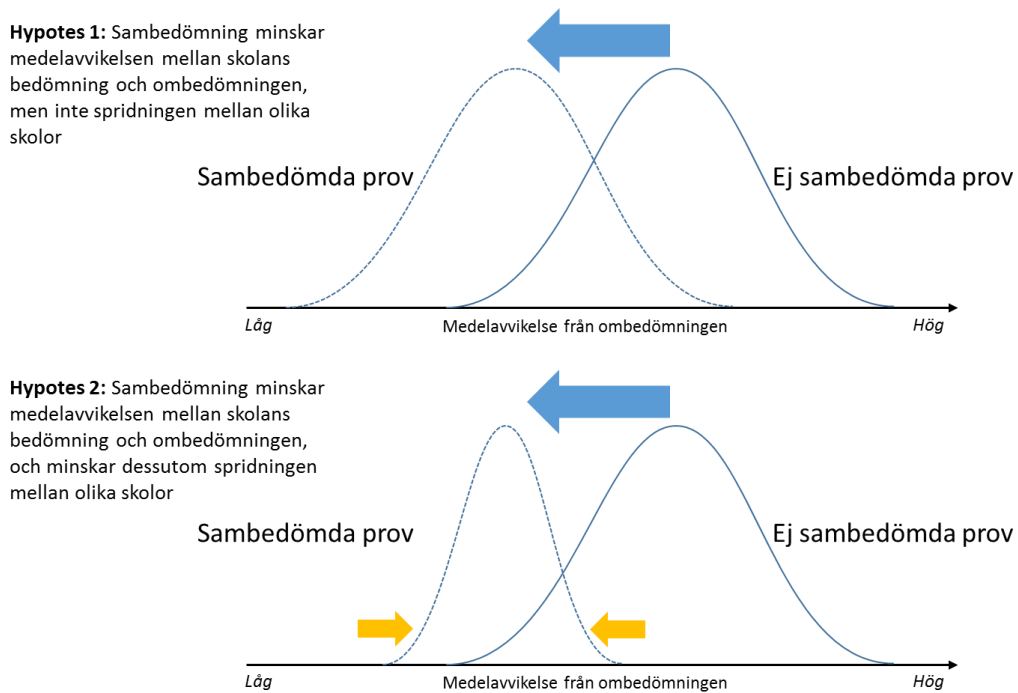
*Signifikant *lägre* genomsnittsavvikelse för avidentifierade prov.

**Signifikant *högre* genomsnittsavvikelse för avidentifierade prov.

Effekten av sambedömning och extern bedömning på skillnader mellan skolor

Sambedömning samt att någon annan än elevens ordinarie ämneslärare bedömer provet tenderar enligt vad vi sett i föregående avsnitt att ha en återhållande effekt på provbetygen. Det leder alltså till något lägre betyg i snitt för motsvarande elevlösning än om ordinarie ämnesläraren gör bedömningen själv och sambedömning inte förekommer. En annan fråga är dock om det också

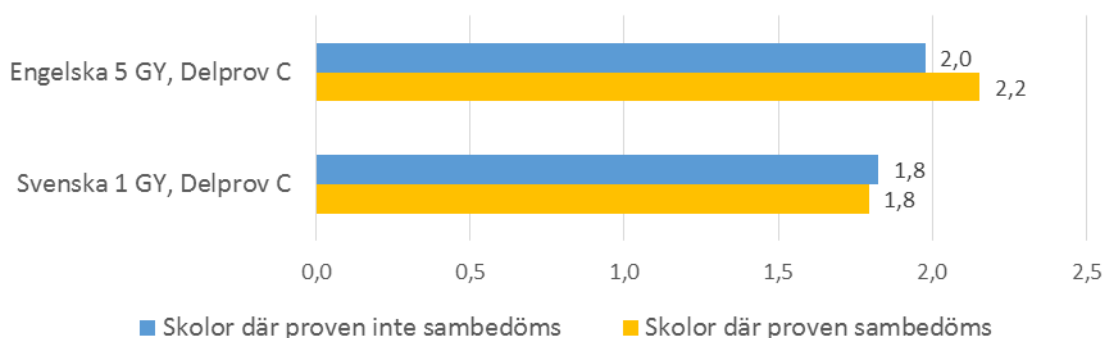
minskar *spridningen* i bedömning mellan olika skolor, eller om det bara är bedömningarnas medelpunkt som flyttas. Ur likvärdighetssynpunkt är det just spridningen mellan skolor som är det intressanta. Var "mittpunkten" i bedömningarna ligger däremot relativt ointressant ur likvärdighetssynpunkt, givet att alla skolor lägger ribban på samma nivå.



Figur 10: Två alternativa hypoteser kring effekten av sambedömning (och extern bedömning)

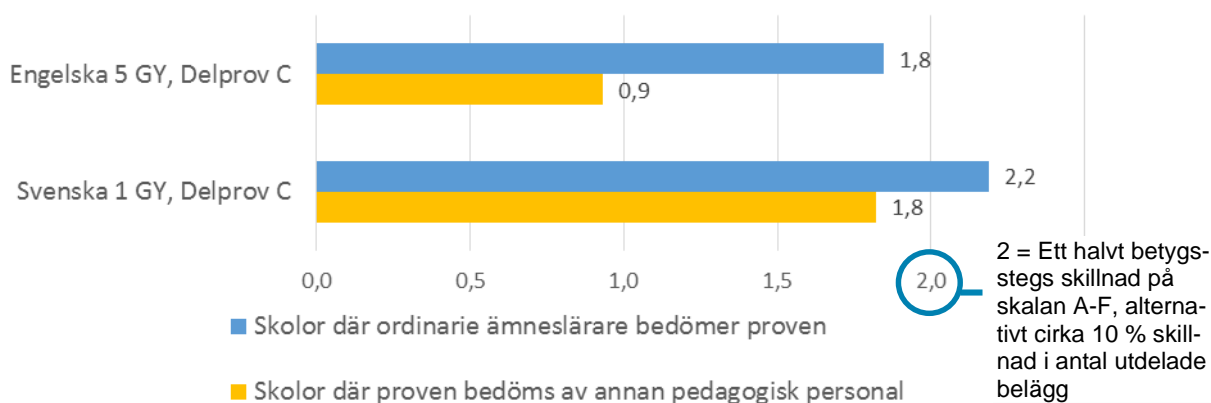
Eftersom denna analys endast kan göras på skolnivå är den bara möjlig att företa för de ämnen där allra flest skolor inkluderats i ombedömningen, nämligen Svenska 1 respektive Engelska 5 på gymnasiet. I övriga fall är antalet skolor i urvalet för litet för att dra några slutsatser. Resultatet redovisas i figur 11 och 12 nedan.

Spridning i medelavvikelse från ombedömningen, p75-p25. Standardiserad skala 0-20



Figur 11: Spridning i medelavvikelse från ombedömningen bland skolor som sambedömer sina prov respektive skolor där sambedömning inte sker. p75-p25 betyder skillnaden mellan den 75:e percentilens och den 25:e percentilens skolor, med avseende på genomsnittlig avvikelse från ombedömningen. 4 på skalan motsvarar ett betygsstegs skillnad på skalan A-F.

Spridning i medelavvikelse från ombedömningen, p75-p25. Standardiserad skala 0-20



Figur 12: Spridning i medelavvikelse från ombedömningen bland skolor där ordinarie ämneslärare bedömer proven respektive skolor där annan pedagogisk personal på eller utanför skolan bedömer proven. p75-p25 betyder skillnaden mellan den 75:e percentilens och den 25:e percentilens skolor, med avseende på genomsnittlig avvikelse från ombedömningen. 4 på skalan motsvarar ett betygsstegs skillnad på skalan A-F.

Resultaten ger en viss indikation om att extern bedömning möjligen kan minska inte bara medelavvikelsen från ombedömningen, utan även spridningen mellan olika skolor. Skillnaden är dock tydlig bara för engelska, och skulle kunna vara slumpmässig. För sambedömning kan ingen effekt alls på

spridningen mellan skolor uppmätas, vare sig för proven i svenska eller engelska. Antalet skolor i analysen är dock begränsat, framförallt avseende skolor där den stora majoriteten av prov bedöms av någon annan än den ordinarie ämnesläraren (24 skolor i studien för Engelska GY och 16 för Svenska GY). Några säkra slutsatser kan därför inte dras från denna analys.

Att vare sig sambedömning eller extern bedömning löser problemet med spridning mellan skolor är på sätt och vis inte särskilt förvånande. En sambedömargrupp har exempelvis goda förutsättningar att *sinsemellan* enas kring uttolkningen av bedömningsanvisningarna, men det hindrar inte att en annan sambedömargrupp någon helt annanstans i landet *sinsemellan* enas om en annan uttolkning.

Eftersom vi däremot sett tydliga tecken på att såväl sambedömning som extern bedömning har en viss återhållande effekt på den genomsnittliga bedömningen av en given elevlösning (det vill säga det tenderar att leda till något lägre betyg i snitt), så finns det en risk att skolors olika tillvägagångssätt vid bedömningen i sig hotar likvärdigheten ur ett elevperspektiv. Elever som går vid skolor som tillämpar extern rättning riskerar att bedömas enligt en något striktare standard än elever vid skolor där den ordinarie ämnesläraren hanterar rättningen.

Skolan, läraren eller provet – vad förklarar avvikelsegraden?

Det faktum att det förekommer skillnader i hur en elevlösning bedöms av olika *sinsemellan* oberoende bedömare säger i sig inget om varför det är så. Flera olika hypoteser är tänkbara, varav vissa delvis kan undersökas utifrån omdömningsresultaten.

Hypotes 1: Systematiska skoleffekter. Denna hypotes innebär att skillnaderna mellan skolor beror på att vissa skolor har *skolövergripande processer* eller en informell *skolkultur* av generös alternativt strikt betygssättning, som färgar av sig på merparten av lärarna. Det finns exempelvis sedan tidigare indikationer på att det är lättare att nå ett högre betyg på skolor där snittnivån är låg, än på skolor där snittnivån är hög.⁶

Hypotes 2: Systematiska lärareffekter. Denna hypotes innebär att vissa *enskilda lärare* är generösa i sin betygssättning, medan vissa är striktare, vilket ger upphov till systematiska skillnader mellan lärare.

⁶ Se t ex Fredriksson, Peter & Vlachos, Jonas (2011), *Reformer och resultat: kommer regeringens utbildningsreformer att ha någon betydelse?* s. 29ff

Hypotes 3: Icke systematiska proveffekter. Denna hypotes innebär att skillnaderna mellan skolor inte är systematiska vare sig på lärar- eller skolnivå, utan utgörs av ett slumpvis "brus" i bedömningen som härrör från provens själva konstruktion (det vill säga uppgifter utan entydigt facit och med bedömningsanvisningar som kräver subjektiv uttolkning).

Det är självklart också tänkbart att en kombination av alla tre effekter förekommer.

Ombedömningen är i grunden inte designad för att kunna besvara frågan om de olika hypotesernas relativa styrka. Till exempel medger inte de tillgängliga register som används att svaren analyseras på lärarnivå, utan endast på skolnivå. Vissa egenskaper hos urvalet gör dock ändå en rudimentär analys möjlig.

För de flesta skolor ombedöms prov från två olika ämnen. Om systematiska skoleffekter spelar en stor roll bör det finnas en korrelation mellan avvikelsegraden i förhållande till ombedömningen för de båda ämnena.

För svenska i årskurs 9 och på gymnasiet ombedömdes två delprov för varje elev (B och C). Samma lärare respektive ombedömare står då typiskt sett för bedömningen av bägge delproven. Båda delproven har ett också ett betydande subjektivt inslag i bedömningen.⁷ Om systematiska lärareffekter spelar en stor roll så bör det finnas en korrelation mellan avvikelsegraden i förhållande till ombedömningen för de båda delproven.

Test av systematiska skoleffekter

För 492 skolenheter finns ombedömningsresultat för två ämnen. För dessa 492 skolor finns en statistiskt signifikant "skoleffekt", definierad som korrelation mellan genomsnittlig avvikelsegrad för det ena ämnet och genomsnittlig avvikelsegrad för det andra. Den uppmätta korrelationen är 0,239.⁸ Är detta mycket? Ett sätt att uttrycka styrkan på sambandet är följande: om en skola ger höga betyg relativt andra skolor för "ämne 1" är det 55 procent sannolikhet att den även ger höga betyg jämfört med andra skolor för "ämne 2". Om en skola å andra sidan ger *låga* betyg jämfört med andra skolor för "ämne 1" är det bara

⁷ Dock inte lika stort. Delprov C består av en stor skrivuppgift, medan delprov B innehåller en blandning av kortare skrivuppgifter och frågor med fasta svarsalternativ.

⁸ Single Measure Intraclass Correlation Coefficient.

44 procent chans att samma skola ger *höga* betyg relativt andra skolor för "ämne 2".⁹ Skoleffekten är alltså tydlig, men inte särskilt stor.

Om sambandsanalysen begränsas till språkproven i svenska och engelska, vilket kan göras i gymnasiet, är sambandet något starkare: korrelationen är här 0,39 mellan de två avvikelsegraderna. Höga betyg relativt andra skolor för provet i engelska innebär 58 procent sannolikhet för höga betyg relativt andra skolor även för provet i svenska, medan *låga* betyg relativt andra skolor på provet i engelska innebär att sannolikhet för *höga* betyg relativt andra skolor för provet i svenska endast är 43 procent.

Beror de uppmätta skoleffekterna på skillnader i bedömningsprocesserna?

En möjlig förklaring till de uppmätta skoleffekterna är skolövergripande skillnader i bedömningsprocesserna, eftersom vi vet att dessa har en effekt på betygssättningen. Till exempel kan det vara så att vissa skolor konsekvent tillämpar sambedömning för alla ämnen, och vissa inte gör det för några.

Antalet skolor som ingår i ombedömningen är för litet för att kunna utreda denna fråga grundligt. Korrelationen mellan de två ämnesresultaten är dock signifikant, och i samma storleksordning som tidigare, även om man kontrollerar för både andelen sambedömda prov och andelen prov rättade av den ordinarie ämnesläraren.¹⁰ Det indikerar att de uppmätta skoleffekterna *inte* i någon större utsträckning kan förklaras av systematiska skillnader i bedömningsprocesser på skolnivå.

Samtidigt är skoleffekten betydligt mindre (0,123), och dessutom inte statistiskt signifikant, om de skolor som tillämpar organiserad sambedömning inkluderas.¹¹ Det skulle indikera att skillnader i bedömningsprocess faktiskt *kan* förklara en del av den uppmätta skoleffekten.

⁹ Andelarna som anges här kan variera något beroende på vilket ämne man väljer att kalla ämne 1 respektive ämne 2 för varje given skola, men ger en bra bild av effektens storleksordning.

¹⁰ Här har vanlig Pearson-korrelation använts för att kunna använda kontrollvariabler. Koefficienten ligger dock nära den i sammanhanget mer korrekta Intraclass Correlation Coefficient som använts ovan.

¹¹ För 80 % eller mer av sina insända prov.

Sammanfattningsvis får konstateras att datamaterialet inte är tillräckligt stort för att nå ett säkert svar på frågan. Eventuellt kan skillnader i bedömningsprocesser mellan skolorna förklara en del av den uppmätta skoleffekten, men de verkar inte kunna förklara hela effekten.

Test av systematiska lärareffekter

För att testa hypotesen om systematiska lärareffekter vore det önskvärt att ha information om vilken lärare på en viss skola som gjort ursprungsbedömningen. För en given skola med flera lärare i samma ämne skulle man då kunna jämföra utfallet visavi ombedömningen per lärare, och på så sätt få ett mått på den enskilde lärarens inflytande på avvikelsegraden. Information om bedömande lärare saknas dock i de betygsregister som finns tillgängliga.

För svenska i årskurs 9 och på gymnasiet ombedömdes dock *två* delprov per elev (delprov B och C).¹² Ett test av systematiska lärareffekter kan därför göras för just detta ämne.

För de 13 138 prov där en avvikelse finns registrerad både för delprov B och delprov C i svenska är korrelationen 0,203 mellan de två avvikelserna (för åk 9 separat är den 0,224 och för gymnasiet 0,188). En positiv avvikelse för delprov B ökar alltså sannolikheten för en positiv avvikelse även för delprov C. Sambandet är statistiskt signifikant.¹³

Samtidigt är det intressant att notera att den uppmätta lärareffekten inte är större än vad den är. Sannolikheten för att ursprungsbedömaren sätter ett högre delprovsbetyg än ombedömaren på delprov C är så hög som 27-29 procent även *givet* att ursprungsbedömaren satt ett *lägre* delprovsbetyg än ombedömaren på delprov B.

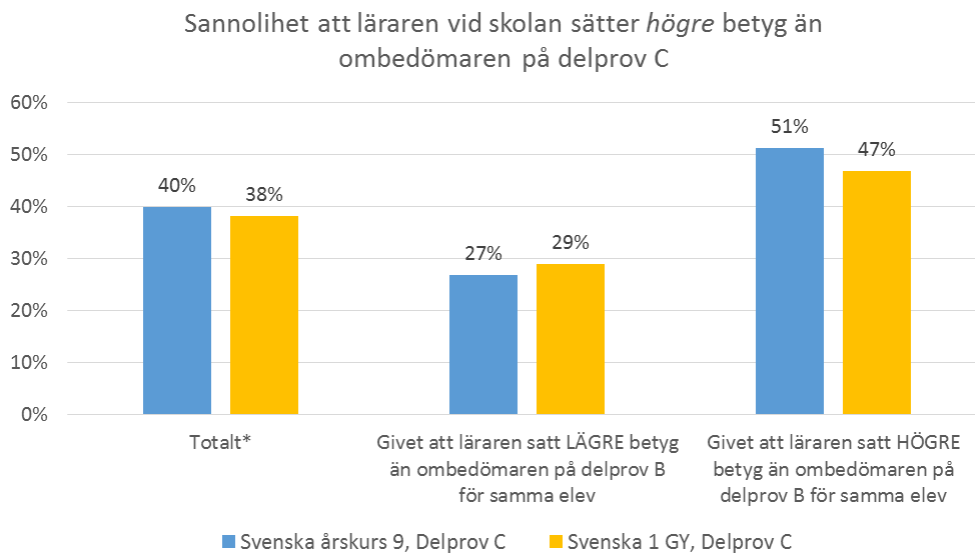
Att en lärare satt ett högre eller lägre delprovsbetyg än ombedömaren på delprov B är alltså långt ifrån någon garanti för att bedömaravvikelsen pekar i samma riktning även för delprov C.

I vilken grad dessa resultat kan generaliseras till fler prov än de aktuella delproven i svenska går inte att säga utifrån ombedömningsresultaten. Det fram-

¹² Resultaten från delprov B är dock inte publicerade på skolnivå.

¹³ Det bör påpekas att den uppmätta lärareffekten emanerar både från systematiska skillnader mellan olika ursprungsbedömare *och* mellan olika ombedömare. Det går inte att bena ut den ena effekten från den andra eftersom samma "bedömarpar" alltid bedömer båda delproven för en viss elev (Delprov B + C).

står dock som rimligt att tänka sig att den latenta lärareffekten inte är dramatisk större eller mindre för just svenskämnet än för något annat ämne. I vilken grad denna effekt sedan kommer till uttryck i bedömningen av ett visst prov beror på utformningen av just det provet.



Figur 13: Sannolikhet att läraren vid skolan sätter högre betyg än ombedömaren på delprov C i Svenska, givet olika bedömaravvikelse för delprov B för samma elev. *Totalsiffran i diagrammet redovisar resultat *exklusive* den andra ombedömning som sker vid stora avvikelser (se Bilaga 1), och skiljer sig därför något från vad som redovisas i figur 1.

Sammanfattning

Resultaten från ombedömningen indikerar att det förekommer både "skoleffekter" (systematiska effekter på skolnivå) och "lärareffekter" (systematiska effekter beroende av enskilda lärare) i bedömningen av proven, men att dessa bara förklarar en viss del av bedömaravvikelsena. Det betyder inte att all återstående variation är osystematisk – flera andra typer av systematiska effekter är tänkbara men kan inte testas inom ramen för detta uppdrag, till exempel systematiska effekter med avseende på lärar- eller elevens egenskaper (ålder, kön, bakgrund och så vidare).¹⁴

Jämförelsen mellan delproven B och C i Svenska för årskurs 9 och Svenska 1 i Gymnasiet indikerar dock att bedömningskillnaderna åtminstone i detta fall till betydande grad förklaras av "icke systematiska proveffekter" (hypotes 3

¹⁴ Vissa av dessa aspekter har dock belysts i tidigare omgångar av ombedömningen, se t ex Skolinspektionen (2012): "Lika för alla: Omrättning av nationella prov i grundskolan och gymnasieskolan under tre år. Resultatbilaga" s. 63 ff

ovan). Vid denna jämförelse är en stor mängd av alla tänkbara systematiska effekter konstanthållna: för ett givet par av delprov är både elev, skola, ursprungsbedömare och ombedömare desamma, och ändå är korrelationen mellan de två avvikelsegraderna inte högre än 0,203. Det tyder på att en betydande del av skillnaderna mellan ursprungsbedömning och ombedömning åtminstone i detta fall varken beror på skolans kultur eller processer, eller på att vissa enskilda lärare konsekvent är "snällare" i sin betygssättning än andra, utan snarare på slumpvisa (i betydelsen osystematiska) skillnader i uttolkningen av provens bedömningsanvisningar.

Rekommendationer utifrån resultaten

Utifrån de resultat som redovisats i denna rapport ger Skolinspektionen följande rekommendationer.

Avseende proven och bedömningsprocesserna

Skolinspektionens rekommendationer avseende proven och bedömningsprocesserna kvarstår i stor utsträckning från tidigare år:

Olika former för sambedömning behöver användas och utprovas i så stor utsträckning som möjligt, som ett stöd för likvärdighet i bedömningarna. Huvudmän kan samverka kring att sambedömargrupperna i så stor utsträckning som möjligt sätts samman av lärare från olika skolor, och gärna också från skolor med olika huvudmän. Detta eftersom sambedömning *inom* skolor endast har begränsad potential att öka likvärdigheten i bedömningarna *mellan* olika skolor, vilket ur ett likvärdighetsperspektiv är lika viktigt som intern samstämmighet. Det är också viktigt att huvudmännen ser till att avsätta tid och resurser för att stärka lärarnas i deras arbete med bedömning och betygssättning.

Skolor och huvudmän behöver generellt sett göra medvetna val i utformningen av sina bedömningsprocesser, eftersom bedömningsprocesserna har en systematisk effekt på betygssättningen. Resultaten från ombedömningen indikerar att inslaget av sambedömning idag varierar kraftigt både mellan skolor och mellan olika årskurser, liksom mellan skolor med kommunal respektive enskild huvudman. Denna "spretighet" på systemnivå riskerar att i sig utgöra ett hot mot likvärdigheten i bedömningarna.

Provutvecklare bör sträva efter att göra bedömningsanvisningarna så lätta som möjligt att uttolka på ett likvärdigt sätt. De delprov där eleven ska ge sitt svar i form av en enda längre sammanhängande text tenderar oavsett ämne att ha en låg bedömaröverensstämmelse. Det är viktigt att hänsyn tas till detta när proven och principerna för sammanvägning till ett samlat provbetyg konstrueras.

Skolinspektionen välkomnar även det arbete som för närvarande bedrivs för att samordna struktur och terminologi mellan de olika ämnesproven. Långsiktigt torde detta ge bättre förutsättningar för systematiska utvärderings- och utvecklingsarbete på systemnivå än vad som idag är fallet.

Avseende kommande ombedömningar

Ombedömningen av nationella prov har över åren bidragit med värdefull kunskap. Skolinspektionen har i ett antal rapporter lyft slutsatser och rekommendationer för ökad likvärdighet i bedömningarna av proven som varit till nytta i

flera olika sammanhang. De huvudsakliga slutsatserna på övergripande nivå har dock varit återkommande år från år.

2013 ändrades därför uppdragsbeskrivningen på Skolinspektionens förslag, så att resultaten bättre skulle kunna användas som underlag i tillsynen på skolnivå. Även detta har gett värdefull kunskap. Ombedömningen är dock ett relativt kostsamt projekt, som också kräver en hel del arbete från skolornas sida med att sammanställa resultat samt kopiera och sända in prov. Trots detta kan bara ett litet stickprov av alla genomförda prov ombedömas.

Skolinspektionens tillämpar även från och med 2015 en ny tillsynsmodell där endast vissa skolor får besök baserat på en riskanalys. På grund av ombedömningens långa ledtider kan urvalet inte helt kopplas till tillsynsplanen, vilket gör att vissa skolor som ombedömts inte får besök, medan resultat från ombedömningen saknas för flera av de skolor som får besök.

Av dessa skäl gör Skolinspektionen bedömningen att en fortsatt årlig ombedömning av nuvarande format riskerar att inte vara kostnadseffektiv, och att ombedömningens övergripande syfte att stödja en likvärdig bedömning och betygssättning av elevers prestationer skulle kunna mötas på ett mer verksamt och kostnadseffektivt sätt genom att en del av de resurser som nu läggs på ombedömningen istället används till andra former av granskningar, samtidigt som möjligheten kvarstår att genomföra mer riktade ombedömningar med utgångspunkt i myndighetens årliga behovsanalys.

Omfattning, urvalsprinciper och metodologiskt upplägg skulle då variera mer över åren och anpassas mer efter identifierade granskningsbehov än vad som hittills varit fallet. Skolinspektionen bedömer att en fortsatt ombedömning på detta sätt skulle kunna stödja myndighetens kärnprocesser bättre än idag, och att det dessutom ger större möjligheter att långsiktigt bygga en allt mera komplett kunskapsbild av området än vad som idag är fallet.

Bilaga 1: Genomförandet av ombedömningen

Urval av prov och skolor

2014 års ombedömning genomfördes för de nationella prov som under året skrevs i årskurs 6 och årskurs 9, under vårterminen på gymnasiets inledande kurser, samt motsvarande kurser inom den kommunala vuxenutbildningen på gymnasial nivå. För proven inom vuxenutbildningen var det dock 2013 års prov som ombedömdes.

Alla skolor ingår inte i ombedömningen ett givet år, utan det görs ett slumpvis urval av skolor bland de huvudmän som det följande året ingår i skolinspektionens regelbundna tillsyn. Det vill säga, urvalet av skolor för ombedömning år 2014 görs bland de skolor som har huvudmän som ingår i tillsynsplanen för 2015, osv. Detta för att så "färska" resultat som möjligt från ombedömningen ska kunna användas i tillsynen.

Ämnen, delprov och årskurser

För att kunna genomföra ombedömningen vid så många skolor som möjligt ombedöms endast ett eller två ämnen per skola, oftast ett språkämne och ett NO/SO-ämne eller matematik. När en skola utvalts till ombedömningen görs en slumpvis tilldelning av ämnen för ombedömning. Därefter görs ett slumpvis urval av elever bland dem som skrivit provet den aktuella terminen. Urvalet görs tillräckligt stort för att resultaten ska vara representativa på skolnivå. Max antal prov att skicka in har dock satts till 100 för grundskolor och 200 för gymnasieskolor. Om fler prov krävs för representativitet för två ämnen dras stället bara ett ämne för den aktuella skolan.

Alla nationella prov är uppdelade i flera delprov. Vissa av delproven innehåller laborativa eller muntliga element, eller uppgifter som ska genomföras i grupp. Denna typ av delprov är inte möjliga att ombedöma. Det är heller inte särskilt intressant att ombedöma delprov som mestadels består av frågor där eleven ska välja mellan ett antal fasta svarsalternativ, eftersom utrymmet för subjektivitet i bedömningen här är mycket litet och tidigare erfarenhet visat att stora avvikelser är mycket sällsynta.

Av dessa skäl ombedömas inte hela proven, utan endast ett eller två delprov per prov. De delprov som väljs ut för ombedömning är sådana som kan ombedömas (det vill säga som inte har laborativa eller muntliga inslag et cetera), men där det inte finns tydliga "rätt och fel svar" på frågorna. I praktiken betyder det att prov som till stor del består av uppsatsfrågor där eleven ska formu-

lera sig fritt i skrift utifrån en öppen frågeställning är de som väljs ut för omdömning. För vissa ämnen, till exempel matematik, finns inga sådana delprov. Då väljs istället det delprov som för det givna ämnet kan antas ha störst utrymme för subjektivitet i bedömningen.

För gymnasiet finns en höstterminsversion och en vårterminsversion av varje prov. Vårterminsversionen skrivs av cirka tre fjärdedelar av eleverna och är den enda som omdömts.

Urvalets sammansättning

Urvalet av skolor för omdömning följer från och med 2014 Skolinspektionens tillsynscykel. Urvalet är alltså inte gjort för att vara representativt på nationell nivå, till exempel avseende region eller huvudmannatyp. Eftersom urvalet följer tillsynscykeln så är det varje givet år bara skolor från ca en tredjedel av landets huvudmän som kan komma ifråga för urval. Urvalet är också draget på ett sådant sätt att andelen omdömda prov från skolor med enskild huvudman är högre än riksgenomsnittet. Framförallt är detta fallet för gymnasieskolan, där 55 procent av omdömda skolor och 45 procent av omdömda prov kommer från skolor med enskild huvudman. Andelen gymnasieskolor med enskild huvudman i riket vid tiden då urvalet drogs uppgick till ca 33 procent.

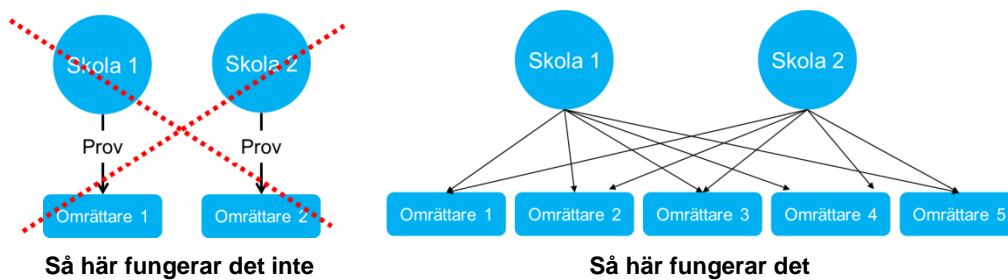
Urvalet exkluderar vidare de allra minsta skolorna. Inga resultat samlas in från skolor där färre än 10 prov för ett visst ämne skrivits, eftersom inga resultat publiceras för skolor där mindre än 10 prov kunnat omdömas.

Alla resultat som presenteras i denna rapport baseras på det faktiska urvalet av omdömda prov, och kan alltså inte med säkerhet generaliseras till riket i stort. Skolinspektionen bedömer dock att urvalets sammansättning och storlek är sådant att det ändå är möjligt och meningsfullt att använda som en indikation om den allmänna situationen i riket.

Insamling av prov och fördelning mellan omdömare

De skolor som väljs ut att delta i omdömningen skickar in kopior av bedömda prov till Skolinspektionen. Proven maskeras och aidentifieras sedan, det vill säga lärarens betyg och anteckningar samt uppgifter om elevens identitet tas bort. Därefter bedöms proven på nytt av yrkesverksamma och erfarna lärare med stor vana att bedöma prov inom sitt ämne som rekryterats av skolinspektionen. Ursprungsbedömningen jämförs sedan med omdömarens bedömning, och resultaten sammanställs per skola och ämne.

Ombedömarna bedömer sina prov enskilt, det vill säga sambedömning tillämpas inte. Däremot sprids varje enskilda skolas elevlösningar över flera ombedömare, för att minimera risken att en enskild ombedömares bedömningar får stort genomslag i de sammantagna resultaten, så att jämförelsen av resultat mellan skolor blir så rättvis som möjligt. Beroende på ämne och årskurs fördelas proven mellan ca 4-40 olika ombedömare. Undantag är Svenska som Andraspråk samt prov skrivna inom Vuxenutbildningen, som endast har 1-3 ombedömare per delprov (detta gäller 2014 års ombedömning).



Figur 14: Schematisk bild över fördelningen av prov till ombedömare. Proven från varje skola fördelas så långt det är möjligt ut till flera olika ombedömare. Fördelningen av prov är utformad på det här sättet för att minimera risken att en enskild ombedömares bedömningar får stort genomslag i de sammantagna resultaten, så att jämförelsen av resultat mellan skolor blir så rättvis som möjligt.

För svenska och engelska finns en särskild kontroll som innebär att om det skiljer mer än två betygsteg mellan bedömning och ombedömning, alternativt om ursprungsbedömaren ger godkänt betyg (A-E) men ombedömaren ger ett F eller vice versa, så går provet vidare till en andra ombedömare. Slutbetyget från ombedömningen blir då den av de två ombedömarens bedömningar som ligger närmast ursprungsbedömningen.

Omfattning och bortfall

Totalt ombedömdes i 2014 års omgång 45 509 delprov från 685 skolor. Ombedömningen genomfördes av totalt 220 lärare, som vardera ombedömde prov inom ett eller två olika ämnen. Fördelningen över delprov framgår av tabellen nedan.

	Antal inbegärda prov	Från antal skolor	Antal inkomna prov	Från antal skolor	Externt bortfall (prov)	Antal kompletta ombeddömda prov	Från antal skolor	Internt bortfall (prov)	Totalt bortfall	Antal ombeddömare
Totalt	55331	710	50150	698	9%	45509	685	9%	18%	220
<i>Delprov:</i>										
B6 A	471	16	443	16	6%	341	13	23%	28%	5
B9 A2	930	23	788	22	15%	682	19	13%	27%	5
E6 C	1947	69	1776	67	9%	1728	66	3%	11%	8
E9 C	7705	135	6740	131	13%	6402	126	5%	17%	24
EG C	9359	201	8841	195	6%	8436	194	5%	10%	34
EV C	316	17	222	15	30%	159	9	28%	50%	2
F6 A	517	17	492	17	5%	267	12	46%	48%	4
F9 A2	729	21	636	20	13%	612	19	4%	16%	5
G6 B	557	18	505	17	9%	431	14	15%	23%	5
G9 A	855	23	763	22	11%	654	21	14%	24%	5
H6 B	552	16	396	15	28%	354	12	11%	36%	5
H9 A	842	24	651	21	23%	576	19	12%	32%	5
K6 A	463	16	426	16	8%	311	13	27%	33%	5
K9 A2	854	24	689	23	19%	612	21	11%	28%	5
M6 D	495	18	457	18	8%	430	17	6%	13%	4
M9 D	786	21	593	19	25%	497	18	16%	37%	6
MGa D	2446	72	2271	71	7%	2133	66	6%	13%	6
MVa D	120	6	98	6	18%	83	4	15%	31%	2
MGb D	2833	60	2582	58	9%	2398	54	7%	15%	9
MVb D	165	10	128	10	22%	46	4	64%	72%	2
MGc D	482	16	407	14	16%	367	13	10%	24%	2
R6 B1+B2	505	17	446	16	12%	333	14	25%	34%	5
R9 B	726	19	618	18	15%	462	15	25%	36%	5
SV6 C1+C2	2320	69	2210	69	5%	2120	69	4%	9%	10

SV9 C	7231	134	6413	126	11%	5992	124	7%	17%	27
SvG C	9419	199	9173	199	3%	8227	193	10%	13%	37
SvV B	602	17	442	17	27%	226	7	49%	62%	2
Sh6 A1+A2	373	16	301	14	19%	161	9	47%	57%	5
Sh9 B	731	22	643	22	12%	469	17	27%	36%	5
---	---	---	---	---	---	---	---	---	---	---
Sv9 B	7231	134	6413	126	11%	5690	123	11%	21%	27
SvG B	9419	199	9173	199	3%	7935	191	13%	16%	37

Resonemang kring bortfallet

Det externa bortfall utgörs av prov som begärts in men inte inkommit från skolorna. En del av detta bortfall är avseende grundskolor beroende av att elevlösningar begärts in från skolorna utifrån föregående års elevlistor. Elever som bytt skola under året utgör alltså en del av det externa bortfallet. Till detta kommer elever som inte skrivit provet på grund av sjukdom eller annan frånvaro. Det förekommer även att skolor inte kan återfinna vissa prov eller bedömningen av proven, till exempel på grund av att ansvarig lärare slutat.

Internt bortfall utgörs av prov som inkommit men inte kunnat analyseras. Exempel på anledningar till det senare är att provkopiorna håller för dålig kvalitet, att sidor saknas, eller att resultatet inte rapporterats in på giltigt format. I det interna bortfallet ingår dessutom alla prov som hör till skolenheter där mindre än 10 prov kunnat ombedömas. Hela skolans resultat exkluderas då från analysen.

Det totala bortfallet varierar kraftigt mellan delprov. Framförallt är det högt för vuxenutbildningen samt för prov i grundskolan där resultatet rapporteras in i form av en summering av antal belägg.

Bortfallsanalyser genomförda i tidigare omgångar av ombedömningen har inte påvisat någon systematik i bortfallet, och det kan antas att detta inte har förändrats då insamlingen av prov genomförts på i stort sett samma sätt över åren.

Bortfallet är dock särskilt stort vad gäller prov skrivna vid den kommunala vuxenutbildningen. Det totala bortfallet är för flera delprov inom vuxenutbildningen mer än 50 procent. Eftersom det totala antalet utvalda skolenheter och antalet insamlade prov dessutom är lågt, bör resultaten avseende VUX-proven tolkas med stor försiktighet. Detsamma gäller ett antal av de NO- och SO-prov,

framförallt Fysik och Samhällskunskap i årskurs 6, där problem med felaktig summering av belägg lett till ett stort internt bortfall.

Frågor om bedömningen

I samband med att proven samlades in från skolorna fick dessa även svara på ett antal frågor om hur bedömningen av provet gick till. Frågorna besvarades för varje enskild insänd elevlösning. Det var dock frivilligt för skolorna att besvara dessa frågor, och att godkänna att de användes i analys och resultatsammanställningar. Det betyder att antal svar som kunnat analyseras är lägre än antalet inkomna prov.

Tabellen nedan visar hur många svar från hur många skolor som inkommit och inkluderats i analysen för respektive fråga. Bortfall anges som 1-(antal svar på frågan/antal kompletta ombedömda prov), där antal svar på frågan betyder antal inkomna prov med korrekt ifyllda uppgifter för den aktuella frågan, där skolan även medgivit att svaren analyseras. Frågorna om bedömningen samlades inte in för proven inom vuxenutbildningen.

Fråga:	1. Vem bedömde provet?			2. Var provet oidentifierat vid bedömningen?			3. Diskuterades bedömningen med annan pedagogisk personal?		
	Antal svar	Från antal skolor	Bortfall	Antal svar	Från antal skolor	Bortfall	Antal svar	Från antal skolor	Bortfall
<i>Delprov:</i>									
B6 A	240	10	30%	320	12	6%	300	11	12%
B9 A2	468	14	31%	465	14	32%	429	13	37%
E6 C	1339	56	23%	1413	57	18%	1410	57	18%
E9 C	3714	85	42%	3919	88	39%	3954	88	38%
EG C	5695	161	32%	5736	161	32%	5730	160	32%
EV C	-	-	-	-	-	-	-	-	-
F6 A	234	11	12%	233	11	13%	234	11	12%
F9 A2	456	14	25%	493	15	19%	502	15	18%
G6 B	296	11	31%	297	11	31%	299	11	31%
G9 A	452	17	31%	448	17	31%	442	17	32%
H6 B	286	10	19%	285	10	19%	284	10	20%
H9 A	338	13	41%	361	14	37%	357	13	38%

SKOLINSPEKTIONEN

2015-11-16
39 (48)

K6 A	237	10	24%	238	10	23%	234	10	25%
K9 A2	472	16	23%	470	16	23%	468	16	24%
M6 D	303	14	30%	365	15	15%	339	14	21%
M9 D	388	16	22%	381	16	23%	392	15	21%
MGa D	1761	60	17%	1793	60	16%	1756	59	18%
MVa D	-	-	-	-	-	-	-	-	-
MGb D	1960	46	18%	1946	46	19%	1966	45	18%
MVb D	-	-	-	-	-	-	-	-	-
MGc D	211	10	43%	210	10	43%	210	10	43%
R6 B1+B2	145	6	56%	210	9	37%	211	9	37%
R9 B	409	14	11%	409	14	11%	408	14	12%
SV6 C1+C2	1281	46	40%	1425	49	33%	1401	49	34%
SV9 C	3965	92	34%	4401	99	27%	4336	96	28%
SvG C	5188	146	37%	5656	155	31%	5638	156	31%
SvV B	-	-	-	-	-	-	-	-	-
Sh6 A1+A2	126	7	22%	126	7	22%	125	7	22%
Sh9 B	336	14	28%	332	14	29%	327	14	30%
---	---	---	---	---	---	---	---	---	---
Sv9 B	3965	92	30%	4401	99	23%	4336	96	24%
SvG B	5188	146	35%	5656	155	29%	5638	156	29%

Bilaga 2: Ett harmoniserat mått för att mäta skillnader i bedömning

Ett nationellt prov består av flera *delprov*, typiskt sett 2-5 stycken. Enligt vad som beskrivs i avsnittet *Urval av prov och skolor* i Bilaga 1 så ombedöms inte alla delprov, utan endast de skriftliga prov där risken för avvikelser kan antas vara störst. Det gör att det samlade provbetyget inte kan användas som resultatmått, utan jämförelsen måste avse endast det aktuella delprovet.

För språkproven sätts delprovsbetyg på skalan A-F, men för alla övriga prov (NO, SO, Matematik) är det endast det samlade provbetyget som anges på skalan A-F. För de olika delproven erhålls istället ett antal *poäng* på E-, C- respektive A-nivå (för NO- och SO-proven är *belägg* det ord som används). Hur många poäng/belägg som kan uppnås på respektive nivå varierar mellan olika ämnen och delprov.

För att kunna göra en översiktlig sammanställning av de ombedömningens resultat för de vitt skilda proven behöver de olika bedömningarna därför på ett eller annat sätt konverteras till en gemensam skala.

Denna har konstruerats på så sätt att bedömningarna av samtliga delprov kan uttryckas på skalan 0-20, där 0 representerar lägsta möjligt bedömning och 20 representerar högsta möjliga bedömning. Skillnaden mellan bedömning och ombedömning kan då på motsvarande sätt uttryckas på skalan -20 till 20. Ett positivt värde indikerar att skolan satt ett högre betyg (språkprov) eller gett fler belägg (SO/NO/Ma) än ombedömaren för en viss elevlösning, och negativa värden indikerar motsatsen.

Beräkning av det harmoniserade resultatmättet

För de delprov där bokstavsbedömning A-F sätts är konverteringen mellan den faktiska bedömningen och den harmoniserade skalan rättfram: A=20, B=16, C=12, D=8, E=4, F=0. En genomsnittsskillnad på 2 mellan bedömning och ombedömning för en viss skola betyder alltså i dessa fall att det i snitt skiljer "ett halvt betygssteg" mellan skolans och ombedömarens bedömning, där skolan i snitt sätter högre betyg.

För de delprov där delprovsbetyg inte sätts omvandlas i stället antalet belägg (poäng) på E- C- respektive A-nivå till skalan 0-20 enligt följande modell:

$$((E*10 + C*15 + A*20) / (\max(E)*10 + \max(C)*15 + \max(A)*20)) * 20$$

Där E, C respektive A står för det antal belägg som skolan respektive ombedö-
maren angivit för delprovet, och $\max(E)$, $\max(C)$ respektive $\max(A)$ står för
det maximala antal belägg som kan uppnås för det aktuella delprovet.

Resultatet av denna beräkning är att även de bedömningar som yttrycks i be-
lägg kan standardiseras till ett enskilt indexvärde på skalan 0-20, där A-belägg
väger dubbelt så tungt som E-belägg i sammanräkningen. Ju färre antal belägg
som utgör maximalt antal belägg på ett delprov, desto större inflytande får
varje enskilt belägg i framräkningen av det harmoniserade måttet.

En snabb approximation som underlättar den intuitiva förståelsen är att 2 på
den harmoniserade skalan motsvarar 10 procent av antalet möjliga belägg för
det aktuella delprovet, 10 motsvarar 50 procent av antal möjliga belägg osv.
Denna approximation tar dock inte hänsyn till att A-, C- och E-belägg viktas
olika i beräkningen.

Det är viktigt att komma ihåg att det harmoniserade resultatmåttet avseende de
prov där det bygger på en sammanräkning av belägg endast är ett grovt sett
sätt att på en gemensam skala ange om en skillnad i bedömning är "stor" eller
"liten". Det finns dock ingen enkel korrespondens mellan detta mått och ett
faktiskt betyg på skalan A-F, eftersom reglerna för hur belägg summeras till ett
samlat provbetyg ser olika ut för olika prov, och dessutom involverar summe-
ring av belägg från flera olika delprov.

Bilaga 3: Metodologisk not avseende beräkning av variationen mellan skolor

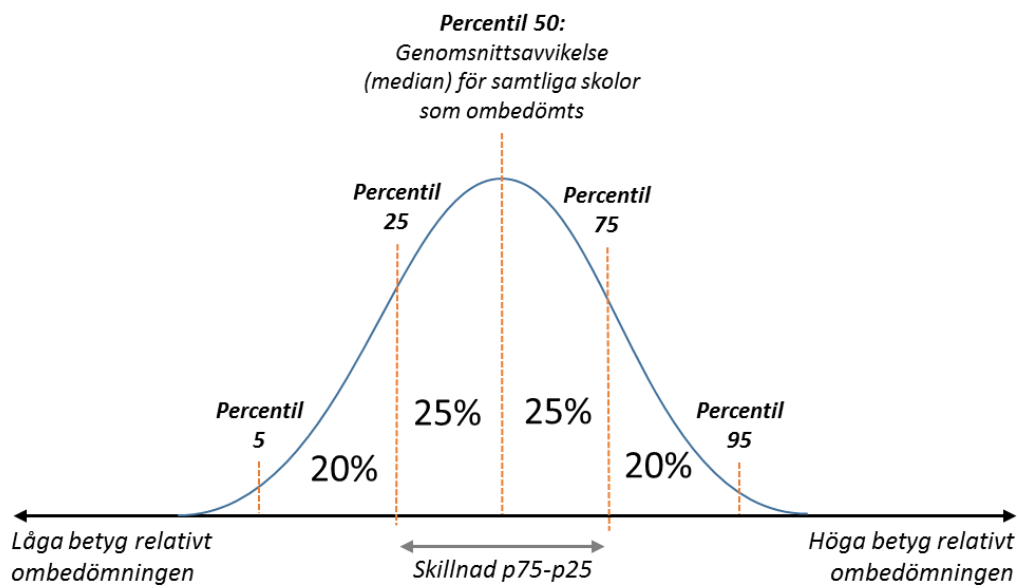
Spridningen mellan olika skolor i bedömningen av samma elevlösning kan inte observeras direkt utifrån ombedömningen, eftersom alla skolor endast bedömer sina egna elevers prov. Resultatet från ombedömningen kan dock användas som en gemensam referenspunkt för att göra en indirekt approximation av skillnaderna mellan skolor. Eftersom proven från varje skola slumpas ut till samma grupp av ombedömare, så borde skolorna om de bedömde likvärdiga elevlösningar på exakt samma sätt också ha ungefär samma medelavvikelse från ombedömningen. Spridningen med avseende på genomsnittsavvikelsen från ombedömningen ger alltså en indikation om hur stora skillnaderna mellan olika skolor skulle vara *om de hade bedömt samma prov*.

Måttet är endast approximativt eftersom sammansättningen av gruppen ombedömare inte ser exakt likadan ut för alla skolor, vilket kan påverka jämförelsen.¹⁵ Skolinspektionen bedömer dock att det ger en tillförlitlig bild av *storleksordningen* på skillnaderna mellan olika skolor. Mindre skillnader mellan olika delprov bör dock tolkas försiktigt. Resultaten redovisas därför endast sammanlagt per ämnesgrupp.

För denna analys har även betygen på språkproven räknats om till skalan 0-20, enligt principen F=0, E=4, D=8, C=12, B=16, A=20. För språkproven är alltså konverteringen mellan den faktiska bedömningen och den harmoniserade skalan rättfram. En skillnad på 4 mellan två bedömningar betyder motsvarar ett betygsstegs skillnad, 8 motsvarar två betygssteg och så vidare.

I analysen jämförs olika percentiler av skolor, med avseende på genomsnittlig avvikelse från ombedömningen. Den hundradel av skolorna som sätter de lägsta betygen i förhållande till ombedömningen utgör percentil 1. På motsvarande sätt utgörs percentil 25 av de skolor som relativt ombedömningen sätter lägre betyg än 75 procent av skolorna, och högre betyg än 25 procent. Figur 15 ger en schematisk illustration av jämförelsen mellan två percentilgrupper.

¹⁵ Se metodavsnittet i Bilaga 2 för en utförligare redogörelse för hur prov från olika skolor fördelas mellan olika ombedömare.



Figur 15: Schematisk illustration av jämförelsen mellan olika percentiler av skolor.

Bilaga 4: Bedömaröverensstämmelse mellan olika omdedömare

I utvärderingssyfte inkluderades för 14 av de omdedömda delproven även så kallade "ankarprov" i omdedöningen. Ankarproven utgörs av 10 autentiska elevlösningar per delprov, utvalda av respektive provkonstruerande institution för att representera en god spridning med avseende på bedömd kvalitet. Proven levereras till Skolinspektionen tillsammans med den provkonstruerande institutionens bedömning av elevlösningen ifråga. Varje institution har sina egna processer för att komma fram till denna bedömning, men typiskt sett sker det i dialog mellan provutvecklarna och en sambedömargrupp bestående av erfarna lärare.

Varje ankarprov distribuerades sedan ut till samtliga omdedömare inom det aktuella ämnet som en del av den ordinarie omdedöningen. Omdedöarna vet inte om vilka prov som är ankarprov, och gör oberoende av varandra sina respektive bedömningar av samma elevlösning. På så vis ges möjlighet att utvärdera spridningen i bedömning av samma elevlösning mellan olika omdedömare.

Skillnader mellan olika omdedömare

Omdedömargruppen för ett visst delprov består typiskt sett av mellan 4 och 40 omdedömare, beroende på hur många prov som omdedöms (Se Bilaga 1). Omdedöarna är erfarna och yrkesverksamma lärare inom sitt ämne. Inte desto mindre är det rimligt att tänka sig att olika omdedömare kan skilja sig åt i sin bedömning av samma elevlösning.

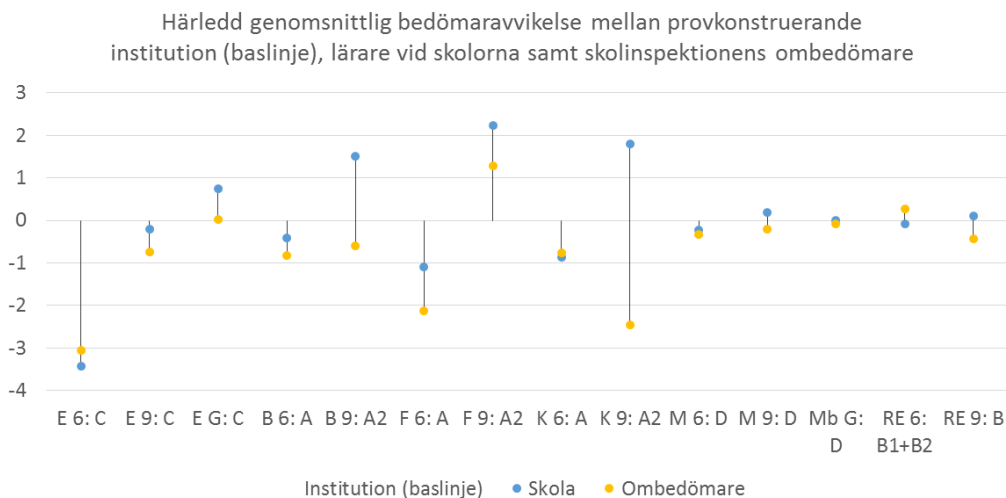
Analysen visar att avvikelserna mellan olika omdedömare är i samma storleksordning som skillnaderna mellan skolor. Sett endast till skolor där färre än 20 elever skrivit provet, det vill säga skolor där alla insända prov typiskt sett bedömts av samma lärare, är variationen mellan olika skolor i stort sett exakt lika stor som motsvarande skillnader mellan olika omdedömare. Det är intuitivt rimligt: i de fall en skola bara har en lärare för ett visst ämne är en jämförelse mellan skolor i princip detsamma som en jämförelse mellan enskilda lärare. I det fall en skola har fler än en lärare för ett visst ämne (det vanliga) så bidrar det till att jämna ut skillnaderna mellan skolor eftersom skolorvalet dras slumpvis från alla lärare på skolan.

Resultaten från ankarproven indikerar alltså att bedömaröverensstämmelsen mellan olika omdedömare varken är substantiellt högre eller lägre än mellan lärare i gemen. Det faktum att proven från respektive skola slumpas ut över alla

tillgängliga ombedömare (se Bilaga 1) är alltså centralt för att jämförelsen av resultat mellan olika skolor ska bli rättvisande.

Skillnader i genomsnittsbedömning skola – ombedömare – institution

Med samma metodik som använts för att skatta skillnaden mellan olika skolor, ger ankarproven även möjlighet att indirekt skatta relationen "Skola (A) – Ombedömare (B) – Institution (C)" med avseende på genomsnittsavvikelse mellan de tre grupperna. Eftersom vi har måttet A-B från den ordinarie ombedömningen och måttet B-C från ankarproven, kan även måttet A-C räknas fram.¹⁶ Man ska dock vara medveten om att resultatet endast är en skattning, eftersom skillnaden B-C mäts utifrån ankarproven, och skillnaden A-B utifrån de prov som sänds in från skolorna i den ordinarie ombedömningen. Resultatet av skattningen redovisas i figur 16 nedan, där institutionens bedömning satts som baslinje (0) och de övriga gruppernas genomsnittsavvikelse gentemot denna baslinje visas.



Figur 16: Härledd genomsnittlig bedömaravvikelse mellan provkonstruerande institution (baslinje), lärare vid skolorna samt Skolinspektionens ombedömare.

Här bör åter nämnas att de ombedömda delproven inte bara skiljer sig åt med avseende på ämne och årskurs, utan också i hög grad med avseende på typ av frågor, bedömningsskalor mm. Meningsfulla jämförelser av resultat mellan

¹⁶ $(A-B)+(B-C)=A-C$. När det i detta avsnitt görs jämförelser mellan skolornas och institutionernas bedömningar (A-C) så handlar det alltså inte om en direkt jämförelse (någon sådan finns inte att tillgå), utan om en härledd sådan.

olika prov kräver en fördjupad analys av respektive provs konstruktion, som inte företas i denna rapport. Några intressanta observationer av mer allmän karaktär kan dock göras.

”Ordningen” mellan skola, ombedömare och institution varierar. För vissa delprov gör institutionen de mest ”generösa” bedömningarna i snitt, för andra de mest ”strikt”. De genomsnittliga avvikelserna mellan de olika grupperna är typiskt sett små, motsvarande +/- 1 på den harmoniserade skalan, det vill säga cirka en fjärdedels betygssteg för delprov med bokstavsbedömning A-F, eller ca 5 procent av totalt antal belägg för övriga prov.

För skrivuppgiften i Engelska i årskurs 6 gör både skolor och ombedömare i medeltal en striktare bedömning av proven än institutionens ankarvärden. Mönstret går inte igen för skrivuppgiften i årskurs 9 och på gymnasiet. För Kemiprovet i årskurs 9 gör ombedömarna en betydligt striktare bedömning än institutionen, som i sin tur gör en betydligt striktare bedömning än skolorna: resultatet blir det största totala spannet mellan de tre grupperna för de 14 aktuella delproven, 4,3 på den harmoniserade skalan, eller ca 20 procent av totalt antal möjliga belägg. Skolorna gör en markant generösare bedömning än institutionen för samtliga NO-prov i årskurs 9, en tendens som inte går igen i årskurs 6.

Överhuvudtaget ses en tendens för samtliga ämnen, utom för Matematik i gymnasiet, att ju högre årskurs, desto generösare bedömning av skolan kontra institutionen. En möjlig tolkning av det är att institutionernas bedömningar utgår från en kraftigare progression mellan årskurser än vad skolornas bedömningar ger uttryck för.

Bilaga 5: Användning av resultaten i tillsynen

Enligt uppdraget ska Skolinspektionen använda resultaten från omdömningen som ett underlag i tillsynen av skolor och huvudmän. Myndigheten ska "utifrån indikationer på betydande avvikelser när det gäller dels rättning av nationella prov, dels betygssättning i ämnen och kurser, inom ramen för sin tillsynsverksamhet vidta lämpliga åtgärder."

Det övergripande syftet med att analysera dessa bedömningsindikatorer är att betygssättningen ska vara rättvis och likvärdig. Det är dock alltid processen som leder fram till resultatet som granskas i skolinspektionens tillsyn, och inte resultaten/provbetygen i sig, eftersom det inte finns några lagstadgade eller på annat sätt reglerade gränser för hur stora avvikelserna kan eller bör vara. En hög avvikelsegrad i förhållande till andra skolor kan dock vara en indikation på att skolan behöver utveckla sina processer för bedömning och betygssättning, och att detta är ett område som bör ägnas extra uppmärksamhet i tillsynen.

Betydande avvikelser på skolnivå

Av de 685 skolor för vilka omdömningsresultat publicerats för ett eller två ämnen i 2014 års omdömning, var det 204 stycken (30 procent) som fick markering för betydande avvikelse i ett eller flera ämnen.

Vad som ska anses vara en "betydande avvikelse" är en bedömningsfråga. Eftersom inget entydigt definierat rätt svar kan anges för de frisvarsbaserade delprov som framförallt är föremål för omdömning, och bedömningen därför även kan variera mellan olika omdömare, så har Skolinspektionen valt att sätta ribban för vad som ska anses vara en betydande avvikelse relativt högt. Det innebär för språkproven en skillnad på *två betygssteg eller mer* mellan skolans bedömning och omdömningen. För de delprov där bedömningen sker i form av belägg är gränsen satt till +/- 6 i skillnad på den harmoniserade skalan 0-20. Det motsvarar en skillnad på approximativt 30 procent eller mer i antal utdelade belägg i de två bedömningarna.¹⁷

För det andra så ska de betydande avvikelserna i den ena eller andra riktningen vara mer vanligt förekommande vid den aktuella skolan än för genomsnittet av de omdömda skolorna. Eftersom omdömningens syfte är likvärdighet i bedömningen *mellan skolor* är det en enskilds skolas avvikelse från andra skolor,

¹⁷ I sammanvägningen väger dock A-belägg tyngre än C-belägg, som i sin tur väger tyngre än E-belägg. Beräkningen beskrivs i Bilaga 2.

inte från omdömningen i sig, som ska minimeras. Om en skolas andel betydande avvikelser i antingen positiv eller negativ riktning skiljer sig med mer än 10 procentenheter från genomsnittet för de omdömda skolorna för det aktuella delprovet, så anges det därför som en signal till tillsynsverksamheten att särskilt granska skolans processer för bedömning och betygssättning vid nästkommande tillsynsbesök.

Sammanfattningsvis ska alltså en enskilds skolas avvikelser i förhållande till omdömningen vara både *stora* och mer *frekventa* än genomsnittligt för att en signal ska ges till tillsynsverksamheten.

Resultaten från omdömningen är ett av de underlag som Skolinspektionens inspektörer granskar inför tillsyn. Skolor med hög avvikelsegrad i förhållande till andra skolor får redogöra för sina processer kring bedömning och betygssättning, och hur man arbetar för att säkerställa likvärdighet i bedömningarna. Om skolans arbetsätt inte uppfyller kraven i lag och förordning så föreläggs skolan att åtgärda bristerna.

Publicering av resultat på skol- och huvudmannanivå

Rapporter på skol- och huvudmannanivå för 2014 års omdömning finns publicerade i skoldatabasen Siris.¹⁸ Materialet på Siris, som är det som utgör underlag i tillsynen, redovisas *terminsvis* utifrån myndighetens tillsynscykel (HT + VT).

Resultaten som redovisas i föreliggande rapport avser de *sammanslagna* resultaten från hela 2014 års omdömning. De totaler som redovisas i den här rapporten skiljer sig därför från de som publicerats i de terminsvisa skol- och huvudmannarapporter som används i tillsynen.

¹⁸ <http://siris.skolverket.se/> > Tillsyn > Omdömning av nationella prov