

Bedömningsprocessernas betydelse för likvärdigheten

Ombedömning av nationella prov 2016

Redovisning av regeringsuppdrag Dnr U2014/7535/GV

Innehåll

FÖRORD	4
SAMMANFATTNING	5
UPPDRAGET	7
BAKGRUND, SYFTE OCH GENOMFÖRANDE.....	7
DENNA RAPPORT	8
PUBLICERING AV RESULTAT PÅ SKOL- OCH HUVUDMANNANIVÅ.....	8
RESULTAT	9
OMBEDÖMDA DELPROV.....	9
URVALET FÖLJER TILLSYNSCYKELN.....	9
BEDÖMNINGSPROCESSEN	9
<i>Vanligast att elevens ordinarie lärare bedömer provet</i>	10
<i>Läraren känner vanligen till elevens identitet vid bedömning</i>	11
<i>Sambedömning tillämpas i närmare tre fjärdedelar av fallen</i>	12
<i>Sambedömning är vanligare på kommunala skolor</i>	13
<i>Vanligare att elevens ordinarie lärare bedömer provet på fristående skolor</i>	14
<i>Avidentifiering lika vanligt i skolor med kommunal som enskild huvudman</i>	15
<i>Större andel avidentifierade delprov i engelska och svenska</i>	15
<i>Bedömningsprocessen kan ha stor betydelse för utfallet</i>	16
<i>Sambedömning, extern bedömning och avidentifiering minskar avvikelsegraden</i>	18
HUR MYCKET SKILJER SIG BEDÖMNINGARNA AV SAMMA EVELÖSNING?	22
<i>Olika bedömningar för mer än varannan elevlösning</i>	22
<i>Något större bedömningsskillnader i svenska än i engelska</i>	22
<i>Störst andel avvikelser i fysik</i>	23
HAR RESULTATET FÖRÄNDRATS JÄMFÖRT MED TIDIGARE OMBEDÖMNINGAR?	24
<i>Små uppmätta skillnader i svenska mellan omgång 6, 7 och 8</i>	24
<i>Andelen avvikelser i engelska överensstämmer i hög grad med omgång 7</i>	25
<i>Små skillnader mellan åren för NO-ämnena</i>	26
STOR VARIATION MELLAN OLIKA SKOLOR FÖR SAMMA EVELÖSNING.....	27
<i>Kan skolan förklara avvikelsegraden?</i>	28
<i>Små avvikelser mellan kommunala och fristående skolor</i>	30
<i>Avvikelserna kan ha en avgörande betydelse för det slutgiltiga provbetyget</i>	30
FAKTORER SOM PÅVERKAR SAMSTÄMMIGHETEN I BEDÖMNINGEN	32
SAMBEDÖMNING, EXTERN BEDÖMNING OCH AVIDENTIFIERING ÖKAR LIKVÄRDIGHETEN ..	32
LIKLANDE PROCESSER OCH SAMVERKAN ÖKAR SAMSTÄMMIGHETEN I BEDÖMNINGEN	32
AVSLUTNINGSVIS.....	33
BILAGA 1: ANVÄNDNING AV RESULTATEN I TILLSYN OCH	
KVALITETSGRANSKNING	34
HUR AVGÖRS VAD SOM ÄR BETYDANDE AVVIKELSER?	34

HUR ANVÄNDS RESULTATEN I SKOLINSPEKTIONENS TILLSYN OCH KVALITETSGRANSKNING?	35
.....	
SKOL- OCH HUVUDMANNARAPPORTER.....	36
BILAGA 2: GENOMFÖRANDET AV OMBEDÖMNINGEN	37
URVAL AV PROV OCH SKOLOR	37
<i>Ämnen, delprov och årskurser som ingår i omdömmen</i>	37
<i>Hela provet omdöms inte, endast längre skrivuppgifter</i>	37
<i>Urvalets sammansättning</i>	38
TERMINOLOGISK ÖVERSIKT	38
FÖRDELNING AV PROV MELLAN OMBEDÖMARE	39
OMFATTNING OCH BORTFALL.....	40
<i>Bortfall</i>	41
<i>Vad beror bortfallet på?</i>	41
<i>Frågor om bedömningen på skolan</i>	41
BILAGA 3: ETT HARMONISERAT MÅTT FÖR ATT MÄTA SKILLNADER I BEDÖMNING.....	43
BERÄKNING AV DET HARMONISERADE RESULTATMÅTTET	43
BILAGA 4: METODOLOGISK NOT AVSEENDE BERÄKNING AV VARIATIONEN MELLAN SKOLOR.....	45
BILAGA 5: REFERENSFÖRTECKNING	47

Förord

Skolinspektionen 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 åttonde årets centrala rättning. Denna rapport avser nationella prov i grund- och gymnasieskolan som genomfördes under 2016 och nationella prov inom den kommunala vuxenutbildningen på gymnasial nivå som genomfördes under 2015.

Vid Skolinspektionen har följande personer arbetat med att genomföra uppdraget. Projektledare: Klara Staffansdotter Klingspor, Jenny Börlin och Fanny Hemmingsson. Utredare/statistiker: Mattias Friström, Peter Gustafsson, Björn Ljung, Jacob Kurkiala, Arjann Akbari och Johan Granath. Avdelningschef: Anna Sellin. Enhetschef: Jonna Andersson.

Sammanfattning

Den här rapporten redovisar resultaten från 2016 års ombedömning av nationella prov, den åttonde i ordningen sedan 2009. Syftet med ombedömningen är att stödja en likvärdig bedömning och betygssättning av proven över landet och därmed befästa elevens rätt till en likvärdig betygssättning.

I omgång 8 av ombedömningen ingår nationella prov i svenska och engelska för grundskolans årskurs 6 och 9, samt prov i biologi, fysik och kemi (NO) för årskurs 9. Vidare innefattar ombedömningen prov i svenska och engelska i gymnasieskolans grundläggande kurser och motsvarande kurser inom den kommunala vuxenutbildningen på gymnasial nivå.

Fokus i årets rapport är bedömningsprocessernas betydelse för likvärdigheten i bedömningen av nationella prov. Slutsatserna i ombedömningen av nationella prov 2016 bekräftar tidigare ombedömningsomgångar.

Bedömningarna skiljer sig åt för mer än varannan elevlösning

I 41 procent av bedömningarna har skolans lärare satt ett delprovsbetyg som är minst ett betygssteg *högre* än ombedömaren och 21 procent av delproven har fått ett delprovsbetyg av skolan som är minst ett betygssteg *lägre*. Av de ombedömda delproven är andelen avvikelser störst bland NO-proven. Resultaten är i linje med tidigare omgångar.

Små förändringar i resultatet jämfört med tidigare ombedömningar

Resultaten från den åttonde ombedömningen samstämmer på det stora hela med tidigare jämförbara ombedömningar, omgång 6 och 7. I engelska och NO-ämnena är resultaten i paritet med tidigare omgångar för alla årskurser, likaså för delprov genomförda i svenska årskurs 6 och på den kommunala vuxenutbildningen. I svenska i årskurs 9 och i gymnasieskolan har andelen elevlösningar som fått ett högre betyg av skolans lärare jämfört med ombedömaren ökat något mellan omgång 7 och 8.

Fortfarande vanligast att elevens ordinarie lärare bedömer provet men sambedömning ökar

Vid bedömning av nationella prov tillämpar skolorna olika processer. Detta betyder att bedömningen kan ske på olika sätt på olika skolor. I likhet med tidigare omgångar är det vanligast att elevens ordinarie lärare bedömer proven. Sambedömning av prov kan ske genom informella samtal mellan lärare men också som en organiserad process. Andelen delprov som sambedömts inom

ramen för en organiserad process har ökat något denna omgång. De flesta delproven var inte avidentifierade, d.v.s. det var vanligast att läraren kände till elevens identitet.

Sambedömning, extern bedömning och avidentifiering minskar den genomsnittliga avvikelsen

Skillnader mellan skolan och de ombedömmande lärarnas bedömningar mäts i genomsnittliga avvikelser. En mindre genomsnittlig avvikelse innebär att skolans lärare och ombedömarna har gjort en liknande bedömning.

Resultaten från omgång 8 befäster resultaten från tidigare ombedömningar: sambedömning och extern bedömning (att någon annan än elevens ordinarie lärare bedömer provet) minskar de genomsnittliga avvikelserna och bidrar till en likvärdig bedömning av nationella prov. Även avidentifierade prov tycks ha en återhållande effekt på avvikelsegraden.

Faktorer som påverkar likvärdigheten

Bedömningsprocesserna har en systematisk effekt på betygssättningen av de nationella proven. Skolornas olika bedömningsprocesser – om provet är avidentifierat, om provet är sambedömt eller om det är elevens lärare eller en extern bedömare som har rättat provet – påverkar likvärdigheten inom betygssättningen. En mer samstämmig bedömningsprocess tillsammans med en ökad samverkan mellan skolor, inom och mellan olika huvudmän, ger större möjligheter för ökad likvärdighet i bedömningen av nationella prov.

Uppdraget

Bakgrund, syfte och genomförande

Skolinspektionen har sedan 2009 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.

Skolinspektionen samlar in kopior av genomförda nationella prov från ett urval av skolor. Proven avidentifieras och bedöms därefter på nytt av erfarna lärare som rekryterats av Skolinspektionen. Därefter jämförs de ursprungliga bedömningarna med ombedömningarna av elevsvaren. Resultatet används sedan av Skolinspektionens inspektörer som ett av flera underlag inför tillsyn av skolor. En utförligare beskrivning av hur resultaten från ombedömningen används i inspektionen finns i bilaga 1.

Bakgrunden till att Skolinspektionen fått i uppdrag att göra en central rättning av nationella prov är att det finns en risk för att bedömningarna av de nationella proven inte är likvärdiga mellan olika lärare och skolor. Bland de delprov där elevens svar ges i form av en text eller ett mindre antal längre texter är risken för skillnader mellan olika bedömningar störst. Tidigare års ombedömningar har visat att det bland dessa delprov är vanligt att det skiljer ett betygsslag eller mer mellan två olika lärares bedömning av samma elevsvar. Delprovsresultatet/-betyget påverkar i sin tur provbetyget. Provbetygen på nationella prov påverkar i nästa led betyget på ämnet eller kursen. Med tanke på betygets 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.

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 problematisk, eftersom ett av provens huvudsyften är att 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 resultat från ombedömningen av de nationella proven med andra skolors resultat, och att Skolinspektionen följer upp dessa indikatorer i sin tillsynsverksamhet.

Denna rapport

Skolinspektionen redovisar med denna rapport resultatet av åttonde årets centrala rättning av nationella prov. I rapporten presenteras både deskriptiva resultat och ett antal statistiska analyser med bäring på frågan om likvärdig bedömning av proven. Den detaljerade redogörelsen för projektets genomförande, urvalet av skolor och delprov samt beräkningstekniska detaljer har sammanställts i bilagor. Den läsare som vill skaffa sig en djupare förståelse för resultaten uppmanas att läsa bilagorna.

Publicering av resultat på skol- och huvudmannanivå

Rapporter på skol- och huvudmannanivå för omdömning av nationella prov från år 2016 finns publicerade i skoldatabasen Siris.¹ Materialet på Siris, som är det som utgör underlag i tillsynen, kan användas av rektorer och huvudmän på de skolor som har ingått i undersökningen i syfte att analysera och utveckla arbetet med bedömning och betygssättning.

Resultaten som redovisas i denna rapport avser de *sammanslagna* resultaten från hela 2016 års omdömning. De totaler som redovisas i den här rapporten skiljer sig något från de som publicerats i skol- och huvudmannarapporterna. Det beror på att skol- och huvudmannarapporterna inte innehåller resultat för skolenheter där antalet omdömda elevlösningar för aktuellt delprov understiger tio till antalet eller där färre än hälften av de prov som skrevs vid skolenheten för det aktuella delprovet och terminen har omdömts.

¹ <http://siris.skolverket.se/siris/ris.omrattning.searchForm>

Resultat

Ombedömda delprov

I 2016 års ombedömning ingår prov i svenska och engelska för grundskolans årskurs 6 och 9, samt prov i biologi, fysik och kemi (NO) för årskurs 9. Vidare innefattar ombedömningen prov i svenska och engelska i gymnasieskolans grundläggande kurser och motsvarande kurser inom den kommunala vuxenutbildningen på gymnasial nivå.

I ämnena engelska, biologi, fysik och kemi ombedöms endast *ett delprov*. Det samma gäller för ämnet svenska, med undantag för årskurs 6, där två delprov valts ut för ombedömning. Delprov som ingår är de där risken för skillnader i bedömning mellan olika lärare bedömts som störst. I praktiken betyder det att de prov som väljs till stor del består av uppsatsfrågor, där eleven fritt ska formulera sig utifrån en öppen frågeställning. En mer utförlig beskrivning av urvalet av prov och delprov för ombedömningen återfinns i bilaga 2.

Urvalet följer tillsynscykel

Urvalet av skolor för ombedömningen följer från och med 2014 Skolinspektionens tillsynscykel. Urvalet är således inte utformat för att ge en representativ bild av avvikelser mellan skolor på nationell nivå. Resultatet som presenteras i denna rapport kan därför inte med säkerhet generaliseras till riket i stort. Skolinspektionen bedömer dock att urvalets sammansättning och omfattning gör det möjligt att använda resultaten som en indikation på den allmänna situationen i landet.

Det totala bortfallet på provnivå är 18 procent. Bortfallet varierar dock kraftigt mellan de olika delproven. Högst är bortfallet för vuxenutbildningen. Resultaten för delproven från vuxenutbildningen får därför anses mindre stabila än för övriga delprov. Bortfallet beskrivs mer ingående i bilaga 2.

Bedömningsprocessen

I samband med insamling av prov för ombedömning gavs skolorna möjlighet att redogöra för hur de genomfört bedömningen av de nationella proven. Information kring bedömningsprocessen efterfrågas för varje enskild elevlösning och det är frivilligt för skolorna att fylla i uppgifterna. Svarsfrekvensen för dessa uppgifter är lägre än för det totala antalet skolor som ingår i 2016 års ombedömning. Mer ingående redovisning av svarsfrekvenser och bortfall för skolornas bedömningsprocess finns i bilaga 2.

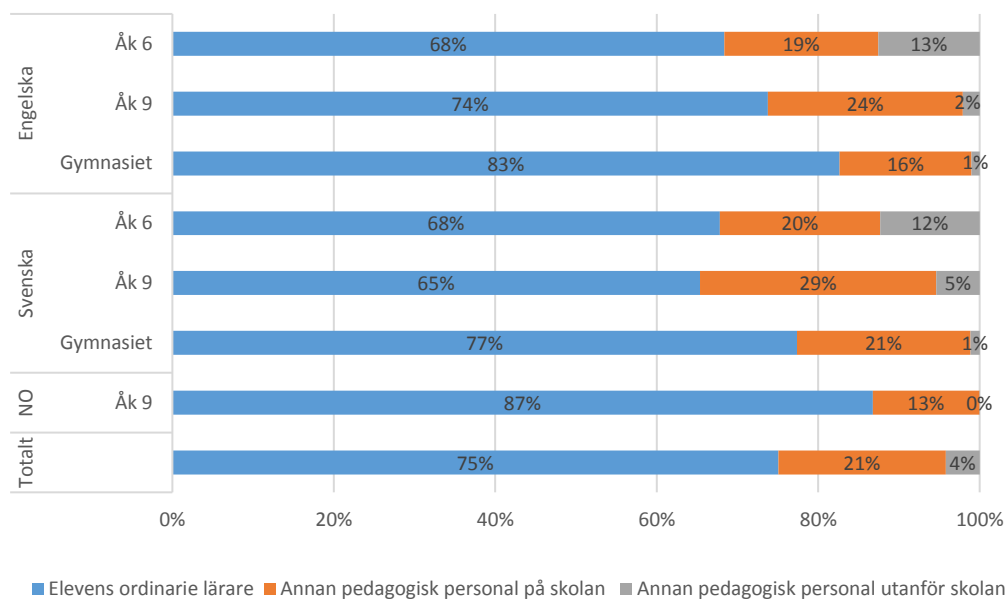
Information om bedömningsprocessen samlas in för prov genomförda inom grund- och gymnasieskolan. Däremot efterfrågas inte dessa uppgifter för prov genomförda inom den kommunala vuxenutbildningen. Dessa prov ingår således inte i analysen av skolornas bedömningsprocess. Resultatet redovisas uppdelat på årskurs 6 och 9 samt gymnasiet i respektive ämne, med undantag för NO-ämnena som redovisas sammanslaget.

Vanligast att elevens ordinarie lärare bedömer provet

I denna ombedömningsomgång framkommer att elevens ordinarie lärare i de flesta fall bedömt provet. Det gäller för 75 procent av alla ombedömda prov där skolan besvarat frågorna om hur bedömningsprocessen gått till. Tidigare ombedömningsomgångar visar på liknande resultat.

Vidare visar resultaten för 2016 års ombedömningsomgång att andelen prov som bedömts av annan än elevens ordinarie lärare på skolan är ungefär lika stor för samtliga årskurser, dock något högre för delproven i svenska och engelska i årskurs 9. Vad det gäller bedömning av prov av annan pedagogisk personal² utanför skolan är det vanligare för prov i årskurs 6 än för övriga årskurser.

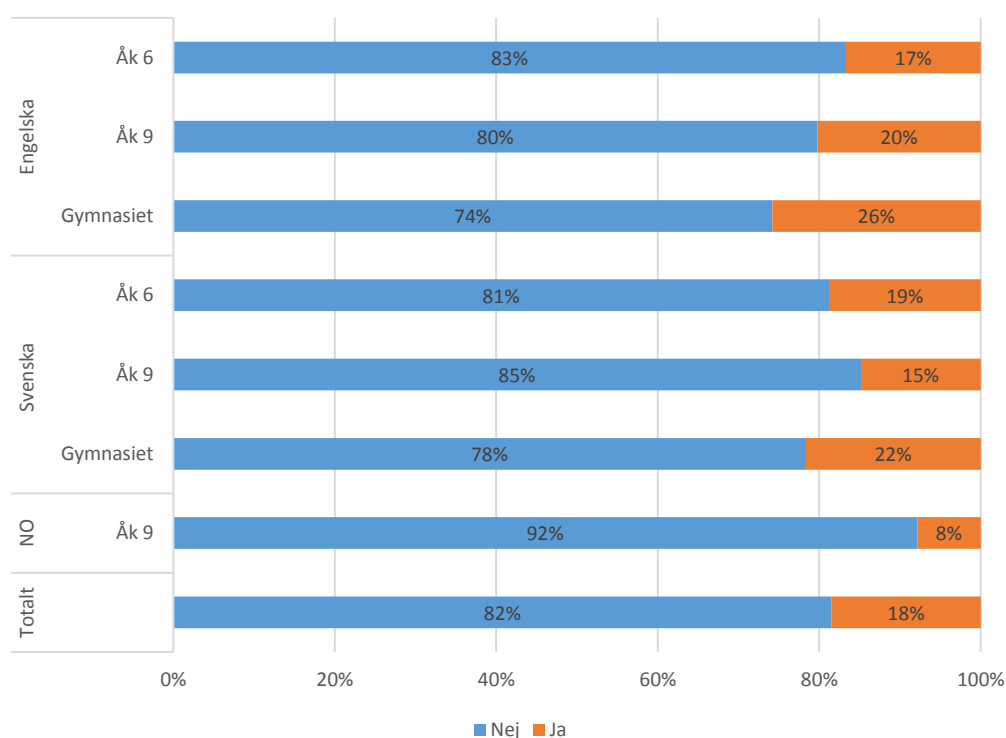
² Avser andra än elevens ordinarie lärare.



Figur 1: Skolornas svar på frågan om vem som bedömt det insända provet. Frågan besvaras för varje enskild elevlösning och är frivillig för skolan att besvara.

Läraren känner vanligen till elevens identitet vid bedömning

Tidigare omedbedömningsomgångar har visat att läraren i de flesta fall känner till elevens identitet vid bedömning. Även denna omedbedömningsomgång visar på detta förhållande. Endast en knapp femtedel av alla prov (18 procent) som ingick i omedbedömningen var avidentifierade då skolans ursprungsbedömning gjordes. Andelen avidentifierade prov var högst på gymnasiet, motsvarande 26 procent för engelska och 22 procent för svenska. Orsaken till detta kan vara att allt fler gymnasieelever gör proven på dator, vilket underlättar avidentifiering.

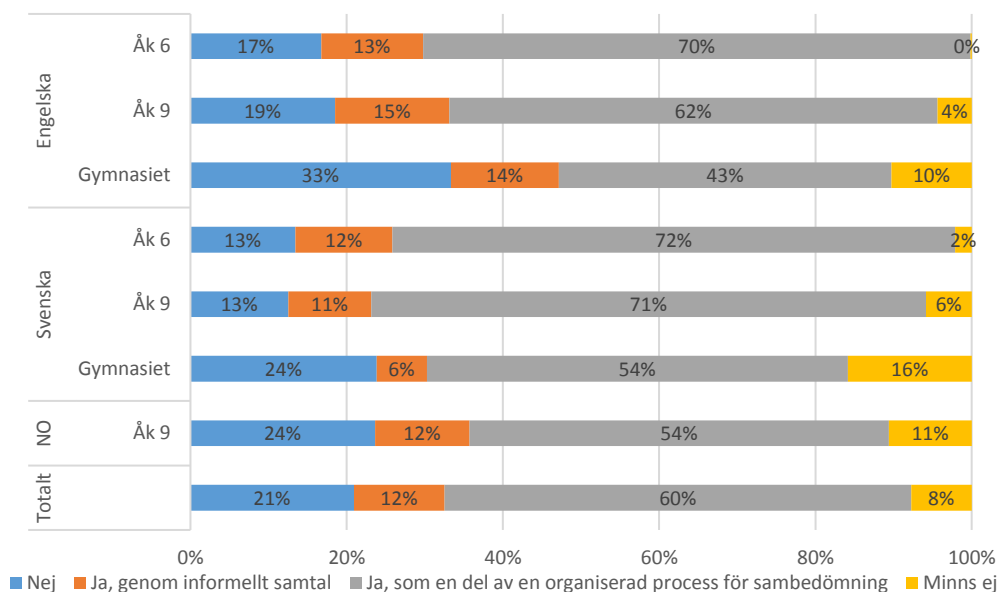


Figur 2: Skolornas svar på frågan om det insända provet var avidentifierat vid bedömningen. Frågan besvaras för varje enskild elevlösning och är frivillig för skolan att besvara.

Sambedömning tillämpas i närmare tre fjärdedelar av fallen

Vid ursprungsbedömningen av de inrapporterade proven bedömdes 72 procent genom någon form av sambedömning. Flertalet av dessa, 60 procent, bedömdes inom ramen för en organiserad process för sambedömning medan resterande tolv procent sambedömdes med hjälp av informella samtal. Vart femte prov bedömdes helt utan någon form av sambedömning. Jämfört med föregående ombedömningsomgångar har andelen prov som bedömts genom någon form av sambedömning ökat något. En möjlig orsak till förändringen kan vara en ökad medvetenhet kring dessa frågor ute i skolorna som en följd av den allmänna diskussionen kring bedömning och betygssättning av nationella prov.

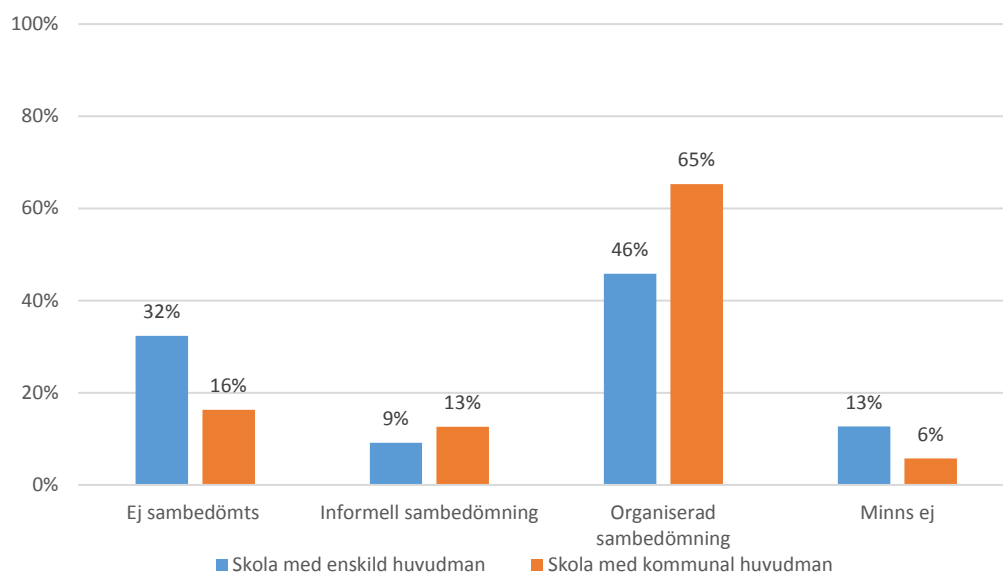
Enligt figur 3 nedan är skillnaden mellan grund- och gymnasieskolan påtaglig, sett till tillämpning av sambedömning. I årskurs 6 tillämpades sambedömning i någon form för 84 procent av proven. För prov i årskurs 9 var andelen 76 procent. Motsvarande siffra i gymnasieskolan var endast 58 procent.



Figur 3: Skolornas svar på frågan om sambedömning skett vid bedömningen av det insända provet. Frågan besvaras för varje enskild elevlösning och är frivillig för skolan att besvara.

Sambedömning är vanligare på kommunala skolor

Organiserad sambedömning är betydligt vanligare på kommunala skolor än på skolor med enskilda huvudmän, något även tidigare års ombedömningsresultat visat. 65 procent av proven genomförda på skolor med kommunal huvudman bedömdes genom en organiserad process för sambedömning. Motsvarande siffra i skolor med enskilda huvudmän var betydligt lägre, 46 procent. Förhållandet är det motsatta vad gäller andelen prov som inte sambedömts. Medan 32 procent av proven genomförda i skolor med enskild huvudman inte sambedömts var motsvarande siffra endast 16 procent för prov genomförda på skolor med kommunal huvudman.

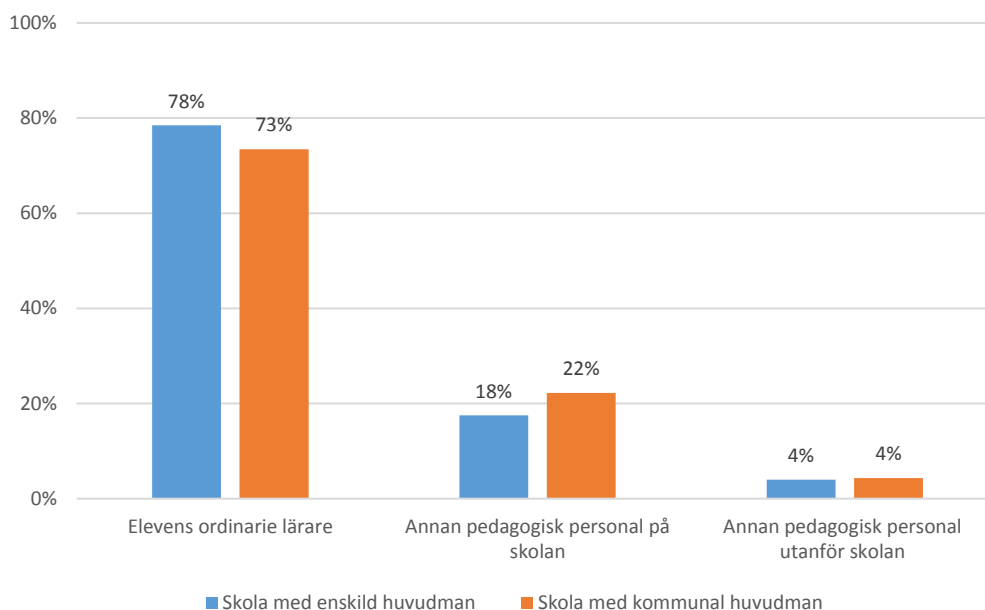


Figur 4: Andelen prov som inte sambedömts, som bedömts genom organiserad sambedömning eller genom sambedömning av mera informell karaktär utifrån skolans huvudmannatyp; enskild eller kommunal. Underlaget till figuren bygger på uppgifter som är frivilliga för skolan att lämna.

Orsaken till att sambedömning av olika slag tillämpades i större utsträckning på skolor med kommunal huvudman än på skolor med enskild huvudman kan vara att de flesta kommunala huvudmän har flera skolor, varpå skolöverskridande sambedömning och organiserade sambedömningsprocesser är enklare att samordna. Storlek på skolan kan även vara av avgörande betydelse för förekomsten av sambedömning. På mindre skolor försvåras till exempel förutsättningarna för sambedömning om det finns få undervisande lärare inom ett ämne.

Vanligare att elevens ordinarie lärare bedömer provet på fristående skolor

Andel prov som bedömts av elevens ordinarie lärare på skolor med enskild huvudman är något större jämfört med kommunal huvudman, motsvarande 78 respektive 73 procent. I föregående ombedömningsomgång var skillnaderna tydligare mellan kommunal och enskild huvudman vad det gäller vem som bedömt provet. I denna omgång är skillnaden mindre uttalad.



Figur 5: Andelen prov som bedömts av antingen elevens ordinarie lärare eller annan pedagogisk personal, utifrån skolans huvudmannatyp; enskild eller kommunal. Underlaget till figuren bygger på uppgifter som är frivilliga för skolan att lämna.

Avidentifiering lika vanligt i skolor med kommunal som enskild huvudman

Andelen avidentifierade prov för kommunala skolenheter uppgår till 18 procent, och för enskilda skolenheter uppgår andelen till 19 procent. Andelen avidentifierade prov skiljer sig inte nämnvärt mellan de två huvudmannatyperna för prov genomförda på gymnasiet respektive i grundskolans årskurs 9. I gymnasiet är andelen avidentifierade prov något större för skolenheter med enskilda huvudmän medan det i årskurs 9 är något vanligare med avidentifierade prov bland skolenheter med kommunal huvudman. För elever i årskurs 6 är skillnaderna större. Andelen kommunala skolenheter med avidentifierade prov uppgår till 18 procent jämfört med en procent för skolenheter med en enskild huvudman.

Större andel avidentifierade delprov i engelska och svenska

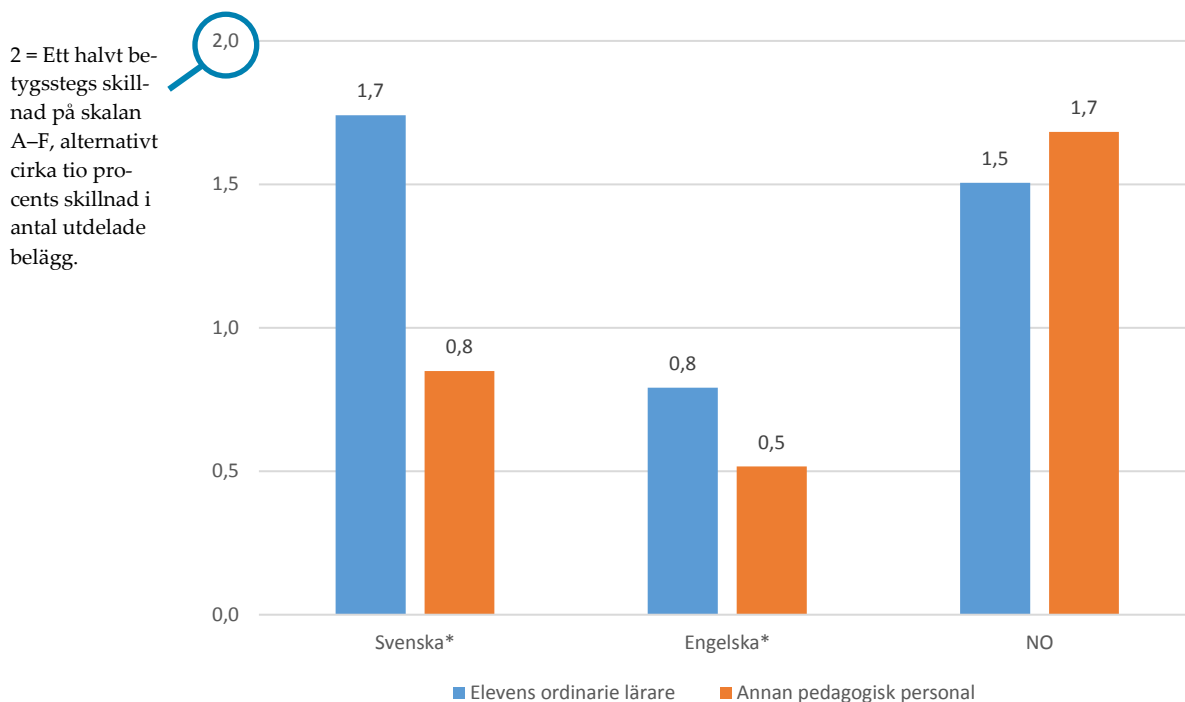
För de båda huvudmannatyperna är andelen avidentifierade prov störst bland delproven i engelska och svenska. Allra störst är skillnaden mellan de båda huvudmannatyperna i svenska årskurs 6, där 20 procent av proven genomförda på skolor med kommunal huvudman var avidentifierade, medan inga prov på skolor med enskild huvudman var avidentifierade. Däremot var andelen avidentifierade prov i engelska respektive svenska på gymnasiet tre procentenheter högre bland enskilda skolor än kommunala.

Det bör poängteras att andelen prov från skolor med enskilda huvudmän är betydligt lägre än den för kommunala skolor, vilket gör att resultatet får tolkas med viss reservation.

Bedömningsprocessen kan ha stor betydelse för utfallet

Som visas i figur 6 nedan är den genomsnittliga skillnaden mellan skolans bedömning och omdömningen statistiskt signifikant mindre för prov som bedömts av någon annan än elevens ordinarie lärare, för två av tre ämnesgrupper.

Den generella tendensen till att skolan gör en annan bedömning än omdömarens minskar generellt sett när någon annan än den ordinarie läraren står för bedömningen när det gäller delproven i svenska och engelska. För NO-proven visar resultaten dock på *motsatt* riktning. Antalet skolor som resultaten bygger på är dock litet för de enskilda NO-ämnena. I NO-proven ses dock ingen statistiskt signifikant effekt av extern bedömning på aggregerad nivå.

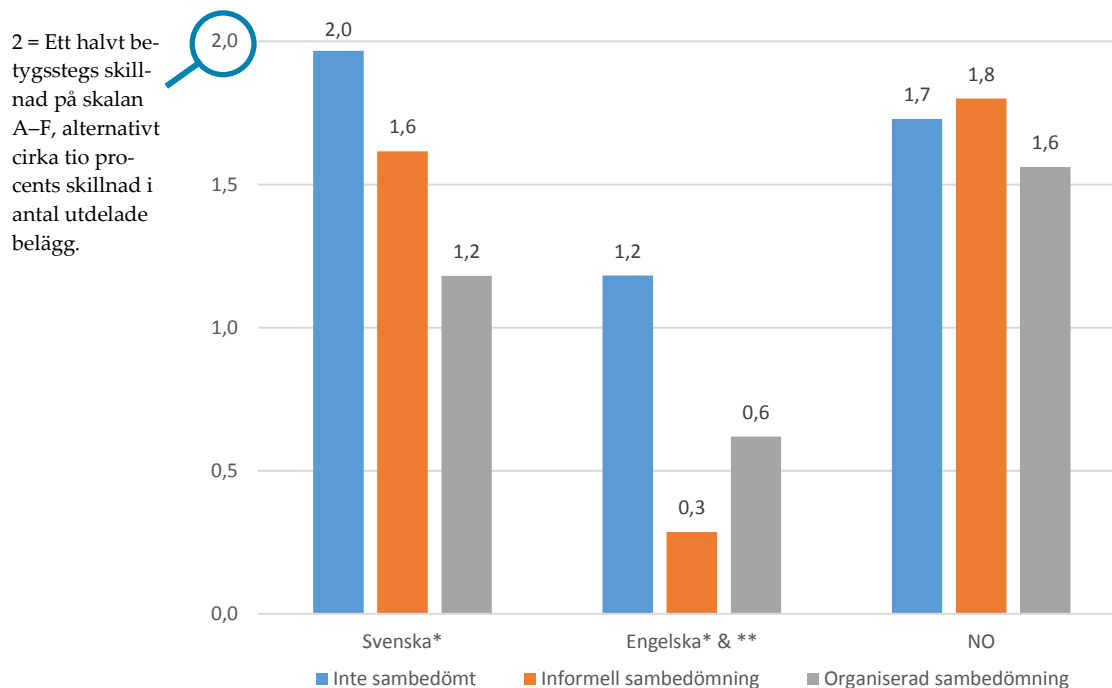


Figur 6: Genomsnittlig skillnad mellan skolans bedömning och omdömningen för prov som bedömts av elevens ordinarie lärare eller av annan pedagogisk personal.

* Statistiskt signifikant *lägre* genomsnittsavvikelse för externt bedömda prov.

På liknande sätt verkar sambedömning, framförallt i svenska och engelska, ha en återhållande effekt på betygssättningen. För dessa ämnen är skillnaden mel-

lan skolan och ombedömarens bedömning statistiskt signifikant mindre för delprov som bedömts i en organiserad process för sambedömning än för delprov där ingen sambedömning skett.



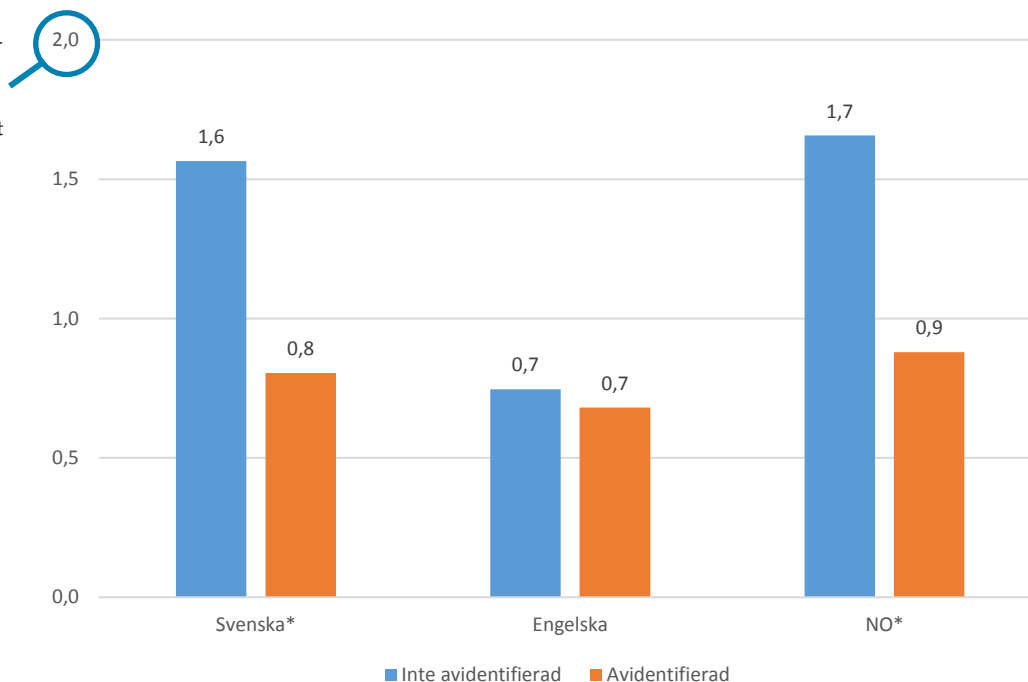
Figur 7: Genomsnittlig skillnad mellan skolans bedömning och ombedömningen för prov där bedömaren inte diskuterat bedömningen med någon annan, för prov där bedömningen diskuterats genom informellt samtal, respektive för prov som bedömts i organiserad sambedömning.

* Statistiskt signifikant *lägre* genomsnittsavvikelse gällande organiserat sambedömda prov jämfört med inte sambedömda prov.

** Statistiskt signifikant *högre* genomsnittsavvikelse gällande organiserat sambedömda prov jämfört med informellt sambedömda prov.

Även aidentifiering verkar ha en viss återhållande effekt på betygsättningen. I två ämnesgrupper, svenska och NO, är skillnaden mellan skolans och ombedömarens bedömning statistiskt signifikant mindre för aidentifierade prov än för icke aidentifierade prov.

2 = Ett halvt betygsstegs skillnad på skalan A–F, alternativt cirka tio procents skillnad i antal utdelade belägg.



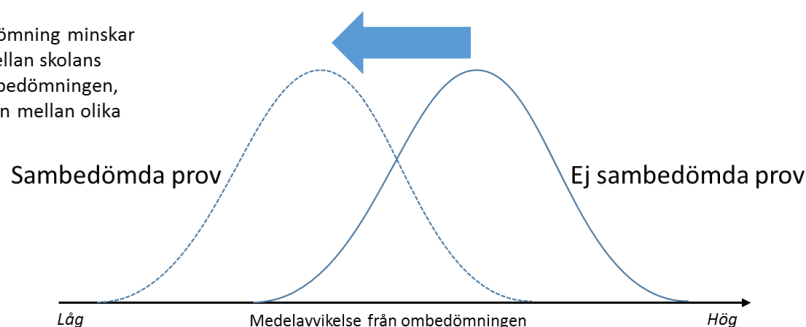
Figur 8: Genomsnittlig skillnad mellan skolans bedömning och omdömningen för prov som var avidentifierade vid bedömningen, respektive för prov där elevens identitet var känd för läraren.

* Statistiskt signifikant *lägre* genomsnittsavvikelse för avidentifierade prov.

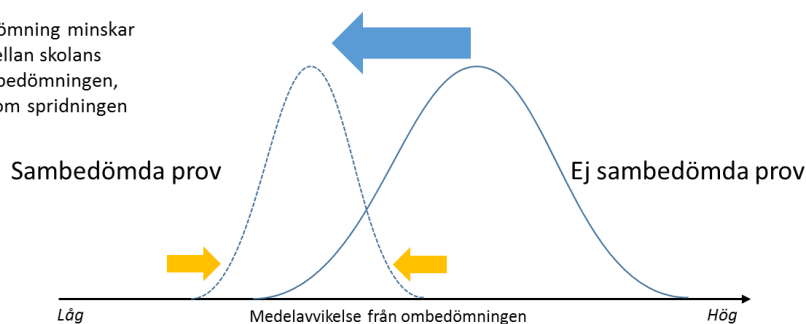
Sambedömning, extern bedömning och avidentifiering minskar avvikelsegraden

Föregående avsnitt visar att extern bedömning, sambedömning och avidentifiering tenderar att generellt sett ha en återhållande effekt på delprovsbetygen. Detta resulterar i ett något lägre betyg i snitt för motsvarande elevlösning än om elevens ordinarie lärare gör bedömningen utifrån en avidentifierad elevlösning och utan sambedömning. Frågan är däremot om det även minskar *spridningen* i bedömning mellan skolor, eller om det bara är bedömningarnas medelpunkt som flyttas. Ur likvärdighetssynpunkt är nämligen spridningen mellan skolor det intressanta. Var "mittpunkten" i bedömningarna ligger är däremot relativt ointressant ur likvärdighetssynpunkt, givet att alla skolor lägger ribban på samma nivå.

Hypotes 1: Sambedömning minskar medelavvikelsen mellan skolans bedömning och ombedömningen, men inte spridningen mellan olika skolor



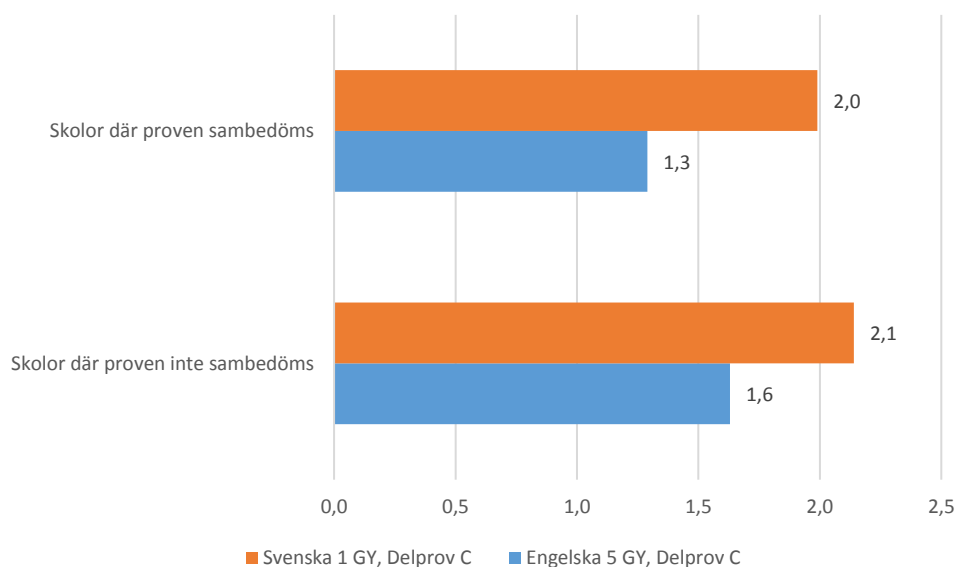
Hypotes 2: Sambedömning minskar medelavvikelsen mellan skolans bedömning och ombedömningen, och minskar dessutom spridningen mellan olika skolor



Figur 9: Två alternativa hypoteser kring effekten av sambedömning och extern bedömning.

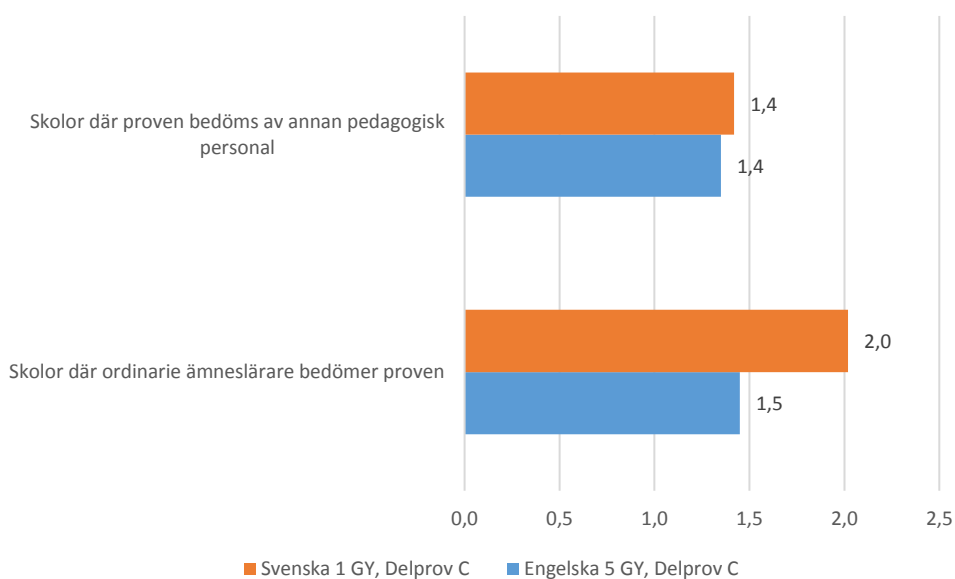
Denna analys är endast möjlig att göra för de ämnen där de allra flesta skolor inkluderats i ombedömningen, vilket i denna ombedömningsomgång endast gäller Svenska 1 respektive Engelska 5 på gymnasiet. I övriga fall är antalet skolor i urvalet för litet för att några tydliga slutsatser ska kunna dras. Resultatet redovisas i figur 10, 11 och 12 nedan.

Resultaten ger en viss indikation på att sambedömning möjligen kan minska, inte bara medelavvikelsen från ombedömningen, utan även spridningen mellan olika skolor. Skillnaden är däremot så pass liten att en slumpmässig tillfällighet inte kan uteslutas.

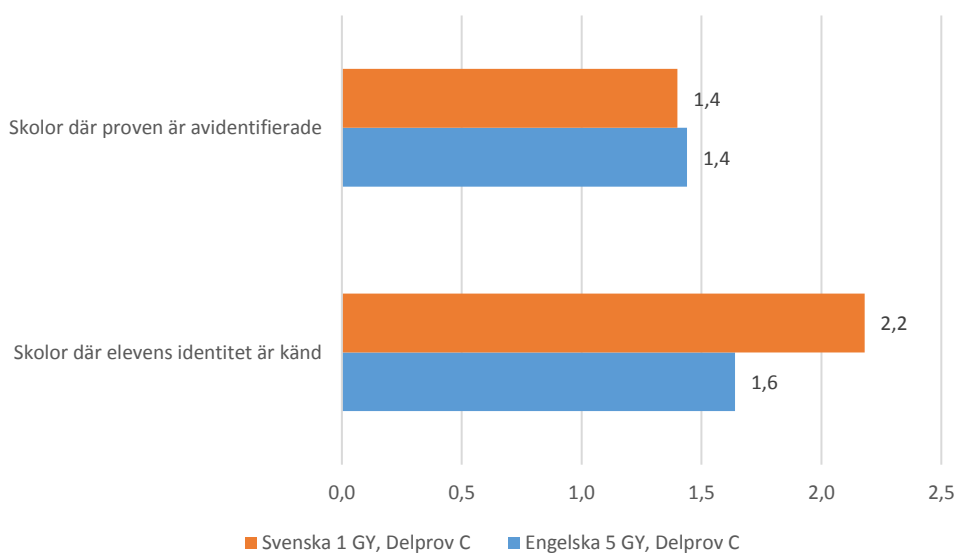


Figur 10: Spridning i medelavvikelse från omdömningen bland skolor som sambedömer sina prov respektive för skolor där sambedömning inte tillämpas. $p_{75}-p_{25}$ = skillnaden mellan den 75:e percentilens och den 25:e percentilens skolor, med avseende på genomsnittlig avvikelse från omdömningen. 4 på skalan motsvarar ett betygsstegs skillnad på skalan A–F.

Även resultatet för extern bedömning (figur 11) och aidentifiering (figur 12) ger en liknande indikation. Skillnaden är dock även här så pass liten att en slumpmässig tillfällighet inte kan uteslutas.



Figur 11: Spridning i medelavvikelse från omdömningen bland skolor där ordinarie ämneslärare bedömer proven respektive skolor där annan pedagogisk personal på eller utanför skolan bedömer proven. p75-p25 = skillnaden mellan den 75:e percentilens och den 25:e percentilens skolor, med avseende på genomsnittlig avvikelse från omdömningen. 4 på skalan motsvarar ett betygsstegs skillnad på skalan A–F.



Figur 12: Spridning i medelavvikelse från omdömningen bland skolor där proven varit avidentifierade respektive skolor där proven inte varit avidentifierade. p75-p25 = skillnaden mellan den 75:e percentilens och den 25:e percentilens skolor, med avseende på genomsnittlig avvikelse från omdömningen. 4 på skalan motsvarar ett betygsstegs skillnad på skalan A–F.

Att vare sig extern bedömning, sambedömning eller avidentifiering löser problemet med spridning mellan skolor är utifrån omständigheterna inte särskilt förvånande. En sambedömargrupp kan nämligen ha goda förutsättningar att *sinsemellan* enas kring uttolkningen av bedömningsanvisningarna. Däremot hindrar det inte att en annan sambedömargrupp på annat håll i landet enas om en helt annan tolkning.

Resultatet från ombedömningen visar att såväl sambedömning, extern bedömning som avidentifiering har en viss återhållande effekt på den genomsnittliga bedömningen av en given elevlösning (tenderar att leda till ett något lägre betyg i snitt). Skolornas olika tillvägagångssätt vid bedömningen kan därför mycket väl undergräva elevens rätt till en likvärdig bedömning.

Hur mycket skiljer sig bedömningarna av samma elevlösning?

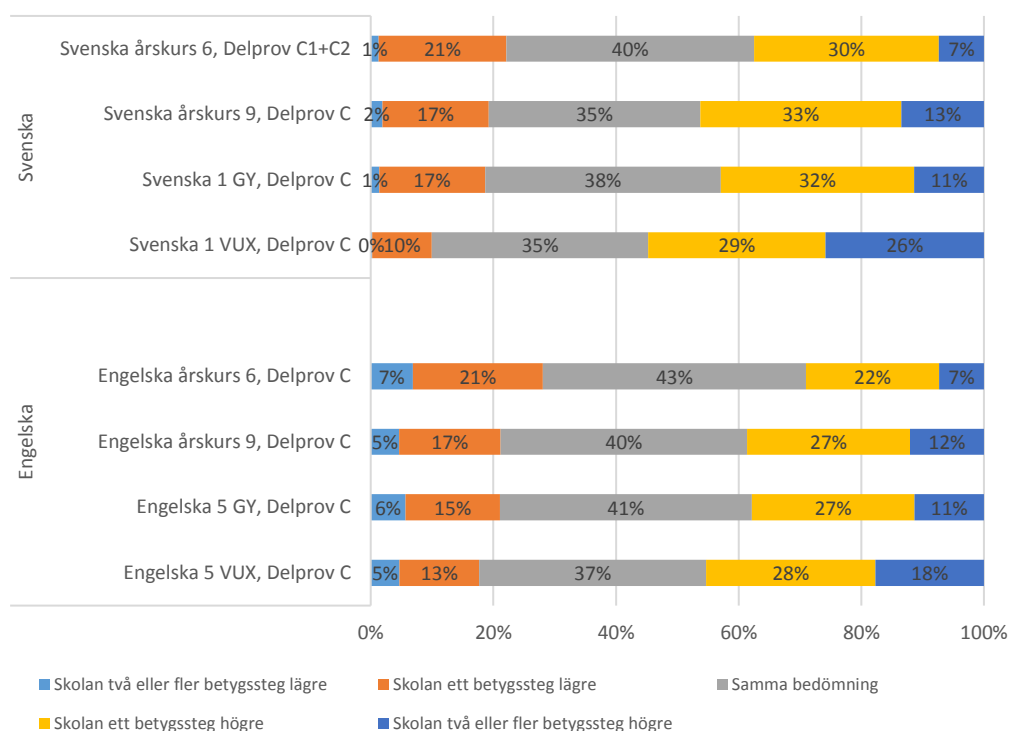
Olika bedömningar för mer än varannan elevlösning

För 62 procent av elevlösningarna skiljer sig ursprungsbedömningen och ombedömningen med minst ett betygssteg. I 21 procent av fallen har skolans lärare satt ett delprovsbetyg som är minst ett betygssteg *lägre* än ombedömaren och i 41 procent av fallen har skolans lärare satt ett delprovsbetyg som är minst ett betygssteg *högre* än ombedömaren. I 38 procent av fallen har skolan och ombedömaren satt samma betyg.

Något större bedömningsskillnader i svenska än i engelska

För proven i engelska och svenska ges delprovsbetyg, som i sin tur vägs samman till ett provbetyg. De ombedömda delproven i dessa ämnen består av uppsatser där eleven har skrivit fritextsvar.

I figur 13 nedan visas andelen avvikelser mellan ursprungsbedömningen och ombedömningen i delprov C i svenska och engelska. Bedömningsskillnaderna är något större bland proven i svenska jämfört med proven i engelska i samtliga årskurser. För 35–40 procent av proven i svenska och för 37–43 procent av proven i engelska har skolans lärare satt samma delprovsbetyg som ombedömaren på skalan A–F. Det är vanligare att skolans lärare har satt ett högre delprovsbetyg än ombedömaren i proven i svenska jämfört med i proven i engelska.



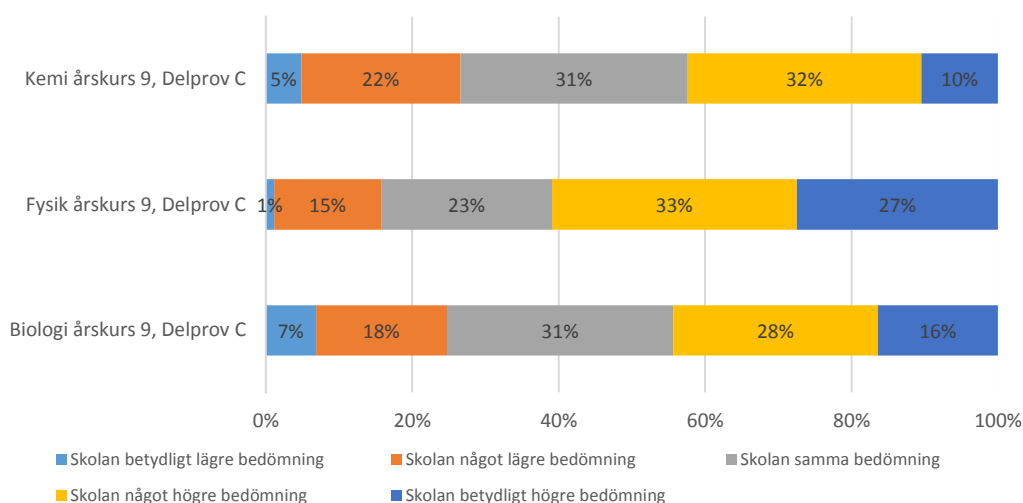
Figur 13: Andel avvikelser mellan bedömningen från skolans lärare och ombedömarnes bedömning i delprov C i svenska och engelska, årskurs 6, årskurs 9, gymnasiet och vuxenutbildningen.

Andelen avvikelser mellan skolans bedömning och ombedömningen följer ungefär samma mönster i alla årskurser, med undantag av vuxenutbildningen. Bland proven i den kommunala vuxenutbildningen är andelen avvikelser där skolans lärare har satt minst två betygssteg högre än ombedömaren något större än i övriga årskurser, både i svenska och i engelska. Resultaten avseende proven i vuxenutbildningen bör dock tolkas med försiktighet eftersom bortfallet är stort och antalet insamlade prov är lågt.

Störst andel avvikelser i fysik

För NO-prov sätts inga delprovsbetyg utan i stället delas det ut *belägg*. I de fall där bokstavs-betyg inte förekommer har beläggen räknats om till en gemensam skala som sträcker sig från 0–20 för att ge en mer lättbegriplig bild över hur avvikelserna ser ut.³

³ I bilaga 3 redovisas i detalj hur resultatmättet beräknats.



Figur 14: Andel avvikelser mellan bedömningen från skolans lärare och ombedömarens bedömning i delprov C kemi, fysik och biologi, årskurs 9.

Fysik är det ämne som resulterat i den största andelen avvikelser av alla ämnen som ingått i ombedömningen (se figur 14). Andelen elevlösningar som bedömts generöst av skolan i jämförelse med ombedömaren är särskilt stor. Bland 60 procent av elevlösningarna har skolans lärare gjort en något högre eller betydligt högre bedömning jämfört med ombedömaren i fysikämnet. Enbart 23 procent av elevlösningarna fick samma bedömning. Så såg det ut även i förra årets ombedömning. I ämnena kemi och biologi gjorde skolan och ombedömaren samma bedömning i 31 procent av fallen.

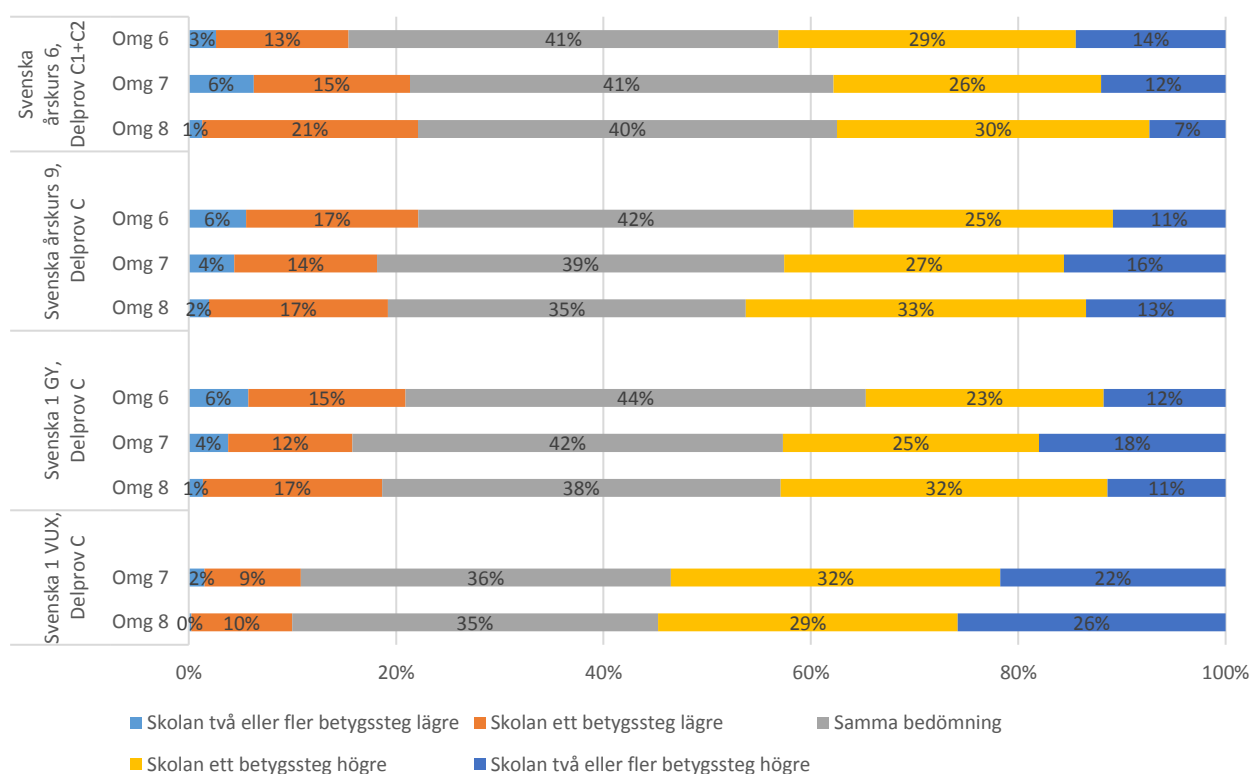
Har resultatet förändrats jämfört med tidigare ombedömningar?

I följande avsnitt jämförs Skolinspektionens senaste ombedömning av nationella prov (omgång 8) med resultatet från de två senaste ombedömningarna av nationella prov (omgång 6 och 7). Tidigare ombedömningar (omgång 1–5) har uteslutits i jämförelsen på grund av att provens utformning och de betygsskalor som används i olika årskurser har förändrats, vilket innebär att resultaten inte är jämförbara.

Små uppmätta skillnader i svenska mellan omgång 6, 7 och 8

En jämförelse med tidigare års ombedömningar av delprov C i svenska visar att andelen avvikelser inte har förändrats påtagligt, se figur 15. Resultatet från omgång 8 är på det stora hela jämförbart med resultatet i omgång 6 och 7 i alla

årskurser.⁴ Några mindre förändringar kan emellertid konstateras. I årskurs 9 och i gymnasiet har andelen avvikelser mellan ursprungsbedömningen och omdedömningen ökat något för varje år. Det har framför allt skett en ökning av andelen prov där skolans lärare satt ett betygssteg högre än omdedömaren. I årskurs 6 är det en större andel av lärarna som har satt ett betygssteg lägre än omdedömarna i omgång 8 jämfört med omgång 6 och 7. Samtidigt har andelen som satt två eller fler betygssteg högre minskat.



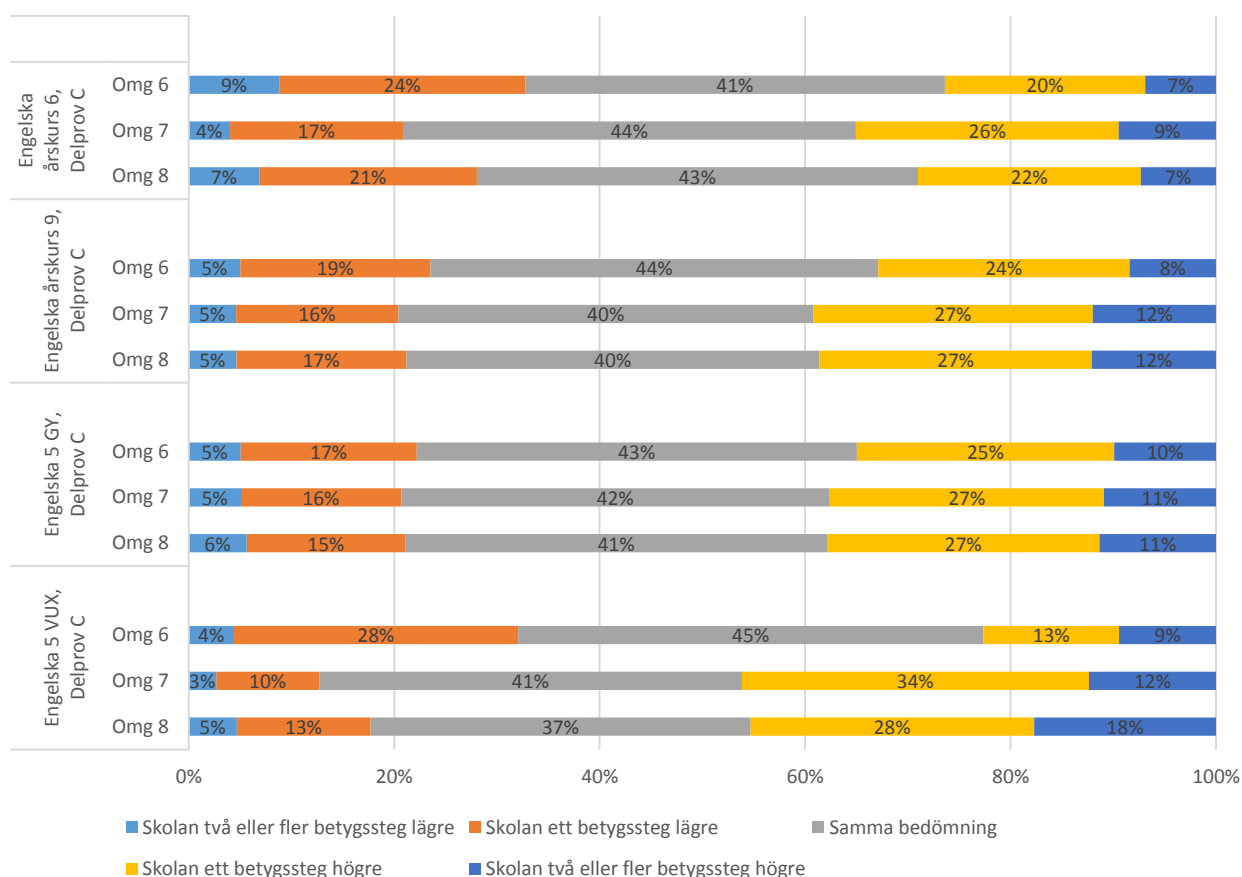
Figur 15: Andel avvikelser mellan bedömningen från skolans lärare och omdedömarens bedömning för delprov C i svenska, omgång 6, 7 och 8.

Andelen avvikelser i engelska överensstämmer i hög grad med omgång 7

Figur 16 nedan visar en jämförelse av andelen avvikelser i bedömningen av delproven i engelska mellan de tre senaste omdedömningarna av nationella prov. Resultaten av omdedömningarna från omgång 7 och 8 är mycket lika i samtliga

⁴ Resultatet från omdedömningen i Svenska 1 i vuxenutbildningen omgång 6 ingår inte i jämförelsen på grund av att delprov B användes i stället för delprov C (som användes i övriga årskurser och omgångar). Delprov B är utformat på ett annat sätt än delprov C. Resultaten är därför inte jämförbara.

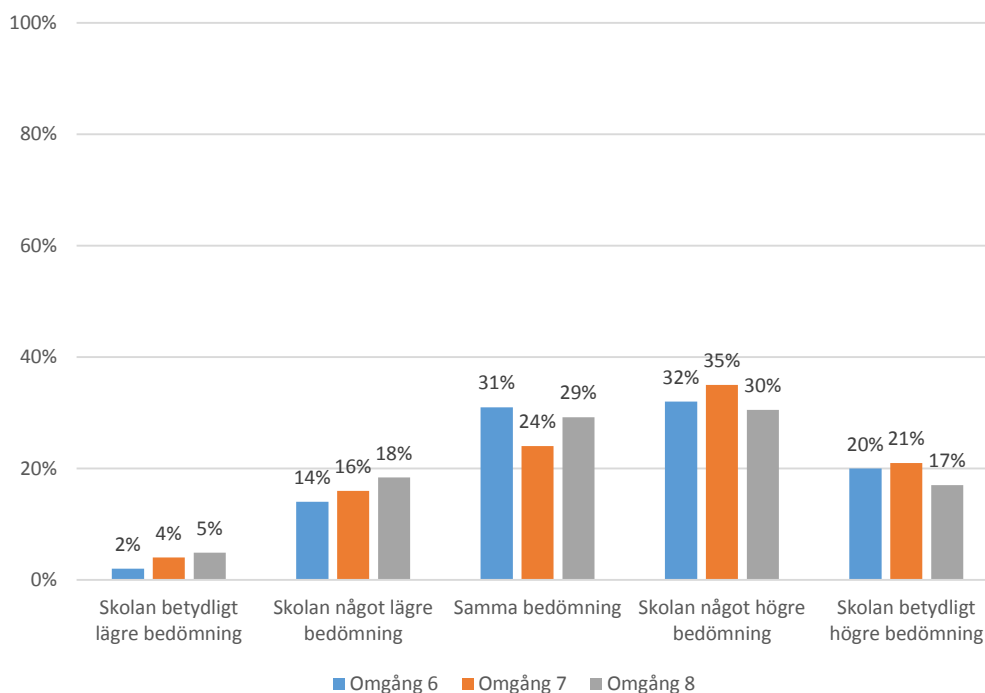
årskurser. Resultatet från omgång 6 visar däremot generellt något färre avvikelser där skolan gjort en generösare bedömning än ombedömaren. Det var alltså något mindre vanligt att skolans lärare satte högre betyg än ombedömaren i denna omgång jämfört med i omgång 7 och 8.



Figur 16: Andel avvikelser mellan bedömningen från skolans lärare och ombedömaren's bedömning för delprov C i engelska, omgång 6, 7 och 8.

Små skillnader mellan åren för NO-ämnena

Resultatet av ombedömningen av NO-ämnena i omgång 8 är i linje med resultatet från omgång 6 och 7, vilket illustreras i figur 17 nedan. Det är vanligt att skolans lärare gör en något högre bedömning än ombedömaren. I den senaste ombedömningen av nationella prov är det dock en något lägre andel elevlösningar där skolan har gjort en högre bedömning och en något högre andel där skolan har gjort en lägre bedömning, jämfört med tidigare år.



Figur 17: Andel avvikelser mellan bedömningen från skolans lärare och ombedömarens bedömning för delprov i NO-ämnena i årskurs 9, omgång 6, 7 och 8.

Stor variation mellan olika skolor för samma elevlösning

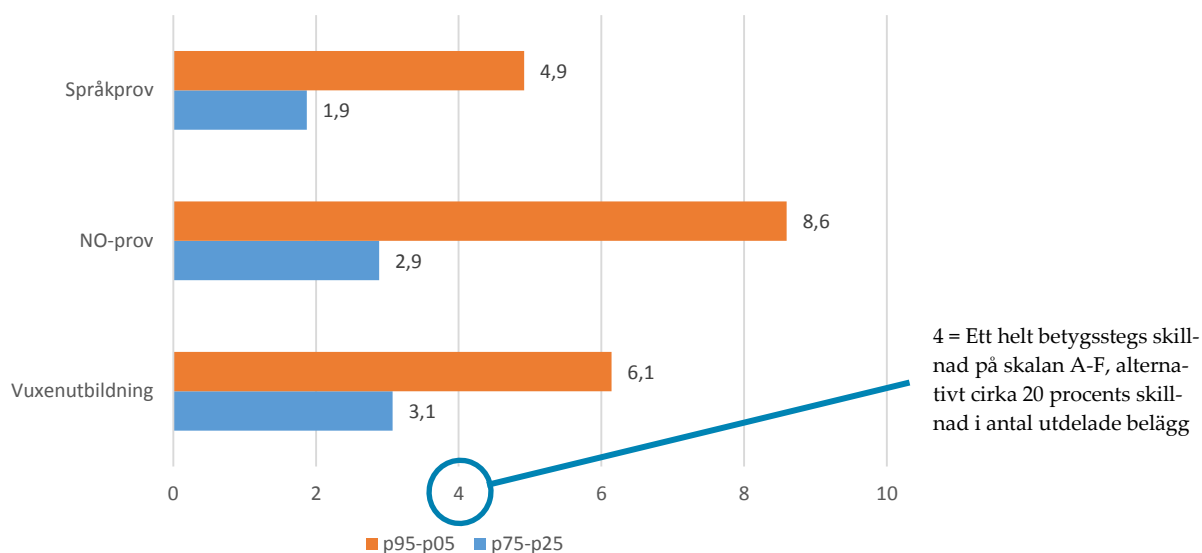
Utifrån ett likvärdighetsperspektiv är det mer intressant att jämföra spridningen mellan *olika* skolor än mellan en viss skola och ombedömningen.

Eftersom de flesta skolorna bedömer sina egna prov är det inte möjligt att utifrån ombedömningen analysera andelen avvikelser mellan olika skolor. Resultaten från ombedömningen kan däremot användas som en gemensam referenspunkt för att undersöka om olika skolor *i genomsnitt* sätter lika höga eller låga provbetyg för likvärdiga elevlösningar.⁵ I bilaga 4 beskrivs tillvägagångssättet i beräkningarna närmare.

I figur 18 nedan redovisas den beräknade genomsnittliga skillnaden av likvärdiga elevlösningar mellan olika skolor, redovisat per ämnesgrupp. Resultaten visar att en elev på en skola som sätter relativt låga delprovsbetyg kan förväntas få ett delprovsbetyg som ligger cirka ett halvt till ett betygssteg lägre än om

⁵ "Likvärdiga elevlösningar" syftar här på en hypotetisk situation där olika skolor bedömt samma elevlösningar.

eleven lämnat in samma svar på en skola som sätter relativt höga delprovsbetyg.



Figur 18: 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 ombedömningen.

Sett till de skolor som placerar sig längst ute på kanterna växer skillnaderna ytterligare. En elev på en skola som tillhör de fem procent mest strikta skolorna i landet när det gäller betygssättning kommer typiskt sett att få ett provbetyg som ligger runt ett till två betygssteg lägre på skalan A–F än om eleven lämnat in samma svar på en av de fem procent mest generösa skolorna.

Spannet mellan de skolor som är relativt strikta respektive relativt generösa i betygssättningen av delprovsbetyg är något större för NO-proven⁶ och delproven genomförda inom den kommunala vuxenutbildningen jämfört med språkproven.

Kan skolan förklara avvikelsegraden?

Att det förekommer skillnader i hur en elevlösning bedöms av olika sinsemellan oberoende ombedömare säger inget i sig om varför så är fallet. En hypotes, som kan testas utifrån ombedömningsresultaten, är ifall det förekommer systematiska skoleffekter.

⁶ Uttryckt på den harmoniserade bedömningsskalan 0–20. För delproven i NO ges inga bokstavs-betyg.

Systematiska skoleffekter. Uppmätta skillnader mellan skolor kan bero på att vissa skolor har *skolövergripande processer* eller en informell *kultur* av generös alternativt strikt betygssättning. Det finns sedan tidigare indikationer på att det är lättare att nå ett högre slutbetyg på skolor där snittnivån är låg än på skolor där snittnivån är hög.⁷ De tillgängliga register som används medger inte att svaren analyseras på lärarnivå, utan endast på skolnivå. För flera 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.

Svag positiv korrelation för de systematiska skoleffekterna

För 289 skolenheter finns ombedömningsresultat för två ämnen. För dessa 289 skolor finns en statistiskt signifikant skoleffekt, definierad som korrelationen mellan genomsnittlig avvikelsegrad för det ena ämnet, det vill säga "ämne 1", och genomsnittlig avvikelsegrad för det andra ämnet, det vill säga "ämne 2". Den uppmätta korrelationen mellan de två ämnena är 0,126.⁸ Ett samband förekommer, men det är inte särskilt starkt.

Om sambandsanalysen begränsas till delproven i svenska och engelska, vilket kan göras i gymnasiet, är sambandet aningen starkare: korrelationen är här 0,203 mellan de två avvikelsegraderna. Skoleffekten förekommer men den är inte särskilt påtaglig, eftersom korrelationen mellan de genomsnittliga avvikelserna är svag.

En möjlig förklaring till de uppmätta skoleffekterna är skolövergripande skillnader i bedömningsprocesserna, eftersom dessa har en effekt på betygssättningen för delproven.⁹ Det kan till exempel vara så att vissa skolor konsekvent tillämpar sambedömning för alla ämnen, och vissa inte gör det.

Antalet skolor som ingår i ombedömningen är för litet för att denna fråga ska kunna utredas grundligt. Korrelationen är statistiskt signifikant men något lägre jämfört med tidigare omgångar, mellan de två ämnesresultaten, vilket antyder att de uppmätta skoleffekterna inte i någon större utsträckning kan förklaras av systematiska skillnader i bedömningsprocessen på skolnivå.

⁷ Se till exempel Fredriksson, Peter & Vlachos, Jonas (2011). Reformen och resultat: Kommer regeringens utbildningsreformer att ha någon betydelse? s. 29.

⁸ Single Measure Intraclass Correlation Coefficient.

⁹ Skolinspektionen (2015). Ombedömning av nationella prov 2014: "Processerna spelar roll".

Sammanfattningsvis bör det poängteras att datamaterialet inte är tillräckligt stort för att med säkerhet kunna ge svar på frågan om det finns systematiska skoleffekter. Eventuellt kan skillnader i bedömningsprocesser mellan skolorna förklara en del av den uppmätta skoleffekten, men de kan inte förklara hela effekten.

Resultaten från ombedömningen indikerar att en så kallad skoleffekt (systematiska effekter på skolnivå) förekommer, men att skoleffekten bara förklarar en viss del av bedömaravvikelseerna. Det betyder inte att återstående variation är osystematisk – flera andra typer av systematiska effekter är tänkbara, exempelvis med avseende på lärar- eller elevens egenskaper (ålder, kön, bakgrund och så vidare).¹⁰ Däremot kan dessa effekter inte kontrolleras inom ramen för detta uppdrag.

Små avvikelser mellan kommunala och fristående skolor

Skillnaden i andelen betydande avvikelser mellan de olika huvudmannatyperna är liten. För skolor med enskilda huvudmän har 82 procent av proven fått en likartad bedömning¹¹ och för skolor med kommunal huvudman är motsvarande andel 85 procent. Däremot är andelen prov som bedömts generösare något högre bland de enskilda skolorna (14 procent) jämfört med de kommunala (11 procent) skolorna.

Avvikelseerna *kan* ha en avgörande betydelse för det slutgiltiga provbetyget

Eftersom endast ett delprov per prov ombedöms är det inte givet att en uppmätt skillnad för ett enskilt delprov får genomslag på det slutliga provbetyget. Ofta räcker det däremot med ett betygsstegs skillnad för ett delprov för att det totala provbetyget ska påverkas. För delprov i NO-ämnen kan en skillnad på 2–3 betyg, vilket är vanligt förekommande för de ombedömda delproven, påverka det totala provbetyget.

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

¹⁰ 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.

¹¹ Med likartad bedömning avses avvikelser på ett betygssteg eller mindre.

De nationella proven ger inte hela bilden av en elevs ämneskunskaper, då endast vissa delar av kursplanen prövas, men är en viktig del i lärarens bedömningsunderlag.¹² Detta innebär att elevens slutbetyg/kursbetyg inte behöver överensstämma med resultatet på det nationella provet. Läraren ska utgå ifrån elevens prestationer på de nationella proven, men förväntas även använda annan tillgänglig information om elevens kunskaper i förhållande till kunskapskraven.¹³

¹² Skolverket (2016). PM – Resultat på nationella prov i årskurs 3, 6 och 9, läsåret 2015/16.

¹³ Skolverket (2011). Läroplan för grundskolan, förskoleklassen och fritidshemmet. Reviderad 2016.

Faktorer som påverkar samstämmigheten i bedömningen

Alla elever har rätt till en likvärdig och rättvis bedömning och betygssättning. Betyg har en direkt koppling till behörighet och urval för vidare studier och kan därmed ha en avgörande betydelse för elevens framtida förutsättningar att lyckas på en konkurrenspräglad arbetsmarknad. Därför är det viktigt att alla elever, oavsett vilken lärare de har eller vilken skola de går i, blir bedömda utifrån samma kriterier.

Utifrån resultaten från denna rapport vill Skolinspektionen belysa några faktorer som kan bidra till att öka samstämmigheten mellan bedömningarna av nationella prov.

Sambedömning, extern bedömning och avidentifiering ökar likvärdigheten

Av resultaten från ombedömningen framgår att olika former av sambedömning och extern bedömning minskar avvikelserna mellan skolans bedömning och ombedömningen. Resultaten visar att organiserad sambedömning leder till en lägre avvikelsegrad jämfört med prov som inte har sambedömts. På liknande sätt indikerar resultaten att prov som bedömts av annan pedagogisk personal, antingen på eller utanför den skola eleven går på, uppvisar färre avvikelser än prov som bedömts av elevens ordinarie lärare. Även avidentifiering av elevens identitet förefaller minska avvikelsegraden och därmed främja en likvärdig bedömning av de nationella proven.

Liknande processer och samverkan ökar samstämmigheten i bedömningen

Bedömningsprocesserna har en systematisk effekt på betygssättningen. Resultaten från ombedömningen visar att inslagen av sambedömning, extern bedömning och avidentifiering varierar kraftigt både mellan skolor, ämnen och årskurser, liksom mellan skolor med kommunal respektive enskild huvudman. Denna "spretighet" på systemnivå riskerar att i sig minska likvärdigheten i bedömningen.

Skolinspektionen vill understryka vikten av samverkan mellan skolor kring bedömning. Samverkan *inom* skolor ökar likvärdigheten inom skolan men har begränsad potential att öka likvärdigheten *mellan* skolor. Ur ett likvärdighetsperspektiv är samstämmigheten minst lika viktig *mellan* som *inom* skolor.

Avslutningsvis

Ombedömningen av nationella prov, som varit ett regeringsuppdrag för Skolinspektionen sedan 2009, har över åren bidragit med värdefull kunskap. Skolinspektionen har i ett antal rapporter lyft slutsatser och utvecklingsområden för ökad likvärdighet i bedömningarna av proven som varit till nytta i flera olika sammanhang. Resultaten från tidigare ombedömningar har bland annat refererats till i såväl betänkandet *Likvärdigt, rättssäkert och effektivt – ett nytt nationellt system för kunskapsbedömning* (SOU 2016:25) som Regeringens proposition *Nationella prov – rättvisa, likvärdiga, digitala* (Prop. 2017/18:14).

Skolinspektionen ser det som angeläget att även fortsatt bibehålla en statlig granskning av skolors och huvudmäns arbete med bedömning och betygssättning för att säkerställa elevernas rätt till likvärdighet i detta avseende.

Dock har de övergripande slutsatserna för ombedömningsuppdraget varit återkommande år från år. I och med att resultaten är likartade vid varje genomförd ombedömning är mervärdet av en fortsatt ombedömning begränsat. Ombedömningens övergripande syfte, att stödja en likvärdig bedömning och betygssättning av elevers prestationer, kan mötas på ett mer verksamt och kostnadseffektivt sätt genom andra former av granskningar. Med utgångspunkt i detta har Skolinspektionen genom tidigare skrivelser, *Ökad kostnadseffektivitet i granskning av likvärdighet i bedömning och betygssättning*¹⁴ och *Budgetunderlag för åren 2017–2019*¹⁵, föreslagit att uppdraget om central rättning av nationella prov kan ändras så att arbetet med likvärdighet i bedömning och betygssättning förstärks genom att bedrivs som kvalitetsgranskningar.

Om ombedömningsuppdraget ska fortgå under och efter digitaliseringens införande behöver uppdraget utformas i förhållande till Skolverkets uppdrag att digitalisera de nationella proven.¹⁶ Dessa nya förutsättningar medför ett stort omställningsarbete för Skolinspektionen, främst i form av utveckling av it-lösning och behandling av sekretessfrågor, varför uppdragsbeskrivningen behöver anpassas i god tid. Vid extern bedömning av nationella prov är resultatet från ombedömningen inte längre på samma sätt relevant för Skolinspektionens tillsynsverksamhet på skolenhetsnivå.

¹⁴ Skolinspektionen (2015). PM Ökad kostnadseffektivitet i granskningen av likvärdighet i bedömning och betygssättning (SI diarienummer 2015:5148).

¹⁵ Skolinspektionen (2017). Budgetunderlag för åren 2017–2019 (SI diarienummer 2016:1582).

¹⁶ Regeringen (2017). Uppdrag att digitalisera de nationella proven m.m. (U2017/03739/GV).

Bilaga 1: Användning av resultaten i tillsyn och kvalitetsgranskning

I enlighet med regeringsuppdraget ska Skolinspektionen använda resultatet 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."¹⁷

Det övergripande syftet med analysen är att främja en rättvis och likvärdig betygssättning. Eftersom det inte finns några lagstadgade gränser för hur stora avvikelserna bör eller får vara, är det alltid processen som leder fram till resultatet som granskas i Skolinspektionens tillsyn, inte provbetygen i sig. En hög avvikelsegrad i förhållande till andra skolor kan däremot vara en indikation på att skolan behöver utveckla sina processer för bedömning och betygssättning och att området bör ägnas extra uppmärksamhet i samband med tillsyn.

Hur avgörs vad som är betydande avvikelser?

Vad som klassas som en "betydande avvikelse" är i grunden en bedömningsfråga. Det går inte att fastslå någon entydig definition av vad som är rätt svar för de frisvarsbaserade delproven.¹⁸ Skolinspektionen har därför valt att sätta ribban för vad som ska anses vara en betydande avvikelse relativt högt.

I ett första steg gäller det att för varje prov avgöra vad som är en avvikelse mellan skolans bedömning och Skolinspektionens omdömning. För delproven i svenska och engelska innebär en *betydande avvikelse* en skillnad på *två betygssteg eller mer* mellan skolans bedömning och Skolinspektionens omdömning. För delprov där bedömningen sker i form av belägg är gränsen satt till en skillnad på cirka 30 procent eller mer i antal utdelade belägg mellan de två bedömningarna.¹⁹

I ett andra steg gäller det att avgöra vad som är en avvikelse mellan skolans andel avvikelser från omdömningen och den genomsnittliga avvikelserna för

¹⁷ Utbildningsdepartementet (2015). Regleringsbrev för budgetåret 2015 avseende Statens Skolinspektion. U2013/5175/GV. U2014/7521/SAM (delvis). U2014/7535/GV.

¹⁸ Vilka delprov som ingår i denna omdömningsomgång framgår av tabell 1 i bilaga 2. De utvalda delproven består av uppsatsfrågor, där eleven fritt ska formulera sig utifrån en öppen frågeställning.

¹⁹ Motsvarar +/- 6 i skillnad på den harmoniserade skalan 0-20. I sammanvägningen väger dock A-belägg tyngre än C-belägg, som i sin tur väger tyngre än E-belägg. Beräkningen av den harmoniserade betygsskalan beskrivs i bilaga 3.

samtliga skolor som ingår i ombedömningen. De betydande avvikelserna i den ena eller andra riktningen ska vara mer vanligt förekommande vid den aktuella skolan än för genomsnittet av de ombedömda skolorna. Eftersom ombedömningens syfte är likvärdighet i bedömningen *mellan skolor* är det den enskilda skolans avvikelse från andra skolor, inte från ombedömningen i sig, som bör minimeras. En skolas resultat uppmärksammas om andelen betydande avvikelser, antingen i generösare eller striktare riktning, skiljer sig med mer än tio procentenheter från genomsnittet för de ombedömda skolorna för det aktuella delprovet. I denna ombedömningsomgång har drygt en tredjedel av skolorna betydande avvikelser på tio procent eller mer över medelvärdet för det aktuella provet.

Hur används resultaten i Skolinspektionens tillsyn och kvalitetsgranskning?

När en skola bedöms ha en stor andel betydande avvikelser mellan ursprungsbedömning och ombedömning skickas en signal till tillsynsverksamheten. Signalen innebär att Skolinspektionen särskilt ska granska skolans processer för bedömning och betygssättning vid nästkommande tillsynsbesök. Det är främst rektorns och huvudmannens arbete som granskas. Resultaten från ombedömningen är ett av flera underlag som Skolinspektionens inspektörer analyserar inför tillsyn av skolor.

Inför ett tillsynsbesök tar inspektören del av resultatet från ombedömningen, i de fall det finns ett sådant för skolan. Om ombedömningen visar en hög avvikelsegrad jämfört med andra skolor, antingen i mer strikt eller generös riktning, kan det vara en indikation på att skolan behöver utveckla sina processer för bedömning och betygssättning. Skolinspektionens inspektörer väger också in resultatet från Skolverkets jämförelse mellan provbetyg på det nationella provet och elevens ämnesbetyg eller kursbetyg. Båda dessa är risk signaler som tas upp i inspektörernas utredningsplaner och som gör att fler frågor kommer att ställas inom bedömning och betygssättning. Om en skola både har bedömt de nationella proven generöst jämfört med genomsnittet och generellt sätter höga ämnes- eller kursbetyg jämfört med provbetyg läggs extra stort fokus på skolans arbete med bedömning och betygssättning. Detsamma gäller i det omvända förhållandet, det vill säga när en skola har bedömt de nationella proven strikt jämfört med genomsnittet och dessutom generellt sätter låga ämnes- eller kursbetyg jämfört med provbetyg.

Det faktum att en skola har en hög avvikelsegrad jämfört med andra skolor i ombedömningen av nationella prov leder inte i sig till en brist, men om Skolinspektionen efter tillsynen bedömer att skolans arbetsätt kring bedömning och

betygsättning inte uppfyller de i lag och förordning fastställda kraven, föreläggs skolan att åtgärda bristerna. Vid eventuella brister följer Skolinspektionen upp att dessa har åtgärdats.

Skolinspektionen gör också regelbundet riktade tillsyner och kvalitetsgranskningar med särskilt fokus på bedömning och betygssättning.²⁰ Urvalet av skolor som ingår i dessa granskningar är ofta styrt av resultatet från omdömningen av nationella prov i kombination med resultatet från Skolverkets jämförelser mellan provbetyg och ämnesbetyg eller kursbetyg. För närvarande pågår en kvalitetsgranskning av huvudmäns och rektorers arbete för en rättvis och likvärdig betygssättning i grundskolan, som ska rapporteras i april 2018.²¹

Skol- och huvudmannarapporter

Resultatet för deltagande skolor och huvudmän inom omdömningen presenteras i särskilda skol- och huvudmannarapporter.²² Resultatet från omdömningen kan användas av rektorer och huvudmän på de skolor som har ingått i undersökningen i syfte att analysera och utveckla arbetet med bedömning och betygssättning.

²⁰ Se till exempel Skolinspektionen (2014): Uppenbar risk för felaktiga betyg, Skolinspektionen (2012): Riktad tillsyn av bedömning och betygssättning hos skolor med stora avvikelser vid omrättning av nationella prov, Skolinspektionen (2014): Förstärkt tillsyn av skolors arbete med bedömning och betygssättning samt Skolinspektionen (2013): Tillsyn av bedömning och betygssättning.

²¹ Skolinspektionen (2017): Direktiv Rättvis och likvärdig betygssättning i grundskolan.

²² De skolvisa rapporterna publiceras varje år på SIRIS: <http://siris.skolverket.se/siris/ris.omrattning.searchForm>

Bilaga 2: Genomförandet av omdedömningen

Urval av prov och skolor

2016 års omdedömning innefattar nationella prov som under året genomfördes i årskurs 6, årskurs 9, gymnasiets grundläggande kurser, samt motsvarande kurser inom den kommunala vuxenutbildningen på gymnasial nivå. För den kommunala vuxenutbildningen omdedömdes prov genomförda 2015.

Skolor som ingår i omdedömningen väljs ut genom ett slumpvist urval bland skolor vars huvudmän följande år ingår i Skolinspektionens regelbundna tillsyn. Urvalet av skolor för omdedömning år 2016 gjordes bland skolor vars huvudmän ingår i tillsynsplanen för 2017. Detta för att ett så aktuellt resultat som möjligt ska kunna användas som underlag vid tillsyn.

När ursprungs rättningen av proven genomfördes kände skolorna ännu inte till att de skulle ingå i urvalet för omdedömningen av nationella prov. Skolinspektionen bedömer därför att risken för att skolorna i förväg skulle kunnat ha manipulerat proven, med avsikt att uppnå ett fördelaktigt resultat, i princip är obefintlig.

Ämnen, delprov och årskurser som ingår i omdedömningen

För att kunna genomföra omdedömningen vid så många skolor som möjligt omdedöms endast ett eller två ämnen per skola, i regel ett delprov i svenska eller engelska i kombination med ett NO-ämne. En slumpvis tilldelning av ämnen görs bland de skolor som valts ut till omdedömningen. Därefter görs ett slumpvist urval bland de elever som genomfört provet den aktuella terminen. Urvalet görs tillräckligt stort för att resultaten ska vara representativa på skolnivå. Det maximala antalet prov som omdedöms har begränsats till i regel 100 för grundskolor och 200 för gymnasieskolor. Om fler prov krävs för att uppnå representativitet för två ämnen väljs endast ett ämne för den aktuella skolan.

Gymnasieskolorna har möjlighet att genomföra antingen en höstterminsversion eller en vårterminsversion av varje prov. Skolinspektionen har valt att endast omdedöma vårterminsversionen, då denna provversion skrivs av cirka tre fjärdedelar av landets elever och skolor och därmed ger ett bredare provunderlag att utgå ifrån.

Hela provet omdedöms inte, endast längre skrivuppgifter

Alla nationella prov är uppdelade i delprov. Vissa delprov innehåller laborativa eller muntliga element, eller uppgifter som genomförs i grupp. Denna typ

av delprov är svåra, om än inte omöjliga att omdöma. Tidigare erfarenhet har även visat att större avvikelser är mycket sällsynta för denna typ av prov.

Av ovanstående skäl omdöms inte hela prov, utan endast ett eller två delprov. De delprov som väljs ut för omdömning är sådana som är möjliga att omdöma (som saknar laborativa eller muntliga inslag), men där det inte finns tydliga "rätt eller fel svar" på frågorna. I praktiken betyder det att de prov som väljs till stor del består av skrivfrågor, där eleven fritt ska formulera sig utifrån en öppen frågeställning.

Urvalets sammansättning

Andelen omdömda prov från skolor med enskild huvudman som ingår i urvalet är högre än andelen enskilda skolor i riket vid tidpunkten för urvalet. Andelen omdömda skolor med enskild huvudman är 24 procent för grundskolor och 48 procent för gymnasieskolor. Vid tiden då urvalet gjordes uppgick andelen grundskolor respektive gymnasieskolor med enskild huvudman till 17 procent respektive 33 procent av det totala antalet grund- respektive gymnasieskolor i landet.

Terminologisk översikt

Skolinspektionen kan inte fastslå att omdömarnas bedömningar är mer korrekta än de bedömningar som ursprungs rättarna ute på skolorna gjort. I denna rapport talas det därför i termer av *avvikelser* mellan skolans ursprungsrättning och Skolinspektionens omdömning, snarare än korrekta eller felaktiga bedömningar. Med avvikelse avses en skillnad i bedömningen av en elevprestation med minst ett betygssteg på en standardiserad skala.²³

Generös bedömning: Ursprungsrättaren på skolan har bedömt att elevens svar motsvarar ett högre betyg än Skolinspektionens omdömare.

Strikt bedömning: Ursprungsrättaren på skolan har bedömt att elevens svar motsvarar ett lägre betyg än Skolinspektionens omdömare.

Betydande avvikelse: En avvikelse på två betygssteg eller mer mellan Skolinspektionens och skolans bedömning.

Då Skolinspektionen inte kan säga vilken av de båda bedömningarna som är mer "korrekt", läggs inte heller någon värdering i begreppen. Det är snarare ett

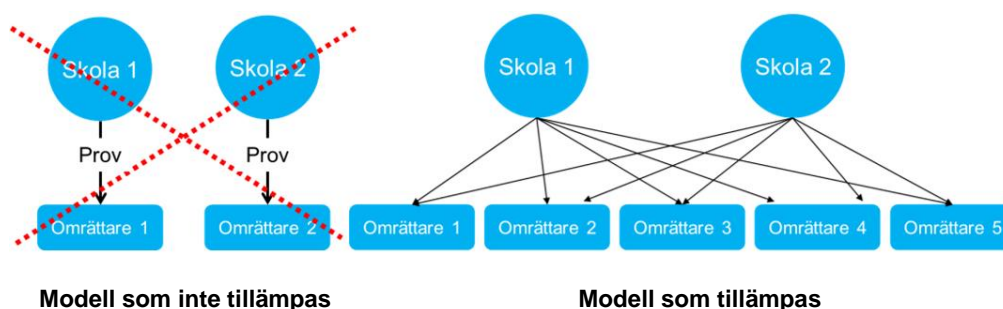
²³ Det harmoniserade måttet för mätandet av skillnader i bedömningen beskrivs i bilaga 3.

sätt att åskådliggöra i vilken riktning skolorna placerat sig i förhållande till Skolinspektionens omdömning.

Fördelning av prov mellan omdömare

Skolor som ingår i omdömningen ombeds skicka in kopior av bedömda prov till Skolinspektionen. Proven maskeras och avidentifieras därefter på personuppgifter, bedömningsanteckningar och andra uppgifter från ursprungs rättningen. Därefter bedöms proven på nytt av yrkesverksamma och erfarna lärare som rekryterats av Skolinspektionen. Ursprungsbedömningen jämförs sedan med omdömningen, och resultatet sammanställs i skolvisa rapporter.

Skolinspektionens lärare omdömer sina prov enskilt. Sambedomning tillämpas inte. För att minimera risken att en enskild omdömares bedömningar får stort genomslag i de sammantagna resultaten sprids elevlösningar för varje enskild skola ut bland flera olika omdömare. Jämförelsen av resultatet mellan skolor blir på så vis så rättvist som möjligt. Beroende på ämne och årskurs fördelas proven mellan 3 och 22 olika omdömare.



Figur 19: Schematisk bild över fördelningen av prov mellan omdömare. Prov från varje enskild skola fördelas så långt som möjligt ut till flera olika omdömare. Detta för att minimera risken för att den enskilda omdömarens bedömningar får för stort genomslag i de sammantagna resultaten.

För prov i svenska, där det skiljer mer än två betygssteg mellan bedömning och omdömning, går provet vidare till en andra omdömning. Det slutgiltiga betyget från omdömningen blir då resultatet från den av de två omdömningarna som ligger närmast ursprungsbedömningen. Syftet med den andra omdömningen är att kvalitetssäkra prov med betydande avvikelser för att garantera att Skolinspektionens lärare inte bedömt proven felaktigt.

Omfattning och bortfall

Totalt 26 927 delprov från 423 skolor har ombedömts i 2016 års omgång. Ombedömningen genomfördes av totalt 115 lärare som anställdes av Skolinspektionen. Fördelningen över ämne och delprov framgår av tabell 1 nedan.

Tabell 1: Externt och internt bortfall av delprov. Antal begärda delprov är det totala antalet delprov som begärdes in från skolorna.

	Antal begärda prov	Från antal skolor	Antal inkomna prov	Från antal skolor	Externt bortfall (prov)	Antal kompletta ombedömda prov	Från antal skolor	Internt bortfall (prov)	Totalt bortfall	Antal ombedömare
Totalt:	32 782	445	27 526	436	16 %	26 927	423	2 %	18 %	115
<i>Ämne och delprov:</i>										
Engelska åk 6 delprov C	3 228	110	2 894	106	10 %	2 851	103	1 %	12 %	8
Engelska åk 9 delprov C	4 453	93	3 814	89	14 %	3 725	85	2 %	16 %	20
Engelska Gy delprov C	5 781	102	4 737	100	18 %	4 629	97	2 %	20 %	22
Engelska Vux delprov C	379	25	261	24	31 %	192	12	26 %	49 %	3
Svenska åk 6 delprov C1+C2	3 115	108	2 756	106	12 %	2 682	101	3 %	14 %	12
Svenska åk 9 delprov C	5 523	98	4 780	95	13 %	4 692	92	2 %	15 %	20
Svenska Gy delprov C	5 803	104	4 745	103	18 %	4 663	100	2 %	20 %	21
Svenska Vux delprov C	652	25	421	22	35 %	391	18	7 %	40 %	5
Biologi åk 9 delprov A2	1 744	54	1 399	48	20 %	1 383	45	1 %	21 %	5
Fysik åk 9 delprov A2	879	28	714	26	19 %	714	26	0 %	19 %	6
Kemi åk 9 delprov A2	1 225	37	1 005	33	18 %	1 005	33	0 %	18 %	4

Bortfall

Det *externa bortfallet* utgörs av delprov som begärts in, men som inte inkommit från skolorna. Elevlösningar begärs in från skolorna utifrån föregående års elevlistor, varför till exempel elever som bytt skola under året utgör en del av det externa bortfallet. Till detta kommer elever som inte genomfört provet på grund av sjukdom eller frånvaro av annan orsak. Det händer även att skolor inte kan återfinna proven eftersom de inte har arkiverats enligt anvisningar till rektor.

Internt bortfall utgörs av prov som inkommit från skolorna, men som av olika orsaker inte kunnat analyseras. De vanligaste orsakerna är att provkopiorna håller för dålig kvalitet (kopieringen har inte tillräcklig svärta), att sidor saknas eller att resultatet inte rapporterats in på korrekt sätt.

Vad beror bortfallet på?

Det sammanvägda bortfallet varierar kraftigt från delprov till delprov. Bortfallet är särskilt högt för den kommunala vuxenutbildningen. Orsaken till varför bortfallet av prov är särskilt stort i den kommunala vuxenutbildningen kan delvis härledas till att det är frivilligt för de kommunala huvudmännen att skicka in kopior på bedömda nationella prov. Eftersom det totala antalet utvalda skolenheter och antalet insamlade prov är lågt, bör resultatet avseende proven som rör den kommunala vuxenutbildningen tolkas med försiktighet.

Bortfallsanalyser som gjorts i samband med tidigare ombedömningsomgångar har inte kunnat påvisa några systematiska bortfall. Då insamlingens genomförande inte förändrats över åren kan det även antas att inte heller detta förändrats.

Frågor om bedömningen på skolan

I samband med insamlingen av prov fick skolorna även möjlighet att svara på ett antal frågor om hur bedömningen av provet gick till. Frågorna besvarades för varje enskild insänd elevlösning. Det är frivilligt för skolorna att besvara frågorna, vilket i sin tur medför att antalet svar som kunnat analyseras är lägre än antalet inkomna prov.

I tabell 2 redovisas resultaten från de skolor som besvarat frågorna om bedömningens genomförande och som gett samtycke till att svaren får användas i analysyfte. Bortfallet utgörs av skolor som inkom med provkopior som gick att ombedöma, men som inte svarat på frågorna om bedömningens genomförande eller godkänt att svaren får användas i analysyfte.

För prov genomförda inom den kommunala vuxenutbildningen ställdes inga frågor om bedömningsprocessen och de är därför inte inkluderade i analysen.

Tabell 2: Bortfall vid frågor om bedömningens genomförande på skolorna.

Frågor om bedömningen:	1. Vem bedömde provet?			2. Var provet avidentifierat 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>Ämne och delprov:</i>									
Engelska åk 6 delprov C	2 340	88	18 %	2 717	100	5 %	2 663	99	7 %
Engelska åk 9 delprov C	2 700	68	28 %	3 560	83	4 %	3 353	82	10 %
Engelska Gy delprov C	3 796	84	18 %	4 201	92	9 %	3 772	91	19 %
Svenska åk 6 delprov C1+C2	2 052	85	23 %	2 552	97	5 %	2 478	97	8 %
Svenska åk 9 delprov C	3 500	80	25 %	4 160	88	11 %	3 871	88	17 %
Svenska Gy delprov C	3 816	89	18 %	4 324	96	7 %	3 644	96	22 %
Biologi åk 9 delprov A2	1 191	41	14 %	1 325	44	4 %	1 205	43	13 %
Fysik åk 9 delprov A2	633	24	11 %	708	26	1 %	628	26	12 %
Kemi åk 9 delprov A2	766	28	24 %	974	32	3 %	824	32	18 %

* Andel kompletta omedömda prov där svar på frågan saknas.

** Svaret "Minns ej" räknas som bortfall i tabellen.

Bilaga 3: 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 2 omdöms 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 endast avse det aktuella delprovet.

För delproven i svenska och engelska sätts delprovsbetyg på skalan A–F, men för delprov inom NO-ämnena är det endast det samlade provbetyget som anges på skalan A–F. För de olika delproven erhålls i stället ett antal *poäng* på E-, C- respektive A-nivå (för NO-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 omdömningsresultat för de vitt skilda proven behöver de olika bedömningarna därför 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 möjliga bedömning och 20 representerar högsta möjliga bedömning. Skillnaden mellan bedömning och omdöming kan då på motsvarande sätt uttryckas på skalan -20 till 20. Ett positivt värde indikerar att skolan satt ett högre betyg (delprov i svenska och engelska) eller gett fler belägg (NO) än omdömarens för en viss elevlösning. Ett negativt värde 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: A=20, B=16, C=12, D=8, E=4, F=0. En genomsnittsskillnad på 2 mellan bedömning och omdöming för en viss skola betyder alltså i dessa fall att det i snitt skiljer "ett halvt betygssteg" mellan skolans och omdömarens 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:
$$((E*10 + C*15 + A*20) / (\max(E)*10 + \max(C)*15 + \max(A)*20)) * 20.$$

E, C respektive A står för antalet belägg som skolan och ombedömaren 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 bedömningar som uttrycks i belägg (NO-prov) kan standardiseras till ett enskilt indexvärde på skalan 0–20, där A-belägg väger dubbelt så tungt som E-belägg i sammanräkningen. Om det maximala antalet belägg som eleven kan få för ett delprov är lågt väger varje enskilt belägg i framräkningen av det harmoniserade måttet desto tyngre.

En snabb approximation som underlättar den intuitiva förståelsen är att 2 på den harmoniserade skalan motsvarar tio procent av antalet möjliga belägg för det aktuella delprovet, 10 motsvarar 50 procent av antal möjliga belägg och så vidare. 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 som bygger på en sammanräkning av belägg endast är ett grovt 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 summering av belägg från flera olika delprov.

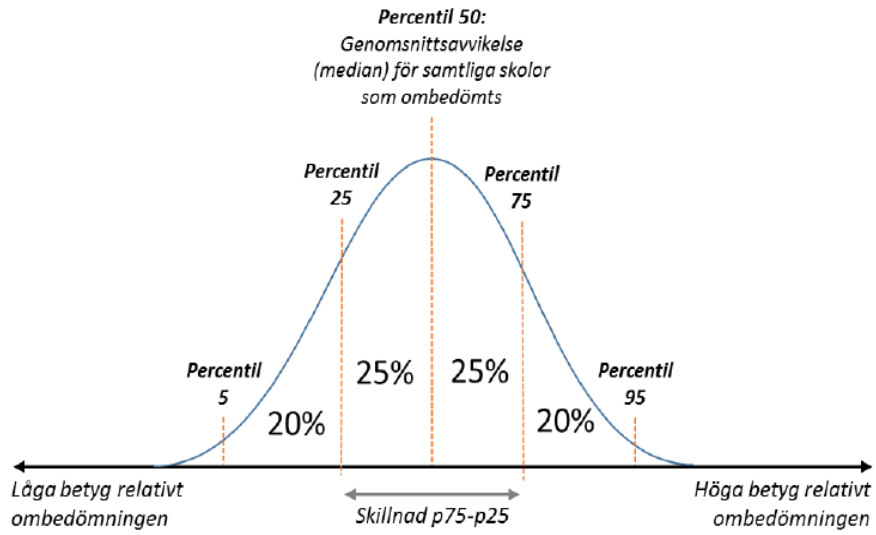
Bilaga 4: Metodologisk not avseende beräkning av variationen mellan skolor

Spridningen mellan olika skolor i bedömningen av samma elevlösning kan inte observeras direkt utifrån ombedömningen, eftersom de flesta skolor endast bedömer sina egna elevers prov. Resultatet från ombedömningen kan dock användas som en gemensam referenspunkt för att göra en indirekt skattning av skillnaderna mellan skolor. Eftersom proven från varje skola slumpas ut till samma grupp av ombedömare 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 därmed en indikation på 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. Skolinspektionen bedömer däremot att det ger en tillförlitlig bild av *storleksordningen* på skillnaderna mellan olika skolor. Mindre skillnader mellan olika delprov bör dock tolkas med försiktighet. Resultaten redovisas därför endast sammanslaget per ämnesgrupp.

För denna analys har även betygen på delprov i svenska och engelska räknats om till skalan 0–20, enligt principen F=0, E=4, D=8, C=12, B=16, A=20. För delproven i svenska och engelska är alltså konverteringen mellan den faktiska bedömningen och den harmoniserade skalan rättfram. En skillnad på 4 mellan två bedömningar 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 20 ger en schematisk illustration av jämförelsen mellan två percentilgrupper.



Figur 20: Schematisk illustration av jämförelsen mellan olika percentiler av skolor.

Bilaga 5: Referensförteckning

Betänkandet Likvärdigt, rättssäkert och effektivt – ett nytt nationellt system för kunskapsbedömning (SOU 2016:25).

Fredriksson, Peter & Vlachos, Jonas (2011). Reformen och resultat: Kommer regeringens utbildningsreformer att ha någon betydelse? Finanspolitiska rådet.

Regeringen (2017). Proposition Nationella prov – rättvisa, likvärdiga, digitala (Prop. 2017/18:14).

Regeringen (2017). Uppdrag att digitalisera de nationella proven m.m. (U2017/03739/GV).

Skolinspektionen (2012). Lika för alla? Omrättning av nationella prov i grundskolan och gymnasieskolan under tre år.

Skolinspektionen (2012): Riktad tillsyn av bedömning och betygssättning hos skolor med stora avvikelser vid omrättning av nationella prov.

Skolinspektionen (2013): Tillsyn av bedömning och betygssättning.

Skolinspektionen (2014): Förstärkt tillsyn av skolors arbete med bedömning och betygssättning.

Skolinspektionen (2014): Uppenbar risk för felaktiga betyg.

Skolinspektionen (2015). Ombedömning av nationella prov 2014: "Processerna spelar roll".

Skolinspektionen (2017): Direktiv Rättvis och likvärdig betygssättning i grundskolan.

Skolverket (2016). PM – Resultat på nationella prov i årskurs 3, 6 och 9, läsåret 2015/16.

Skolverket (2009). Likvärdig betygssättning i gymnasieskolan? En analys av samband mellan nationella prov och kursbetyg.

Skolverket (2011). Läroplan för grundskolan, förskoleklassen, och fritidshemmet. Reviderad 2016.

Utbildningsdepartementet (2015). Regleringsbrev för budgetåret 2015 avseende Statens Skolinspektion. U2013/5175/GV. U2014/7521/SAM (delvis). U2014/7535/GV.