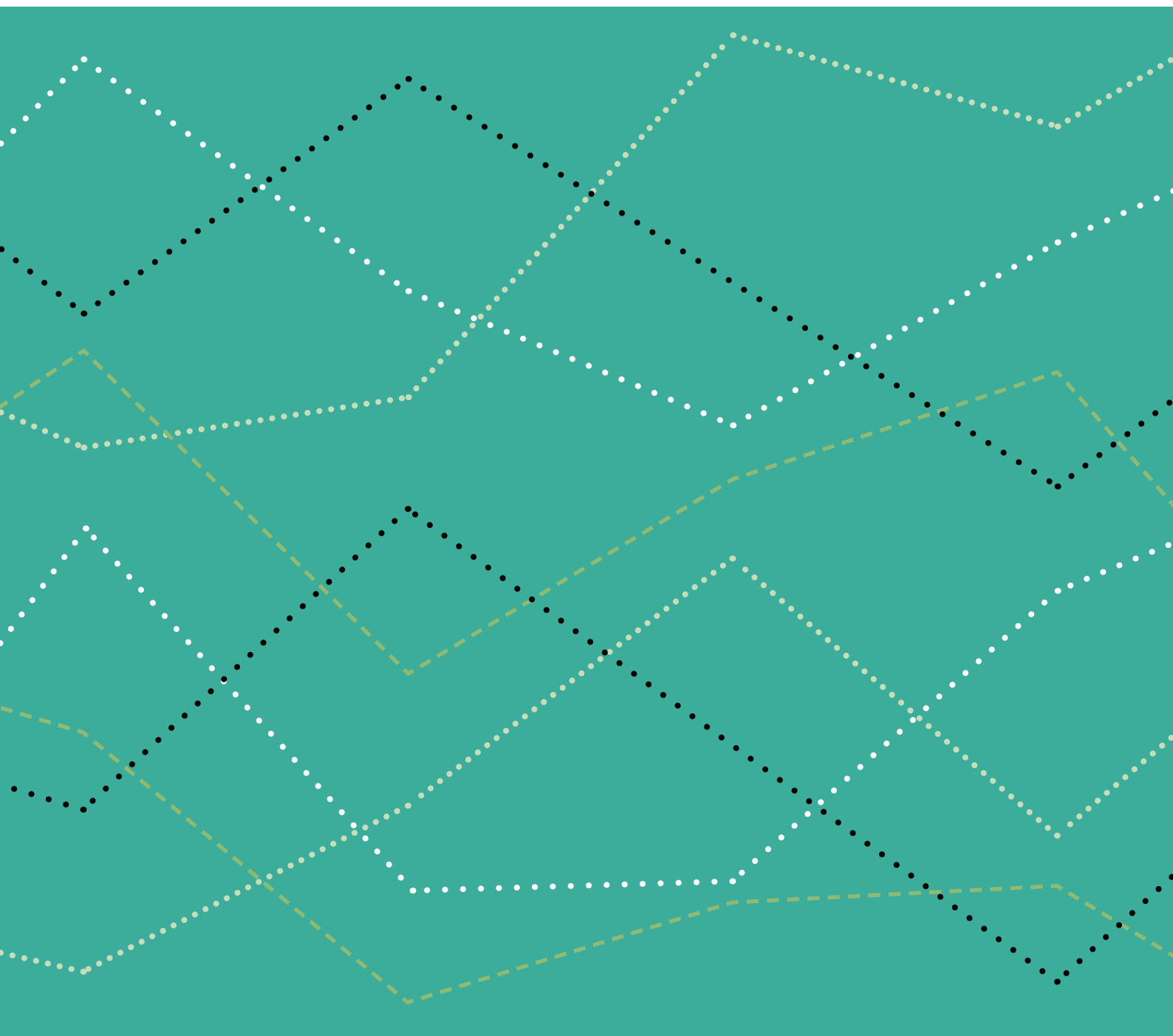


# Provbetygens stabilitet

Om nationella prov – Åk 9 1998–2012





# Provbetygens stabilitet

Om nationella prov – Åk 9 1998–2012

Publikationen finns att ladda ner som  
kostnadsfri PDF från Skolverkets webbplats:  
[skolverket.se/publikationer](http://skolverket.se/publikationer)

ISSN: 1652-2508

ISBN: 978-91-7559-182-7

Grafisk produktion: AB Typoform  
Skolverket, Stockholm 2015

## Förord

Den här rapporten ingår i en serie rapporter som handlar om de nationella proven. Rapporterna kan ha lite olika inriktning, somliga analyserar och diskuterar mer innehållsliga aspekter eller redovisar resultat och resultatsammanställningar som inte ingår i den löpande redovisningen av provresultat. Andra rapporter är inriktade mot mer principiella och i vissa fall tekniska frågeställningar kring prov och provkonstruktion. Denna typ av rapporter vänder sig främst till läsare som har ett särskilt intresse för prov och provkonstruktion i stor skala.

Den här aktuella rapporten syftar till att utifrån den samlade bilden av nationella prov genomförda i årskurs 9 åren 1998–2012 ge en beskrivning och analys av variationen i provbetyg under perioden. Det handlar då inte om den långsiktiga utvecklingen utan om de årliga variationer som kan antas bero på svårigheten att i förväg, innan provets genomförande, kunna bestämma över tid stabila gränserna för olika provbetyg. Syftet är inte att ge svar på vilka variationer som är acceptabla – det är varken en empirisk eller vetenskaplig fråga – utan att ge underlag för att fatta beslut om vad som kan ses som acceptabelt ur ett kvalitets-säkringsperspektiv.

Rapporten innehåller en hel del tabeller, men den som så önskar kan bortse från flertalet av dessa och nöja sig med den information som finns i de olika diagrammen. Eller, alternativt nöja sig med kapitlet *Sammanfattning, slutsatser och diskussionspunkter*.

Kristian Ramstedt har skrivit rapporten. Anders Auer, Lars-Åke Bäckman, Ingrid Jerkeman, Mattias Ludvigsson, Karin Nyqvist och Eva Wirén har lämnat värdefulla kommentarer.

Stockholm i april 2015

*Karin Hector-Stahre*  
Enhetschef

*Kristian Ramstedt*  
Expert



# Innehåll

## Sammanfattning 6

## Inledning 8

Feltyper och felkällor 9

Syfte 10

Metod, antaganden och underlag 11

## Resultat 13

Stabilitet i genomsnittlig betygspoäng 13

Stabilitet på betygsnivå 18

Provbetygens fördelning över tid 20

## Sammanfattning, slutsatser och diskussionspunkter 38

Hur stora är de genomsnittliga avvikelserna? 39

Vilken stabilitet kan anses tillräcklig? 40

Konsekvenser för den nya sexgradiga betygsskalan 40

Olika feltypers betydelse 44

De använda modellernas relevans 50

## Referenser 51

## Bilaga 1 52

## Sammanfattning

Statistik över betygen på de nationella proven visar att de tenderar att variera mellan olika år antingen man utgår från genomsnittlig betygspoäng (GBP) eller andel elever som får olika provbetyg. Syftet med den här studien är att mer systematiskt granska de nationella provbetygens årliga variation (brist på stabilitet) i ämnena svenska, engelska och matematik för årskurs 9 under den period 1994 års betygsskala gällt för nationella prov, det vill säga 1998–2012. Därmed fås ett bättre underlag för att bedöma provens kvalitet som underlag för en över tid stabil betygssättning.

Som referens för de årliga svängningarna har en till samtliga års resultat anpassad trendlinje använts, dels för GBP, dels för andel elever med respektive provbetyg. Den långsiktiga trenden vad gäller till exempel stigande eller fallande betygsnivå berörs inte utan endast de årliga svängningarna i relation till den långsiktiga trenden analyseras.

Resultaten visar att GBP är ett trubbigt instrument som inte medger någon mer ingående analys av till exempel hur många elever som berörs av de årliga svängningarna i provbetygen. Det beror på betygsskalans asymmetriska poängsättning. De analyser som baseras på variationer i andelen elever med olika provbetyg ger däremot möjlighet att bedöma såväl andel felklassificeringar som andel elever (hälften så stor)<sup>1</sup> som respektive år fått avvikande provbetyg i relation till den långsiktiga trenden.

Proven i engelska och svenska visar en genomsnittlig årlig avvikelse i i relation till trendvärdet på cirka 3–4 000 elever medan matematik visar en något högre avvikelse på drygt 6 000 elever per år (räknat på en årskull på 100 000 elever). Det betyder att så många elever per år beräknas bli klassificerade i en annan betygskategori än de skulle ha fått om provbetygen varit stabila i linje med den långsiktiga utvecklingen av betygen. Variationen i avvikelse mellan olika år kan dock i vissa fall vara betydande.

Man kan alltså konstatera att felklassificeringarna legat på 3–6 procent i genomsnitt, vilket betyder att cirka 2–3 procent av de elever som genomför ett nationellt prov fått ett provbetyg som avviker från det utifrån rådande trend förväntade. Någon nämnvärd ändring över tid har inte skett med avvikelserna, de har legat på ungefär samma nivå hela perioden. En betydande orsak till dessa variationer är svårigheten att i förväg fastställa betygsgränser och krav för olika betyg på ett prov, i synnerhet gäller det prov som är baserade på poäng. Där innebär varje poängsteg i allmänhet att andelen elever i berörda betygsgrupper ändras med flera procentenheter. En väsentlig fråga är om en total avvikelse på 3–6 procent ska bedömas som stor eller liten.

Vid en jämförelse med de övriga typer av fel som förekommer vid prov – slumpfel (*SEM*)<sup>2</sup> och avvikande bedömningar mellan olika rättare ("bedömningsfel") – framstår "stabilitetsfelen" som små. För ett normalt nationellt prov kan slumpfelet uppgå till cirka 20–30 procent, dvs. cirka 20–30 procent av

---

1 Om en betygsgrupp har för många elever och en annan för få blir det *klassifikationsfel* i båda grupperna. *Antalet elever* som fått fel betyg är dock endast de som utgör överskott i betygsgruppen med för många elever. De elever som finns i betygsgruppen med för få elever har alla korrekta betyg.

2 Standard Error of Measurement eller på svenska "den enskilda mätningens standardfel".



eleverna får av slumpskäl ett annat provbetyg än de skulle ha. Bedömningsfelet (olika rättare gör olika bedömningar) uppskattas leda till att cirka 10–30 procent av eleverna bedöms få avvikande betyg, med större andel för svårbedömda uppgifter (uppsatser etc.) och mindre för prov med korta mer entydigt besvarade uppgifter. Skattningen av bedömningsfel är dock mer osäker än skattningen av slumpfel och stabilitetsfel. Det är inte möjligt att inom ramen för de använda metoderna och utifrån använda data ange hur och i vilken utsträckning de olika felen påverkar varandra.

Slutligen visas att bytet från den fyrgradiga skala som gällde under den undersökta perioden till den sexgradiga som gäller efter 2011 innebär att provbetygen blir mer osäkra på alla aggregationsnivåer. Vissa felmarginaler kan antas öka med upp till hälften.

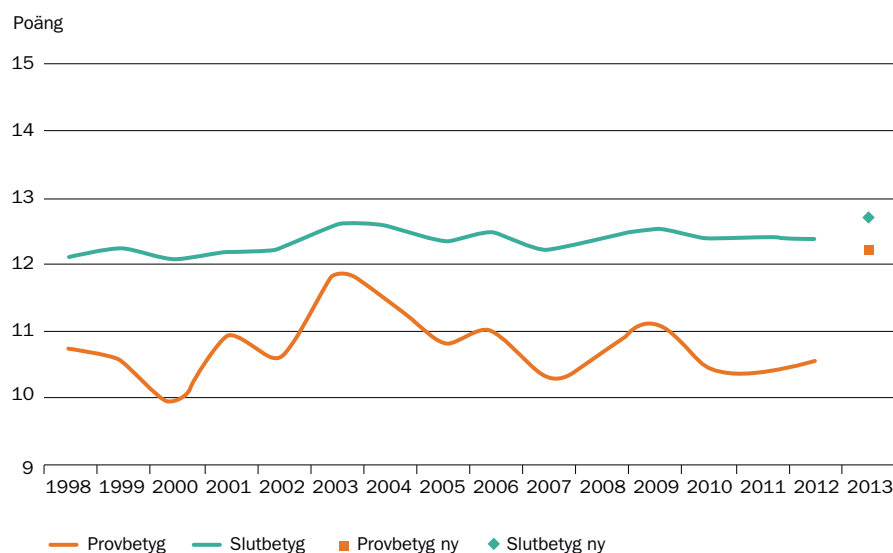
När det gäller stabilitetsfelets storlek är resultatet av granskningen att det varit i stort sett oförändrat under hela perioden. Detta kan tyda på att det finnas behov av ökat forsknings- och utvecklingsarbete för att öka provens stabilitet över tid. Eller alternativt att den stabilitet som gäller bedöms som acceptabel. I relation till övriga fel framstår kravgränsfelet som marginellt, i varje fall på elevnivå.

## Inledning

Hösten 1995 började den första årskullen elever med undervisning enligt Lpo 94 i årskurs 7. Enligt gällande grundskoleförordning skulle nationella ämnesprov ges i slutet av årskurs 9 i ämnena *svenska/svenska som andraspråk*, *engelska* och *matematik* och våren 1998 genomfördes de första proven baserade på Lpo 94. Proven genomfördes därefter årligen fram till och med våren 2012 då de sista gavs. Proven våren 2013 baserades på den nya läroplanen Lgr 11.

Skolverket har under alla år samlat in elevresultat i form av provbetyg. Åren 1998–2002 gällde insamlingen ett slumpmässigt urval av cirka 10 000 elevers betyg, vilket innebar ungefär 10 procent av den aktuella elevpopulationen. Från och med våren 2003 samlas alla elevers provbetyg in. Resultaten redovisas årligen och illustreras i diagram av det slag som visas i figur 1.<sup>3</sup> Den undre gröna linjen anger genomsnittlig betygspoäng<sup>4</sup> på provet och den övre svarta genomsnittlig betygspoäng för lärarnas betyg. I olika rapporter har relationen mellan provbetyg och lärarbetyg analyserats och diskuterats,<sup>5</sup> men i den här promemorian är det endast den gröna trendlinjen för provbetyg och dess variationer som är av intresse.

**Figur 1.** Genomsnittlig betygspoäng för provbetyg och lärarbetyg enligt 1994 års betygsskala samt 2011 års betygsskala (2013).



Syftet med de nationella proven har varierat en aning under den aktuella tidsperioden, men de huvudsakliga syftena har varit att proven ska vara ett *stöd för lärarnas betygssättning* samt ge *underlag för bedömning av måluppfyllelsen*. Det vill säga proven ska tjäna lärarna i deras uppgift att sätta rättvisa och likvärdiga betyg samt tillhandahålla resultat som kan vara grund för att bedöma och värdera utfallet av skolans arbete på övergripande nationell nivå, på huvudmannanivå och på skolnivå, och därmed tjäna som konsumentupplysning i ett alltmer konkurrensutsatt skolsystem.

3 Ur Skolverket (2014)

4 Förklaras senare.

5 T.ex. Skolverket (2004, 2007)

Som resultatmått redovisar Skolverket andelen elever som fått respektive betyg på provet, dvs. Godkänd (G), Väl godkänd (VG) och Mycket väl godkänd (MVG). För elever som inte nått G redovisas inget betyg. Denna kategori betecknas här EUM ("ej uppnått målen"). Ovanstående gäller 1994 års betygssystem. Efter 2011 tillämpas den nya sexgradiga betygsskalan<sup>6</sup> vilket innebär att proven 2013 är de första med denna typ av provbetyg. Den här promemorian gäller dock främst provbetyg enligt 1994 års betygsskala.

## Feltyper och felkällor

Nationella prov och andra prov är alltid behäftade med olika typer av fel. Prov syftar till att generera utsagor från elever utifrån vilka det sedan är möjligt att dra slutsatser om elevers kunskaper inom det kunskapsområde provet gäller. Inget prov kan dock inom rimlig tid täcka in hela kunskapsområden av någorlunda omfattning. Varje prov blir därmed ett stickprov av alla frågor som kan ställas. Detta är en felkälla för att dra slutsatser om en elevs kunskaper. Beroende på vilka frågor provet ställer kan elevens resultat med lite tur leda till en överskattning av förmågan, med lite otur till underskattning. En annan felkälla kan vara att en elev råkar ha en dålig dag. En ytterligare felkälla kan vara att den bedömning som görs av elevens svar, i synnerhet om de är mer omfattande, är olika för olika bedömare.

Detta är felkällor av *slumpmässig natur* som är kopplade till prov som underlag för att bedöma *enskilda* elever. Den typen av fel kan uppskattas med hjälp av olika metoder<sup>7</sup> baserade på antingen klassisk testteori eller, vilket är mer ovanligt, modern testteori.<sup>8</sup> Dessa teorier gäller främst prov som är uppbyggda av många förhållandevis korta uppgifter ("item"). Sådana fel, "den enskilda mätningens standardfel",<sup>9</sup> skattas vanligen utifrån de reliabilitetsmått som används inom respektive provteori. Detta är också fel som i huvudsak kan sägas ligga på *individnivå*, även om *bedömningsfel* kan vara kopplade till bedömare som bedömer hela klasser av elever.

Den felkälla som gäller själva bedömningen av ett svar är ofta ett mindre problem för korta uppgifter, men ju mer komplexa och omfattande svaren är desto större tenderar variationen mellan olika bedömare att bli. Detta är också en form av slumpartat fel beroende på vilken bedömare eleven, eller en grupp elever råkat få. Den här typen av fel uttrycks vanligen i termer av interbedömaröverensstämmelse eller interbedömarreliabilitet. Eftersom det i Sverige oftast är den egna läraren som bedömer proven kan man säga att denna typ av fel gäller på *gruppnivå*, förutsatt att läraren bedömer alla prov efter samma normer.<sup>10</sup>

En svårighet vid bedömningar av mer komplicerade uppgifter är att bedömningsanvisningarna inte kan ange specificerade rätta svar utan endast mer allmänna kriterier för bedömning och exempel på olika typer av svar. Bedömaren har alltså att tolka såväl elevens utsaga som bedömningsanvisningarna. Skol-

6 A till F, där F innebär ej godkänt resultat och A är det högsta betyget.

7 Vilka finns beskrivna i standardlitteraturen på området t.ex. Crocker & Algina (1986)

8 Item response theory (IRT)

9 "Standard error of measurement", se t.ex. Crocker & Algina (1986).

10 Detta kan man dock inte ta för givet eftersom det är väl belagt att bedömare som känner dem som gjort provet tenderar att påverkas av denna kännedom (haloeffekter). Om detta är en fördel eller nackdel när det gäller hur väl bedömningen speglar elevens förmåga kan diskuteras.

inspektionen genomför kontrollrättning av nationella prov som visar exempel på sådana bedömningsvårigheter och på de osäkerhetsmarginaler som råder, såväl mellan ursprungs rättare och kontrollrättare som mellan olika kontrollrättare.<sup>11</sup> Denna feltyp kommer dock inte att närmare granskas i den här rapporten.

En ytterligare typ av fel är mer systematiska och allmängiltiga, dvs. de är inte främst relaterade till enskilda elever eller deras lärare (bedömare) utan gäller alla provdeltagare. I mål- och kunskapsrelaterade<sup>12</sup> prov av det slag som i dag används i Sverige är den mest påtagliga och konkreta felkällan relaterad till *betygsättningen* av provresultaten. Den princip som gäller är att betygsgränser och betygskrav ska vara fastställda när proven genomförs, det är alltså inte möjligt att avvakta provresultaten innan betygskraven fastställs. Att fastställa betygsgränser under sådana villkor är en grannliga uppgift som åligger de universitetsinstitutioner som på Skolverkets uppdrag konstruerar de nationella proven.<sup>13</sup>

Ett systematiskt fel<sup>14</sup> i provresultatet eller provbetyget skulle inte ha någon nämnvärd betydelse om betygen endast skulle gälla för jämförelser mellan elever och elevgrupper det aktuella året, eftersom alla då drabbas av samma fel. Om resultaten däremot ska användas i sammanhang där de jämförs med provresultat från andra år får systematiska fel betydelse, till exempel när elever från olika år konkurrerar om samma utbildningsplatser. Om betygsgränserna legat för lågt ett år gynnar det elever som fått provbetyg det året i relation till elever som fått sina betyg ett år när betygsgränserna legat högre (dvs. varit svårare att nå).

Någon systematisk granskning av provbetygens årliga variationer eller svängningar har inte gjorts utöver beskrivningar och illustrationer av det slag som visas i figur 1. Det finns därför goda skäl att undersöka resultaten från de prov som genomförts och till exempel försöka ange hur många elever som berörs av de årliga variationerna i provbetyg. Den från 2011 nya betygs-skalan med fler betygssteg gör frågan än mer angelägen eftersom varje betygsgräns måste ses som en potentiell felkälla, vilket gör att fler betygsgränser kan antas leda till fler felkällor och större sammantaget fel.

## Syfte

Den här studien handlar främst om en felkälla på systemnivå, nämligen den som gäller provbetygens stabilitet mellan olika år. Studien syftar i första hand till att sammanfatta och analysera *provbetygens* utveckling och främst deras *variation* under den period nationella prov funnits inom ramen för de mål- och kunskapsrelaterade betygsystemet enligt 1994 års utformning. En ambition med och en viktig kvalitetsaspekt för nationella prov är att både de slumpmässiga och systematiska felen ska vara så små som möjligt. För att åstadkomma detta krävs systematiska studier av tidigare resultat.

Ett annat syfte med studien är att ta fram ett underlag som kan argument för att fastställa *acceptabla felgränser*. Det vill säga vilken nivå är rimlig nivå för hur stor andel av eleverna som kan få fel provbetyg. Detta är dock inte främst en empirisk fråga utan ett normativt beslut som bör fattas när resultat föreligger.

11 Skolinspektionen(2013)

12 Brukar i internationell litteratur benämnas "criterion referenced" eller "standards referenced".

13 Det finns olika metoder för att göra detta, men det skulle leda för långt att här gå in på dessa.

14 Ofta kallad "bias".

Syftet med promemorian innebär således att sammanställa och analysera provresultaten i årskurs 9 under perioden 1998–2012 med särskilt fokus på provbetygens *stabilitet* över tid. Har stabiliteten förändrats? Är stabiliteten lika för prov i olika ämnen? Finns det några intressanta mönster?

Stabilitet har i det här sammanhanget två innebörder, dels i vilken utsträckning genomsnittlig betygspoäng för proven varierar olika år i relation till ett på alla års resultat modellbaserat mått, dels hur andelen elever som får respektive provbetyg förhåller sig olika år. En *liten variation* i genomsnittlig betygspoäng mellan olika (närliggande) år och en liten variation mellan andelen elever som får respektive betyg ses som en indikation på *hög stabilitet*. Provresultat mäts således i andel elever med olika betyg olika år och kan antingen anges sammanfattande som *genomsnittlig betygspoäng* eller uppdelat som *andelen elever med respektive provbetyg*. Båda redovisningsformerna kommer att användas.

## Metod, antaganden och underlag

Prov i ämnena *svenska*, *engelska* och *matematik* för årskurs 9 som funnits under hela den aktuella perioden utgör underlag för studien, men även de 2009 införda proven i naturorienterande ämnen (NO-ämnen) för årskurs 9 tas upp kortfattat.

Som underlag används den redovisning av provbetyg som publiceras på Skolverkets hemsida.<sup>15</sup> Redovisningen anger i procent hur stor andel av provdeltagarna som fått respektive betyg. För åren 2002–2004 redovisas inget samlat provbetyg<sup>16</sup> i svenska och engelska på hemsidan och därför anges inga data för dessa år.

Data ordnas för respektive ämne i en tidsserie och en trendlinje anpassas till de årliga värdena. Residualen (skillnaden) mellan modellvärde för ett visst år och det observerade värdet för samma år används sedan som underlag för att beräkna ett mått på stabiliteten (eller dess motasats; variationen, svängningarna, volatiliteten eller vad man väljer att kalla de kortsiktiga variationerna). En trendlinje används för att få en modellbaserad referenspunkt att mäta de årliga värdena mot. Därmed fås ett mått på provbetygens stabilitet. Förändring över längre tid, till exempel beroende på betygsinflation, kunskapsminskning eller kunskapsförbättring ingår inte i syftet för den här granskningen utan det är *den årliga variationen* som står i fokus.

Det är inte rimligt anta att elevernas faktiska kunskaper varierar mellan olika årskullar i någon större utsträckning och därmed inte heller att resultat på de nationella proven ska visa några större variationer mellan närliggande år. Med de stora grupper och det långa perspektiv det handlar om – cirka 100 000 provdeltagare som under nio år i skolan läst de aktuella ämnena – sker knappast tvärsnitt i den samlade kunskapsmängden hos den aktuella elevgruppen mellan enskilda år. Antagandet är därför att de förändringar som sker, sker långsamt och visar sig i de trender man kan se över perioder på flera år. När det sker snabba förändringar mellan närliggande år tolkar vi det i det här sammanhanget som en form av mätfel, i den meningen att betygsgränser har fått placeringar

<sup>15</sup> <http://siris.skolverket.se/siris/f?p=SIRIS:150:0::NO::>

<sup>16</sup> Under 2002–2004 fanns inga centrala anvisningar om hur delprovresultaten skulle sammanfattas till ett samlat provbetyg.

som ger resultat som avviker från den långsiktiga trenden.<sup>17</sup> Därigenom kan de ses som en form av systematiskt mätfel eftersom det är ett fel som gäller lika för alla provdeltagare (men får effekt för dem som har provresultat i anslutning till en betygsgräns).

De andra två typerna av fel, de slumpmässiga mätfel som gäller på individnivå och rättarnas bedömningsfel på gruppnivå gäller förstås också, men härvidlag antar vi att dessa fel är konstanta över tid på nationell nivå och att positiva och negativa fel tar ut varandra på nationell nivå. Dessa typer av fel antas därför inte inverka på de aggregerade provbetygens årliga avvikelser från trendlinjen.<sup>18</sup>

---

17 Viss del av felet kan förstås också bero på att ingen poänggräns passar precis.

18 Ett diskutabelt antagande, men det får gälla tills vidare.

## Resultat

Först ges en översiktlig bild av den genomsnittliga betygspoängen för alla tre provämnena och därefter redovisas varje ämne för sig.

### Stabilitet i genomsnittlig betygspoäng

Tabell 1 visar som exempel en sammanställning av andelen elever med olika provbetyg i svenska årskurs 9 för åren 1998–2012 samt beräknad genomsnittlig betygspoäng.<sup>19</sup>

**Tabell 1.** Sammanställning av andel elever med olika provbetyg 1998–2012 samt beräknad genomsnittlig betygspoäng, Svenska åk 9.

År	Rapporterade värden				
	EUM	G	VG	MVG	GBP
1998	5,9	45,7	40,8	7,6	12,2
1999	4,0	49,0	40,0	7,0	12,3
2000	4,0	49,0	37,0	9,0	12,3
2001	4,1	48,7	39,1	8,1	12,4
2002					
2003					
2004					
2005	5,1	44,8	40,9	9,1	12,4
2006	3,8	42,6	42,3	11,3	12,9
2007	3,7	46,0	41,7	8,6	12,6
2008	3,3	40,8	45,2	10,7	13,0
2009	3,3	44,9	43	8,9	12,7
2010	5,1	47,9	39,6	7,4	12,2
2011	2,7	39,1	48,5	9,8	13,1
2012	2,6	37,7	50,1	9,6	13,2

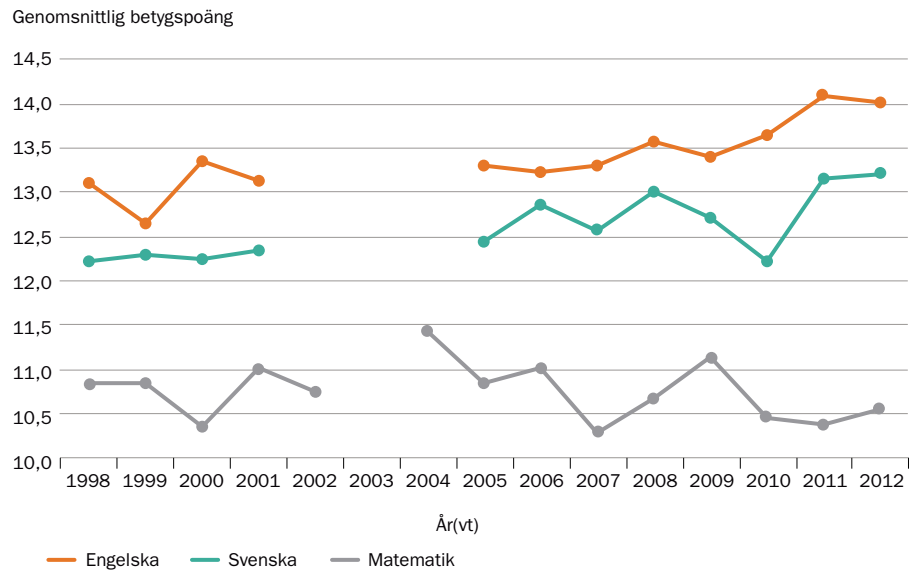
I tabellen kan man notera att andelen elever med EUM och G tycks sjunka medan andelen elever med VG (i synnerhet) respektive MVG tenderar att öka. Även den genomsnittliga betygspoängen förefaller öka, vilket förstås överensstämmer med betygsutvecklingen. Det är dock svårt att bedöma och värdera stabilitetens utveckling enbart utifrån tabell 1.

I figur 2 anges genomsnittlig betygspoäng för de tre provämnena svenska,<sup>20</sup> engelska och matematik i diagramform.

19 Vid beräkningen av GBP används formeln  $GBP = (0 * p(EUM) + 10 * p(G) + 15 * p(VG) + 20 * p(MVG)) / 100$  där  $p(EUM)$  är andel elever med "betyget" EUM i procent osv.

20 Elever med provbetyg i Svenska som andraspråk ingår inte i underlaget.

**Figur 2.** Genomsnittlig betygspoäng på nationella prov 1998–2012. Åren 2002–2004 redovisades inget samlat provbetyg i svenska och engelska.



I figuren kan man notera att proven i engelska varje år har den högsta genomsnittliga betygspoängen. Därefter följer svenska och matematik, som har de lägsta värdena. Man kan också notera att matematik förefaller ha den största variationen mellan olika år, medan variationen för engelska och svenska tycks mer likartad.<sup>21</sup>

Av figuren får man också intrycket att det skett en ökning av den genomsnittliga betygspoängen över tid för engelska och svenska men inte för matematik.

Figur 2 ger en något mer tolkningsbar bild än tabell 1, men även figur 2 är svår att tolka mer i detalj, till exempel när det gäller hur många elever som berörs av svängningarna i betygspoäng.

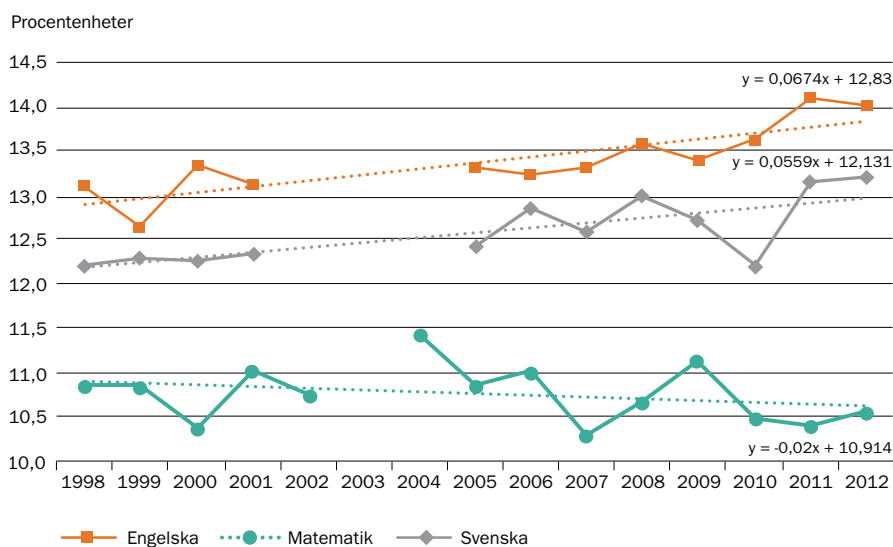
För att ytterligare underlätta bedömningen av såväl den långsiktiga utvecklingen som de årliga variationerna kan man lägga in en trendlinje för respektive prov i diagrammet.<sup>22</sup> Figur 3 visar resultatet.

21 Värdet för matematik år 2003 är missvisande eftersom det på grund av rykten om att provet fanns tillgängligt på nätet var ett stort bortfall i resultatredovisningen. Bortfallet var inte slumpmässig utan en övervikt av skolor med låga provresultat avstod från redovisning. Varför den genomsnittliga betygspoängen i svenska avviker 2010 är mer oklart.

22 Värdet framför x i linjen ekvation är ett mått på linjens lutning. Minustecken ”-” anger att värdet sjunker med tiden.



**Figur 3.** Genomsnittlig betygspoäng olika år med trendlinje och ekvation.

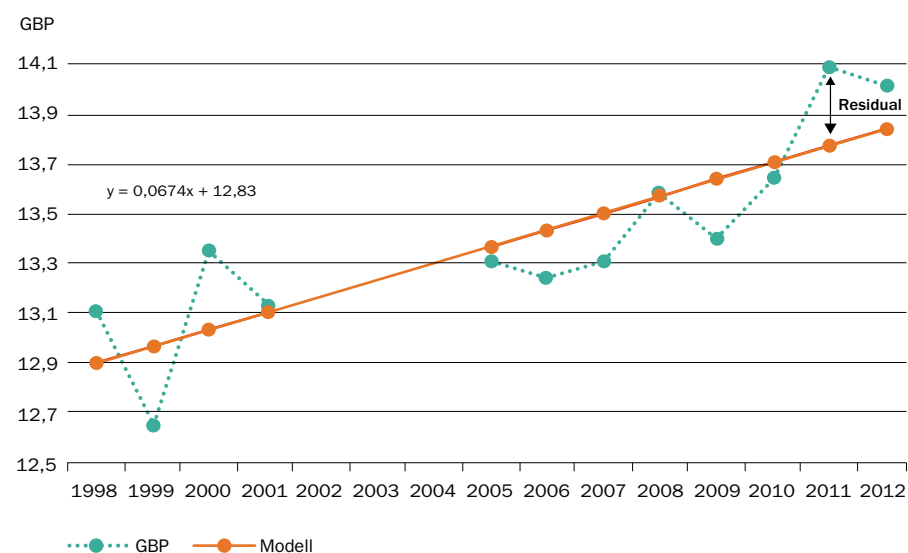


Figuren stärker synintrycket att engelska haft den största ökningen och att matematik<sup>23</sup> haft en nedåtgående trend.<sup>24</sup>

Med hjälp av trendlinjen är det möjligt att beräkna ett mått på avvikelser mellan trendlinjen och det för året aktuella värdet på den genomsnittliga betygspoängen GBP. Detta kan göras med hjälp av ekvationen för trendlinjen. Skillnaden (*residualen*) mellan den aktuella genomsnittliga betygspoängen och det beräknade "modellvärdet" kan då ses som ett mått på skillnaden det aktuella året.

För exempelvis engelska ser en figur med högre upplösning ut som figur 4. Den blå pilen vid 2011 representerar residualens storlek detta år. En skattning utifrån diagrammet tyder på att residualen är cirka 0,3 mätt i genomsnittlig betygspoäng (GBP).

**Figur 4.** Genomsnittlig betygspoäng olika år med trendlinje, ekvation och residual för 2011. Engelska



<sup>23</sup> Resultatet för 2003 uteslutet.

<sup>24</sup> Det framgår även av koefficienterna framför x i ekvationerna där matematik har ett negativt värde.

Att skatta värden ur en graf är dock en ganska osäker metod. Bättre då att  
a) beräkna trendlinjens ekvation och det värde den ger för ett aktuellt år samt  
b) beräkna värdet på GBP samma år utifrån de på Skolverkets hemsida angivna  
fördelningarna av provbetyg (tabell 1). Skillnaden mellan de två typerna av  
värdena ger då residualen för det aktuella året. I tabell 2 finns en sammanställ-  
ning av aktuella värden (GBP), beräknade modellvärden och skillnader mellan  
dessa (räknat som *rapporterat* värde minus modellvärde, G-M),<sup>25</sup> för de tre  
ingående ämnena över den aktuella tidsperioden.

**Tabell 2.** Aktuell genomsnittlig betygspoäng (GBP-S) och modellbaserad GBP (Modell) samt skillnaden mellan dessa (G-M).

År	Svenska			Engelska			Matematik		
	GBP-S	Modell	G-M	GBP-S	Modell	G-M	GBP-S	Modell	G-M
1998	12,2	12,2	0,02	13,1	12,9	0,21	10,8	10,9	-0,05
1999	12,3	12,2	0,06	12,7	13,0	-0,31	10,9	10,9	-0,02
2000	12,3	12,3	-0,05	13,4	13,0	0,32	10,4	10,9	-0,50
2001	12,4	12,4	0,00	13,1	13,1	0,04	11,0	10,8	0,17
2002							10,7	10,8	-0,07
2003									
2004							11,4	10,8	0,66
2005	12,4	12,6	-0,14	13,3	13,4	-0,06	10,8	10,8	0,09
2006	12,9	12,6	0,23	13,2	13,4	-0,20	11,0	10,7	0,28
2007	12,6	12,7	-0,12	13,3	13,5	-0,19	10,3	10,7	-0,43
2008	13,0	12,7	0,25	13,6	13,6	0,01	10,7	10,7	-0,03
2009	12,7	12,8	-0,08	13,4	13,6	-0,24	11,1	10,7	0,46
2010	12,2	12,9	-0,65	13,6	13,7	-0,07	10,5	10,7	-0,19
2011	13,1	12,9	0,23	14,1	13,8	0,32	10,4	10,6	-0,25
2012	13,2	13,0	0,24	14,0	13,8	0,17	10,5	10,6	-0,07
Sum(ABS)			2,07			2,14			3,28
Std			0,24			0,21			0,31

Av tabellen framgår exempelvis (liksom av figur 2 och 3) att avvikelserna i svenska är störst 2010 då det aktuella värdet ligger 0,65 genomsnittliga betygspoäng under modellvärdet (G-M=-0,65). För engelskans del är den största avvikelserna +0,32 GBP (2000 och 2011) och -0,31 (1999) och för matematiken ±0,61 (2000) och 0,58 (2004).

Sum(ABS) anger summan av avvikelsernas absolutbelopp (alla värden räknas som positiva). Där framgår att de samlade avvikelserna är störst för matematik. Detsamma gäller för standardavvikelsen som anger att den årliga variationen i avvikelser från modellvärdet är störst i matematik (0,31) och ungefär lika stor i svenska (0,24) och engelska (0,21), (om vi bortser från resultatet i svenska 2010 blir skillnaden betydligt mindre).

<sup>25</sup> Dvs Residualen = GBP(rapporterat)-Modellvärde = G-M

## Hur värdera skillnader i GBP olika år?

Den fråga man ställer sig är hur resultatet i tabell 2 ska eller bör värderas. Är skillnaderna små eller stora? Vad innebär de när det gäller hur många elever som får olika betyg olika år? Och är genomsnittlig betygspoäng ett lämpligt mått för att bedöma betygsvariationen (eller betygsstabiliteten som kan ses som betygsvariationens motsats)?

Ett problem med den genomsnittliga betygspoängen är att den bygger på en asymmetrisk skala där EUM värderas till 0 betygspoäng, G till 10, VG till 15 och MVG till 20 betygspoäng. Detta innebär att betygen får en oregelbunden viktning vid beräkningen av genomsnittlig betygspoäng ("avståndet" mellan EUM och G blir dubbelt så stort som mellan G och VG respektive VG och MVG). Detta gör tolkningen av GBP svår, i synnerhet om man vill veta hur många elever som berörs av de olika årens variationer. Om till exempel GBP stiger eller faller 0,2 "GBP-enheter" kan detta inte förstås i konkreta termer av typen antal elever med högre betyg eller liknande. En höjning från EUM till G ger 10 betygspoäng, medan exempelvis en höjning från G till VG eller från VG till MVG endast ger 5 betygspoäng. GBP duger som indikation, men är i övrigt ett svårtolkat mått. Av dessa skäl ska vi i nästa avsnitt titta närmare på de enskilda betyg som ingår i GBP.

Men först lite om användningen av en trendlinje som modell. Hur tillförlitlig är den?

## Trendlinjen som modell

Trendlinjen är en regressionslinje, som bestäms i Excel med minsta kvadratmetoden enligt gängse metoder för linjär regression. Vanligtvis används regressionslinjer för att göra prediktioner. Så är dock inte fallet här utan här används den för att få en referenslinje och referenspunkter mot vilken årliga variationer kan jämföras, en form av linjär modell för tidsserier. Man kan förstås fråga sig om en regressionslinje är relevant för ett sådant syfte och om det är rimligt att anta att betygstrender är linjära. Det finns dock ingenting i de tidigare figurerna över GBP som tyder på att GBPs utveckling över tid inte skulle vara åtminstone approximativt linjär över de aktuella åren. Eftersom statistiken, i varje fall efter 2003 baserar sig på hela årskurser på cirka 100 000 elever är det inte heller troligt att det skulle förekomma särskilt stora variationer i kunskapsutvecklingen, och därmed betygsfördelningen, mellan närliggande år. Antagandet om en linjär regression (trend) för betyg och genomsnittlig betygspoäng förfaller därför plausibel.

Figur 4 tyder inte heller på att trenden skulle avvika för åren 1998–2002 då resultaten baserades på stickprov av elever (ca 10 000) och inte heller förefaller de tre år då anvisningar för sammanfattande betyg inte gavs, och sådana inte samlades in i svenska och engelska (2002–2004), markera någon avvikelse i trenden.<sup>26</sup> Såväl samplingsfel som eventuella andra statistiska fel torde därmed

26 För 2002–2004 redovisas resultat för de ingående delproven "läsförståelse", "muntlig förmåga" och "skriftlig förmåga" i svenska och "muntlig förmåga", "receptiv förmåga" och "skrivande förmåga" i engelska. Skolverket har i vissa sammanhang redovisat sammanfattande provresultat där 2002–2004 ingår och har då beräknat ett sammanfattande provbetyg enligt tidigare gällande regler. T.ex. gäller det för 2012 års "Redovisning av uppdrag om avvikelser mellan provresultat och betyg i grundskolans årskurs 9" (Dnr 75-2012:311)

vara försumbara i relation till de mer uppenbara felkällor som finns i proven och som vi återkommer till.

Den här aktuella analysen baseras på aggregerade data som redovisas på Skolverkets hemsida och exemplifieras i nästa avsnitt. Naturligtvis kunde det vara en fördel att ha tillgång till mikrodata men för den här studiens syfte bedöms det aggregerade underlaget vara tillfyllest.

## Stabilitet på betygsnivå

Som underlag används som framgått provdata från Skolverkets hemsida. Redovisningen har delvis sett olika ut genom åren. Den var inledningsvis tämligen sparsmakad och kan exempelvis se ut som tabellen nedan för år 2001 då resultat samlades in från ett stickprov av skolor.

**Figur 5.** Procentuell fördelning av sammanvägda provbetyg i svenska, matematik och engelska år 9 våren 2001

Ämne	Andel (%) elever som erhållit visst provbetyg eller ej nått målen i ämnet enligt provet											
	Totalt				Pojkar				Flickor			
	G	VG	MVG	Ej nått målen	G	VG	MVG	Ej nått målen	G	VG	MVG	Ej nått målen
Svenska/ svenska som andraspråk <sup>2</sup>	49	39	8	4	58	32	5	5	39	47	12	3
Matematik	51	27	9	13	51	26	10	13	51	27	9	13
Engelska	39	45	13	4	42	43	10	5	35	47	15	3

### Förklaringar och definitioner

Samtliga uppgifter avser elever i år 9 vårterminen 2001

\* Ämnesprov årskurs 9

Fr o m vårterminen 1998 ges inom ramen för det nationella provsystemet varje år obligatoriska ämnesprov i år 9 i svenska, matematik och engelska. Fr.o.m. vår-terminen 2001 ska ämnesprovet i svenska även användas för elever som läser svenska som andraspråk. Syftet med ämnesproven skall vara att kontrollera om eleverna uppnått kursplanernas kunskapskrav.

\* Sammanvägt provbetyg

Varje ämnesprov består av ett antal delprov. Resultatet på de olika delproven vägs sedan ihop till ett provbetyg. För att få ett sammanvägt provbetyg ska eleven gjort samtliga delprov.

- 1) Skattade nationella nivåer baserade på provresultat från ett representativt urval av 150 skolor.
- 2) Resultatet för svenska som andraspråk särredovisas ej, pga. för få inrapporterade elevresultat.

Närmast föregående publicering: Skolverkets rapport nr 195.

Efter 2003, då samtliga elevers resultat började samlas in, har informationen efterhand blivit betydligt mer omfattande. För matematikens del ser redovisningen ut som nedan 2012.

**Figur 6.** Procentuell fördelning av provbetyg i matematik år 9, våren 2012.

Matematik	Totalt							Flickor					Pojkar					
	Antal elever som gjort provet		Andel (%) av elever som gjort provet				Antal elever som gjort provet		Andel (%) av elever som gjort provet			Antal elever som gjort provet		Andel (%) av elever som gjort provet				
	Antal elever	gjort provet	G	VG	MVG	Ej nått målen	Antal elever	gjort provet	G	VG	MVG	Ej nått målen	Antal elever	gjort provet	G	VG	MVG	Ej nått målen
Provresultat <sup>2</sup>	100 521	92 543	47,7	25,4	9,8	17,0	49 183	45 513	46,7	26,4	10,2	16,7	51 338	47 030	48,8	24,5	9,4	17,4
därav elever med svensk bakgrund	80 869	75 502	47,6	26,8	10,6	14,9	39 629	37 112	46,4	28,1	11,1	14,4	41 240	38 390	48,8	25,7	10,1	15,5
utländsk bakgrund	19 652	17 041	48,2	19,2	6,3	26,2	9 554	8 401	47,6	19,2	6,5	26,7	10 098	8 640	48,8	19,3	6,1	25,8
Totalt exklusive Nynivandrade elever <sup>4</sup>	97 764	90 819	47,8	25,6	9,8	16,7	47 985	44 746	46,7	26,6	10,3	16,4	49 779	46 073	48,9	24,7	9,4	17,0
Nynivandrade elever	2 757	1 724	43,9	16,0	7,1	33,0	1 198	767	44,5	15,9	6,9	32,7	1 559	957	43,5	16,0	7,3	33,2

2) I matematik ges inga delprovsvbetyg. För att få ett provbetyg i matematik ska eleven ha deltagit i samtliga delprov.

4) Nynivandrade elever har kommit till Sverige de senaste 4 åren. De har inte bott i Sverige eller gått i svenska skola tidigare.

I svenska och engelska redovisas även betyg på olika delprov förutom ett sammanfattande provbetyg (kallat ”provresultat” i tabellen).<sup>27</sup> Det är dessa sammanfattande provbetyg som används i den fortsatta undersökningen. Under åren 2002–2004 redovisades som nämnts inga sammanfattande provbetyg. Skälet för detta angavs i de då gällande föreskrifterna vara att det var lärarnas uppgift att sammanfatta delprovsvbetygen utifrån sina egna bedömningar av hur de borde värderas.

I tabellen ovan från 2012 kan man lägga märke till att provresultat saknas för cirka 8 procent av eleverna. Detta är en rätt normal andel. De nationella proven är obligatoriska för läraren att använda, men har inte varit obligatoriska för den enskilda eleven och en elev som har varit sjuk vid något provtillfälle kan därför bli utan provbetyg.<sup>28</sup> Jämförelser av lärarnas betyg på elever som deltagit i provet respektive inte deltagit visar genomgående att elever som deltagit har högre betygsmedelvärden. Om detta beror på att elever med stor sjukfrånvaro får sämre kunskaper och större sannolikhet att vara sjuk vid ett provtillfälle eller på något annat är inte känt. Men eftersom bortfallet av provresultat inte är slumpmässigt är det viktigt vid jämförelser av provbetyg och lärarbetyg att endast ta med de elever som har båda betygen. Det görs till exempel i Skolverkets jämförelser. I det här sammanhanget ingår dock endast provbetygen och om bortfallen sker av samma orsaker (vilket antas) under den aktuella perioden kan de förbises i den här studien.

Låt oss då se på de enskilda provbetygens fördelning och utveckling över tid.

27 Se bilaga 1 för tabeller i engelska och svenska.

28 Den förordning (eller tolkningen av den) som styr detta har sedermera förändrats så att proven numera ses som en obligatorisk del av undervisningen och därmed är obligatoriska också för eleven.

## Provbetygens fördelning över tid

I följande avsnitt kommer samma metod att tillämpas som för den genomsnittliga betygspoängen. Det innebär att den *procentuella andelen* elever med olika provbetyg de aktuella åren används som underlag för en modellerad tidsserie varefter residualer beräknas baserade på rapporterade värden och modell-anpassade värden. Först redovisas betygsfördelningen i svenska, därefter i engelska och matematik.

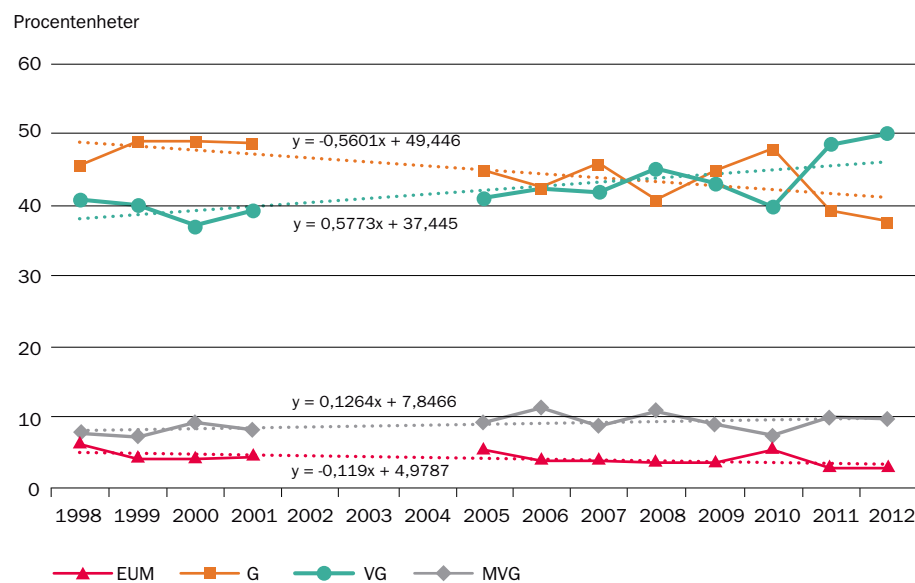
### Svenska

I figur 7 illustreras fördelningen av provbetyg åren 1998–2012 i ämnet svenska. Förutom de aktuella procentensatserna innehåller diagrammet också trendlinjer (streckade) med vidhängande ekvationer<sup>29</sup> för vart och ett av betygen.<sup>30</sup>

Man kan se att G och VG har varit de vanligaste betygen där mellan 40 och 50 procent av eleverna haft respektive betyg. Det framgår också att andelen elever med betyget G har minskat över tid, vilket syns av att trendlinjen lutar neråt och av att koefficienten  $k$  framför "x" är negativ ( $k=-0,56$  avrundat till två decimaler)<sup>31</sup>. Andelen elever med VG har omvänt stigit i samma grad ( $k=0,58$ ).

Andelen elever med EUM och MVG ligger också nära varandra. Andelen elever med MVG har stigit svagt mot 10 procentenheter (med en genomsnittlig ökning på drygt en tiondels procentenhet per år ( $k=0,13$  procentenheter/år). Andelen elever med "provbetyget" EUM har i motsvarande grad sjunkit ( $k=-0,12$ ). mot cirka 3 procentenheter 2012.

**Figur 7.** Procentuell fördelning av provbetyg olika år med trendlinje och ekvation. Svenska åk 9.



29 Ekvationerna finns med därför ett de används för beräkningar senare.

30 Trendlinjens ekvation är av typen  $y = kx + m$ . Där anger värdet på y-axeln (i det här fallet uttryckt i "procentenheter") för olika värden på x som representeras av år på x-axeln (1998 får värdet 1, 1999 värdet 2 osv.).  $k$  är en konstant som är ett mått på linjens lutning och värdet ges i diagrammet.  $m$  är en annan konstant som anger värdet där linjen korsar y-axeln.

31 Det betyder att andelen elever med betyget G har sjunkit med drygt en halv procentenhet (0,56) per år.

**Tabell 3.** Aktuell andel elever med rapporterade provbetyg i vänstra tabellen. Modellbaserade värden i den högra. Svenska åk 9.

Rapporterade värden						Modellbaserade värden					
År	EUM	G	VG	MVG	GBP	Löpnr	EUM	G	VG	MVG	GBP
1998	5,9	45,7	40,8	7,6	12,2	1	4,9	48,9	38,0	8,0	12,2
1999	4,0	49,0	40,0	7,0	12,3	2	4,7	48,3	38,6	8,1	12,2
2000	4,0	49,0	37,0	9,0	12,3	3	4,6	47,8	39,2	8,2	12,3
2001	4,1	48,7	39,1	8,1	12,4	4	4,5	47,2	39,8	8,4	12,4
2002						5					
2003						6					
2004						7					
2005	5,1	44,8	40,9	9,1	12,4	8	4,0	45,0	42,1	8,9	12,6
2006	3,8	42,6	42,3	11,3	12,9	9	3,9	44,4	42,6	9,0	12,6
2007	3,7	46,0	41,7	8,6	12,6	10	3,8	43,8	43,2	9,1	12,7
2008	3,3	40,8	45,2	10,7	13,0	11	3,7	43,3	43,8	9,2	12,7
2009	3,3	44,9	43	8,9	12,7	12	3,6	42,7	44,4	9,4	12,8
2010	5,1	47,9	39,6	7,4	12,2	13	3,4	42,2	44,9	9,5	12,9
2011	2,7	39,1	48,5	9,8	13,1	14	3,3	41,6	45,5	9,6	12,9
2012	2,6	37,7	50,1	9,6	13,2	15	3,2	41,0	46,1	9,7	13,0
Medel	4,0	44,7	42,4	8,9	12,6	Medel	4,0	44,7	42,4	8,9	12,6
Std	1,0	3,9	3,9	1,3	0,4	Std	0,6	2,6	2,7	0,6	0,3

Graferna ger ett något rörigt intryck och tjänar knappast som underlag för närmare granskning och bedömning. För att kunna göra en mer tillförlitlig analys behöver vissa beräkningar göras genom att de observerade andelarna elever med olika betyg jämförs med de modellbaserade värdena. Tabell 3 visar de andelar som rapporteras på Skolverkets hemsida och motsvarande modellbaserade värden.

Ovanför den högra tabellen finns de konstanter som använts för att beräkna modellvärdena. De är hämtade ur diagrammet i figur 7. Man kan notera att värdena i de två delarna av tabellen på det hela taget stämmer rätt väl överens, men också att det blir skillnader i GBP för vissa år, exempelvis 2006, 2008 och särskilt 2010.<sup>32</sup>

Standardavvikelsen<sup>33</sup> indikerar att de största förändringarna över tid skett för betygen G och VG där andelen G minskat (från 48,9 till 41,0 procent enligt modellen) och andelen VG ökat från 38,0 till 46,1 procent, vilket också framgår av figur 7. För EUM och MVG har ändringarna varit mindre.

Det intressanta i sammanhanget är emellertid främst hur stora *de årliga variationerna* är uttryckta i andel elever som klassificerats i en betygskategori som avviker från den de i realiteten placerats i. För att få en tydligare bild av

32 Man kan också notera att den genomsnittliga betygspoängen för de modellbaserade värdena stämmer väl med den trendlinje som finns i figur 3.

33 Standardavvikelsen är här beräknad i relation till medelvärde för respektive betyg över perioden 1998–2012.

**Tabell 4.** Skillnad i procentandelar av provdeltagare som i relation till modellvärden fått "för låga" (negativa värden) eller "för höga" (positiva värden) provbetyg. "Sum(ABS)" anger summan av absolutvärdena. Den vänstra tabellen anger procentenheter. Den högra anger antalet avvikelser om man antar att årskursen är 100 000 elever (med värden avrundade till 100-tal, elever 50-tal). Svenska åk 9.

År	EUM	G	VG	MVG	Sum(ABS)
1998	1,04	-3,19	2,78	-0,37	7,38
1999	-0,74	0,67	1,40	-1,10	3,91
2000	-0,62	1,23	-2,18	0,77	4,81
2001	-0,40	1,52	-0,64	-0,30	2,85
2002					
2003					
2004					
2005	1,07	-0,17	-1,16	0,24	2,64
2006	-0,11	-1,81	-0,34	2,32	4,57
2007	-0,09	2,16	-1,52	-0,51	4,27
2008	-0,37	-2,48	1,40	1,46	5,72
2009	-0,25	2,18	-1,37	-0,46	4,26
2010	1,67	5,74	-5,35	-2,09	14,84
2011	-0,61	-2,50	2,97	0,18	6,27
2012	-0,59	-3,34	4,00	-0,14	8,08
Summa	0,00	0,00	-0,01	0,00	69,62
Medel	0,00	0,00	0,00	0,00	5,80
Std	0,80	2,76	2,63	1,15	3,29

År	EUM	G	VG	MVG	Sum(ABS)	Antal elever
1998	1 000	-3 200	2 800	-400	7 400	3 700
1999	-700	700	1 400	-1 100	3 900	1 950
2000	-600	1200	-2200	800	4 800	2 400
2001	-400	1500	-600	-300	2 800	1 400
2002						
2003						
2004						
2005	1 100	-200	-1 200	200	2 700	1 350
2006	-100	-1 800	-300	2300	4 500	2 250
2007	-100	2 200	-1 500	-500	4 300	2 150
2008	-400	-2 500	1 400	1500	5 800	2 900
2009	-300	2 200	-1 400	-500	4 400	2 200
2010	1 700	5 700	-5 300	-2100	14 800	7 400
2011	-600	-2 500	3 000	200	6 300	3 150
2012	-600	-3 300	4 000	-100	8 000	4 000
Sum(ABS)	7 600	27 000	25 100	10 000	69 700	34 850
Medel(ABS)	600	2300	2100	800	5 800	2 900
Std	800	2 800	2 600	1 200	3 300	1 600

\* På grund av avrundningsfel kan Sum(ABS) i den vänstra och högra tabellen ge lite olika värden.

detta visar tabell 4 nedan i den vänstra tabellen *skillnaden* mellan de observerade andelarna och de modellberäknade andelarna uttryckta i *procentenheter*. I den högra delen av tabellen räknas värdena (skillnaderna) i stället om till *antal felklassificeringar*. Därvid utgår från en antagen "standardpopulation" på 100 000 elever, vilket ungefär brukar motsvara en årskull i Sverige.

"Sum(ABS)" anger summan av absolutantalet avvikelser från modellvärdet.<sup>34</sup> Både avvikelse uppåt och neråt är avvikelse och därför blir den totala avvikelserna summan av absolutvärdena. Observera dock att summan av andelarna i de fyra betygskategorierna är 100 procent. Det betyder om man skulle flytta en betygsgräns så att andelen ökar i en betygskategori, måste den minska motsvarande i någon annan. Flyttas till exempel 100 elever från kategori IG till kategori G minskar den ena gruppen med 100 och ökar den andra med 100. Det har alltså skett två ändringar med 100 (totalt 200 i klassificeringarna) men endast 100 elever har i realiteten bytt betyg. En förändring av en poänggräns skulle alltså i realiteten endas påverka *hälften* så många elever som de som anges i tabellerna. I tabell 4 är alltså totalt 7 400 felklassificeringar noterade 1998. En justering av poänggränserna så att den totala avvikelserna blir noll skulle således endast påverka 3 700 elever.

Man kan i högra delen av tabell 4 notera att för till exempel år 1998 var det

<sup>34</sup> Dvs. Sum(ABS) är den summa som fås om alla värden som adderas räknas som positiva.



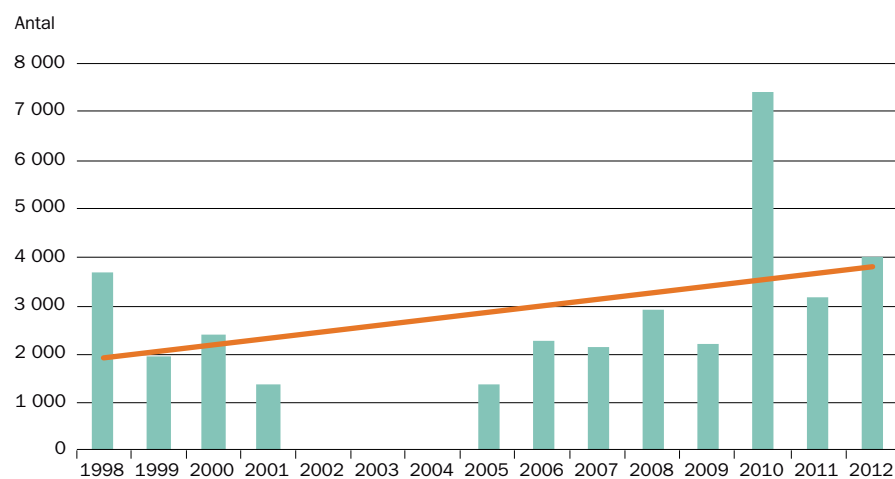
cirka 1 000 elever fler än modellvärdet som inte uppnådde godkänd nivå (EUM) medan 3 200 elever färre än förväntat fick G (av dessa fick alltså 1 000 EUM). Å andra sidan var det 2 800 fler än väntat som fick VG (varav 2 200 borde fått G), och 400 färre än väntat som fick MVG (och i stället antagligen fick VG). Och så vidare för andra år.<sup>35</sup>

I genomsnitt är det cirka 2900 elever per år som fått ett för högt eller för lågt provbetyg om man tar modellvärdet som utgångspunkt. Antalet elever med ”fel provbetyg” varierar årsvis mellan cirka 1 350 (år 2005) och 7 400 (år 2010) med standardavvikelsen<sup>36</sup> 1 550 elever, men 2010 års provbetyg sticker då ut markant. Det är svårt att ur tabellen se någon tydlig trend i variabiliteten, eller variabilitetens motsats – stabiliteten, över tid.

Figur 8 ger en bild av utvecklingen. Observera också att när vi här talar om ”för högt”, ”för lågt” eller ”fel” provbetyg och ser det som en avvikelse från modellvärdet säger vi ingenting om vilka felkällor som kan finnas och vad som kan tänkas förklara avvikelsen. Till detta återkommer vi. Här utgår vi endast från inrapporterade data och antar tillsvidare att de är korrekta liksom vårt antagande om ett giltigt linjärt modellvärde.

Även om man bortser från 2010 års extrema resultat visar trendlinjen i figur 8 på en svag ökning i variabiliteten. Det finns med andra ord ingenting som tyder på att provbetygen i svenska för årskurs 9 har blivit mer stabila under de 15 år proven användes i 1994 års system. Genomsnittsvärdet för avvikelserna ligger på 5,8 procentenheter och standardavvikelsen på 3,8 (tabell 4)<sup>37</sup>, vilket uttryckt i antal elever med fel betyg innebär ett genomsnitt på cirka 2 900 elever per år och en standardavvikelse på 1 550 elever.

**Figur 8.** Antal elever (i en antagen årlig population på 100 000) som fått ett i förhållande till modellvärdet avvikande provbetyg. Svenska åk 9.



I följande avsnitt görs motsvarande redovisning för provbetygen i engelska och därefter för matematik.

35 På grund av avrundningsfel förekommer avvikelser vissa år (summan av avvikelser ≠ noll)

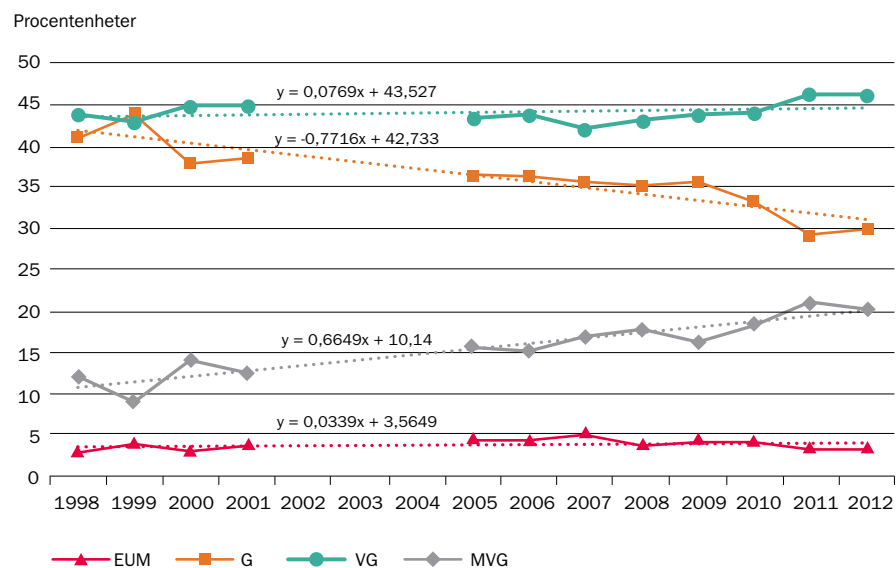
36 Det kan uttryckas som att två tredjedelar av eleverna ligger i intervallet  $2\,900 \pm 1\,550$ .

37 Standardavvikelsen dras dock upp av 2010 års extremvärde.

## Engelska

Förändringen av andelen elever med olika provbetyg i engelska uppvisar nedanstående mönster (figur 9).

**Figur 9.** Procentuell fördelning av provbetyg olika år med trendlinje och ekvation. Engelska åk 9.



Liksom fallet var för proven i svenska är bilden något rörig och svårtolkad även i engelska. Resultaten redovisas därför även här i tabellform (tabell 5).

**Tabell 5.** Aktuell andel elever med rapporterade provbetyg i vänstra tabellen. Modellbaserade värden i den högra. Engelska åk 9.

Rapporterade värden						Modellbaserade värden							
År	EUM	G	VG	MVG	GBP	k =	m =	Löpnr	EUM	G	VG	MVG	GBP
1998	2,9	41,0	43,9	12,1	13,1	0,0339	3,5649	1	3,6	42,0	43,6	10,8	12,9
1999	4,0	44,0	43,0	9,0	12,7	-0,7716	42,733	2	3,6	41,2	43,7	11,5	13,0
2000	3,0	38,0	45,0	14,0	13,4	0,0769	43,527	3	3,7	40,4	43,8	12,1	13,0
2001	3,7	38,6	45,1	12,6	13,1	0,6649	10,14	4	3,7	39,6	43,8	12,8	13,1
2002								5					
2003								6					
2004								7					
2005	4,4	36,6	43,4	15,7	13,3			8	3,8	36,6	44,1	15,5	13,4
2006	4,5	36,5	43,8	15,1	13,2			9	3,9	35,8	44,2	16,1	13,4
2007	5,0	35,8	42,2	17,0	13,3			10	3,9	35,0	44,3	16,8	13,5
2008	3,7	35,2	43,2	17,9	13,6			11	3,9	34,2	44,4	17,5	13,6
2009	4,2	35,7	43,8	16,3	13,4			12	4,0	33,5	44,4	18,1	13,6
2010	4,2	33,3	44,2	18,4	13,6			13	4,0	32,7	44,5	18,8	13,7
2011	3,3	29,3	46,3	21,1	14,1			14	4,0	31,9	44,6	19,4	13,8
2012	3,3	30,1	46,3	20,3	14,0			15	4,1	31,2	44,7	20,1	13,8
Medel	3,9	36,2	44,2	15,8	13,4			Medel	3,9	36,2	44,2	15,8	13,4
Std	0,6	4,1	1,3	3,5	0,4			Std	0,2	3,8	0,4	3,3	0,3

**Tabell 6.** Skillnad i procentandelar av provdeltagare som i relation till modellvärden fått "för låga" (negativa värden) eller "för höga" (positiva värden) provbetyg. "Sum(ABS)" anger summan av absolutvärdena. Den vänstra tabellen anger procentenheter. Den högra anger antalet avvikelser om man antar att årskursen är 100 000 elever (med värden avrundade till 100-tal, elever 50-tal). Engelska åk 9.

År	EUM	G	VG	MVG	Sum(ABS)	År	EUM	G	VG	MVG	Sum(ABS)	Antal elever
1998	-0,70	-0,96	0,30	1,30	3,25	1998	-700	-1 000	300	1 300	3 300	1 650
1999	0,37	2,81	-0,68	-2,47	6,33	1999	400	2 800	-700	-2 500	6 400	3 200
2000	-0,67	-2,42	1,24	1,87	6,19	2000	-700	-2 400	1 200	1 900	6 200	3 100
2001	0,04	-1,05	1,23	-0,20	2,53	2001	0	-1 100	1 200	-200	2 500	1 250
2002						2002						
2003						2003						
2004						2004						
2005	0,56	0,04	-0,74	0,24	1,59	2005	600	0	-700	200	1 500	750
2006	0,63	0,71	-0,42	-1,02	2,78	2006	600	700	-400	-1 000	2 700	1 350
2007	1,10	0,78	-2,10	0,21	4,19	2007	1100	800	-2 100	200	4 200	2 100
2008	-0,24	0,95	-1,17	0,45	2,81	2008	-200	1 000	-1 200	400	2 800	1 400
2009	0,23	2,23	-0,65	-1,82	4,92	2009	200	2 200	-600	-1 800	4 800	2 400
2010	0,19	0,60	-0,33	-0,38	1,50	2010	200	600	-300	-400	1 500	750
2011	-0,74	-2,63	1,70	1,65	6,72	2011	-700	-2 600	1 700	1 700	6 700	3 350
2012	-0,77	-1,06	1,62	0,19	3,64	2012	-800	-1 100	1 600	200	3 700	1 850
Summa	0,00	0,00	0,00	0,00	46,45	Sum(ABS)	6 200	16 300	12 000	11 800	46 300	23 150
Medel	0,00	0,00	0,00	0,00	3,87	Medel(ABS)	500	1 400	1 000	1 000	3 900	1 950
Std	0,62	1,68	1,21	1,31	1,81	Std	600	1 700	1 200	1 300	1 800	900

\* På grund av avrundningsfel kan Sum(ABS) i den vänstra och högra tabellen ibland skilja sig åt.

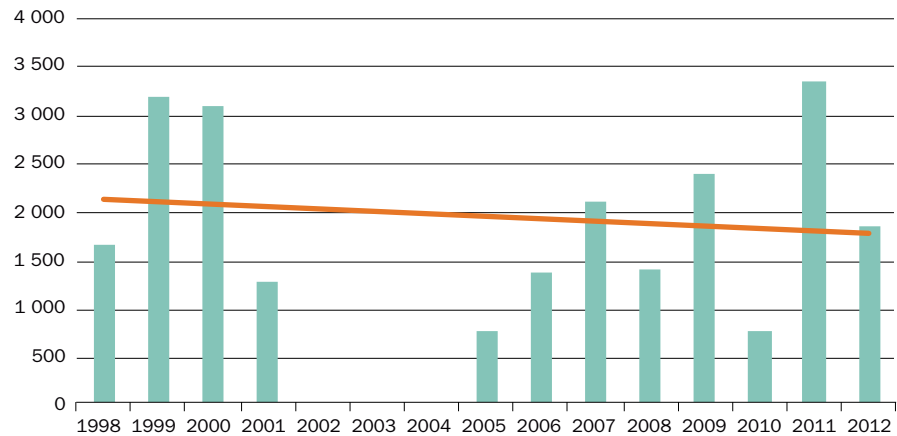
I likhet med provbetygen i svenska ger modellvärdena något andra värden på GBP jämfört med de värden som fås av inrapporterade data. Detta är rimligt med tanke på den asymmetriska betygsskalan. Man kan även för engelskans del notera att förändringen över tid i huvudsak rör två betyg, vilket framgår av standardavvikelsen för modellvärdena, nämligen andelen G som minskat från drygt 40 procent av eleverna till drygt 30 och andelen MVG som ökat från drygt 10 procent till cirka 20. För svenskans del var det andelen VG som ökade med cirka 10 procentenheter.

En sammanställning av förändringarna visas i tabell 6. Här kan man konstatera att det i genomsnitt är cirka 3 900 elever per år som i relation till modellvärdet får för lågt eller högt provbetyg. Värdena varierar mellan 1 500 elever och 6 700<sup>38</sup> med en standardavvikelse på cirka 1 800 elever.

Även för engelskan är det svårt att ur tabellen se om stabiliteten förändrats över tid. Figur 10 visar antalet elever per år som fått för högt eller lågt betyg i relation till modellvärdet samt trenden för stabiliteten.

38 Under antagandet att varje årskull omfattar 100 000 elever.

**Figur 10.** Antal elever (i en antagen årlig population på 100 000) som fått ett i förhållande till modellvärdet avvikande provbetyg. Svenska åk 9,  
Antal

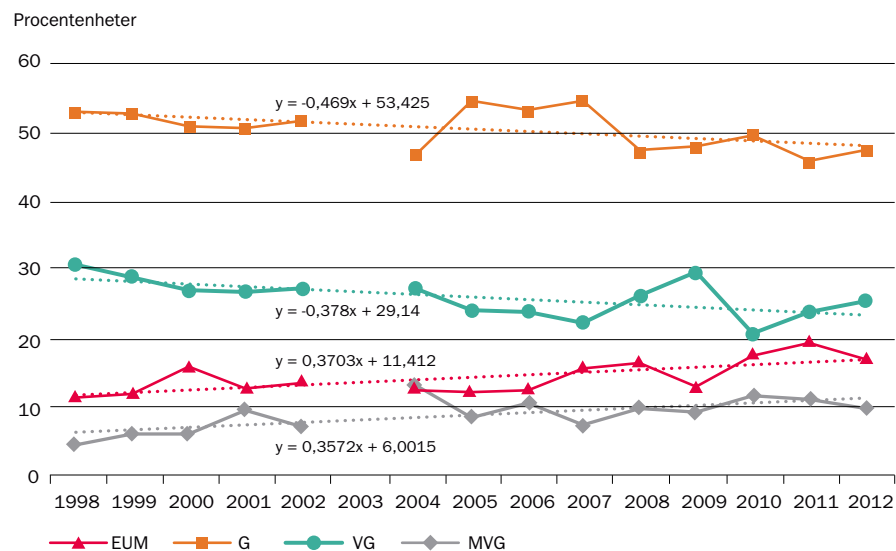


Trendlinjen visar viss tendens till minskning av variabiliteten över tid, men i huvudsak kan man konstatera att den genomsnittliga långsiktiga stabiliteten i det närmaste legat på samma nivå under den aktuella femtonårsperioden. Tabell 6 visar en genomsnittlig avvikelse på 3,9 procentenheter (cirka 1950 elever) och en standardavvikelse på 1,8 procentenheter (motsvarande 900 elever).

## Matematik

För det tredje ämnet matematik har provbetyg rapporterats under hela den aktuella perioden.<sup>39</sup> De inledande åren för ett stickprov på cirka 10 000 elever, därefter för samtliga.

**Figur 11.** Procentuell fördelning av provbetyg olika år med trendlinje och ekvation. Matematik åk 9



<sup>39</sup> Värdena för 2003 har dock uteslutits eftersom resultatavvikelserna det året till stor del beror på faktorer som inte har med provets konstruktion och de angivna kravgränserna att göra.

**Tabell 7.** Aktuell andel elever med olika provbetyg i vänstra tabellen. Modellbaserade värden i den högra. Matematik åk 9.

Rapporterade värden					
År	EUM	G	VG	MVG	GBP
1998	11,4	53,1	30,9	4,5	10,8
1999	12,0	53,0	29,0	6,0	10,9
2000	16,0	51,0	27,0	6,0	10,4
2001	12,9	50,8	26,9	9,5	11,0
2002	13,5	51,9	27,5	7,2	10,7
2003					
2004	12,6	46,8	27,4	13,2	11,4
2005	12,3	54,9	24,2	8,6	10,8
2006	12,4	53,4	23,8	10,5	11,0
2007	15,6	54,7	22,2	7,4	10,3
2008	16,5	47,5	26,2	9,9	10,7
2009	13,0	48,0	29,8	9,3	11,1
2010	17,5	49,8	20,8	11,8	10,5
2011	19,3	45,9	23,8	11,1	10,4
2012	17,0	47,7	25,4	9,8	10,5
Medel	14,4	50,6	26,1	8,9	10,8
Std	2,5	3,0	2,9	2,5	0,3

k =	0,3703	-0,3469	-0,378	0,3572	
m =	11,412	53,425	29,14	6,0015	
Löpnr	EUM	G	VG	MVG	GBP
1	11,8	53,1	28,8	6,4	10,9
2	12,2	52,7	28,4	6,7	10,9
3	12,5	52,4	28,0	7,1	10,9
4	12,9	52,0	27,6	7,4	10,8
5	13,3	51,7	27,3	7,8	10,8
6	13,6	51,3	26,9	8,1	10,8
7	14,0	51,0	26,5	8,5	10,8
8	14,4	50,6	26,1	8,9	10,8
9	14,7	50,3	25,7	9,2	10,7
10	15,1	50,0	25,4	9,6	10,7
11	15,5	49,6	25,0	9,9	10,7
12	15,9	49,3	24,6	10,3	10,7
13	16,2	48,9	24,2	10,6	10,7
14	16,6	48,6	23,8	11,0	10,6
15	17,0	48,2	23,5	11,4	10,6
Medel	14,4	50,6	26,1	8,9	10,8
Std	1,7	1,6	1,7	1,6	0,1

Även för matematik framträder en rätt varierande bild. Andelen G och VG visar tendens att sjunka medan andelen EUM och MVG visar svag ökning. Tabell 7 visar rapporterade betygsandelar och GBP samt motsvarande modellbaserade värden.

Här är standardavvikelserna för de modellbaserade värdena små, vilket indikerar att förändringen över tid varit små.<sup>40</sup> Däremot är standardavvikelserna större för de rapporterade värdena vilket speglar den rätt stora årliga variationen.

GBP-värdena visa tämligen små skillnader utom 2004, 2007 och 2009.

<sup>40</sup> Det vill säga förändringen i relation till medelvärdet för samtliga år.

**Tabell 8.** Skillnad i procentandelar av provdeltagare som i relation till modellvärden fått "för låga" (negativa värden) eller "för höga" (positiva värden) provbetyg. "Sum(ABS)" anger summan av absolutvärdena. Den vänstra tabellen anger procentenheter. Den högra anger antalet avvikelser om man antar att årskursen är 100 000 elever (med värden avrundade till 100-tal, elever 50-tal). Matematik åk 9.

År	EUM	G	VG	MVG	Sum(ABS)		År	EUM	G	VG	MVG	Sum(ABS)	Antal elever
1998	-0,38	0,02	2,14	-1,86	4,40		1998	-400	0	2 100	-1 900	4 400	2 200
1999	-0,15	0,27	0,62	-0,72	1,75		1999	-200	300	600	-700	1 800	900
2000	3,48	-1,38	-1,01	-1,07	6,94		2000	3 500	-1 400	-1 000	-1 100	7 000	3 500
2001	-0,01	-1,29	-0,74	2,05	4,09		2001	0	-1 300	-700	2 000	4 000	2 000
2002	0,24	0,16	0,23	-0,63	1,26		2002	200	200	200	-600	1 200	600
2003							2003						
2004	-1,40	-4,20	0,91	4,70	11,20		2004	-1 400	-4 200	900	4 700	11 200	5 600
2005	-2,07	4,25	-1,92	-0,26	8,50		2005	-2 100	4 300	-1 900	-300	8 600	4 300
2006	-2,34	3,10	-1,94	1,28	8,66		2006	-2 300	3 100	-1 900	1 300	8 600	4 300
2007	0,48	4,74	-3,16	-2,17	10,56		2007	500	4 700	-3 200	-2 200	10 600	5 300
2008	1,01	-2,11	1,22	-0,03	4,37		2008	1 000	-2 100	1 200	0	4 300	2 150
2009	-2,86	-1,26	5,20	-0,99	10,30		2009	-2 900	-1 300	5 200	-1 000	10 400	5 200
2010	1,27	0,88	-3,43	1,15	6,74		2010	1 300	900	-3 400	1 200	6 800	3 400
2011	2,70	-2,67	-0,05	0,10	5,52		2011	2 700	-2 700	0	100	5 500	2 750
2012	0,03	-0,52	1,93	-1,56	4,04		2012	0	-500	1 900	-1 600	4 000	2 000
Summa	0,00	0,00	0,00	0,00	88,35		Sum(ABS)	18 500	27 000	24 200	18 700	88 400	44 200
Medel	0,00	0,00	0,00	0,00	6,31		Medel(ABS)	1 200	1 800	1 600	1 200	6 300	3 150
Std	1,81	2,57	2,30	1,82	3,20		Std	1 800	2 600	2 300	1 800	3 200	1 550

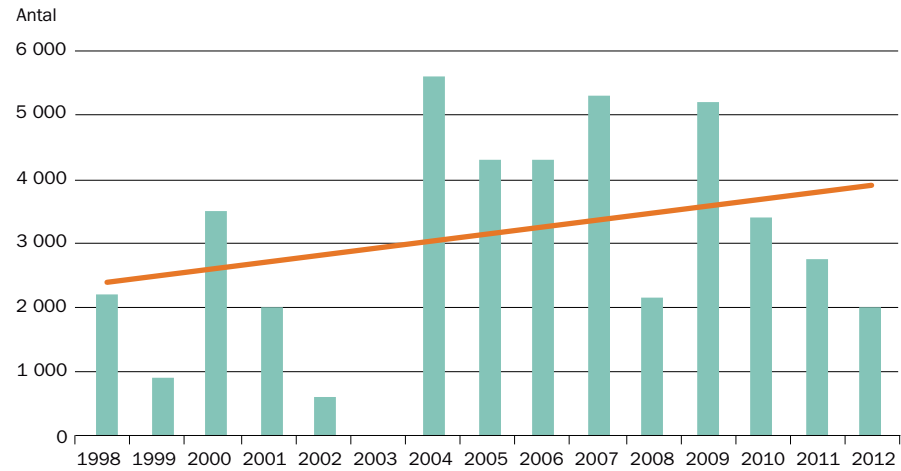
\* På grund av avrundningsfel kan Sum(ABS) i den vänstra och högra tabellen ibland ge lite olika värden.

Tabell 8 visar skillnaderna mellan de rapporterade värdena och modellvärdena.

Man kan notera att avvikelserna och skillnaderna mellan olika år är störst i matematik. Vid fyra tillfällen har avvikelserna varit större än 10 procentenheter. Genomsnittet för perioden är 6 300 avvikelser (3 150 elever med fel provbetyg) med standardavvikelsen 3 200 avvikelser (1 550 elever).

Man kan också notera de förhållandevis små avvikelserna de fem första åren då resultaten baseras på stickprov. Då var den genomsnittliga avvikelsen 3 700. För perioden 2004–2012 då totalinsamling gällde var den genomsnittliga avvikelsen mer än dubbelt så stor, 7 800. Någon saklig förklaring till denna skillnad är svår att finna utan närmare analys. Kanske har provens utformning, bedömningsanvisningarna eller något annat förhållande ändrats på något sätt. Eller kanske är det helt enkelt en slumpeffekt. I vilket fall har avvikelserna ökat, vilket illustreras i figur 12.

**Figur 12.** Antal elever (i en antagen årlig population på 100 000) som fått ett i förhållande till modellvärdet avvikande provbetyg. Matematik åk 9,



Här pekar trendlinjen mot att skillnaden mellan provbetyg och modellbetyg efterhand har ökat, dvs. att stabiliteten i provbetygen skulle ha minskat. Dock tyder utvecklingen 2009–2012 möjligen på att en ny trend med ökande stabilitet var på gång. På proven 2013 tillämpas emellertid den nya betygsskalan och eftersom det inte då finns underlag för någon modellering med trendlinje kan ingen granskning göras av det slag som redovisas här.



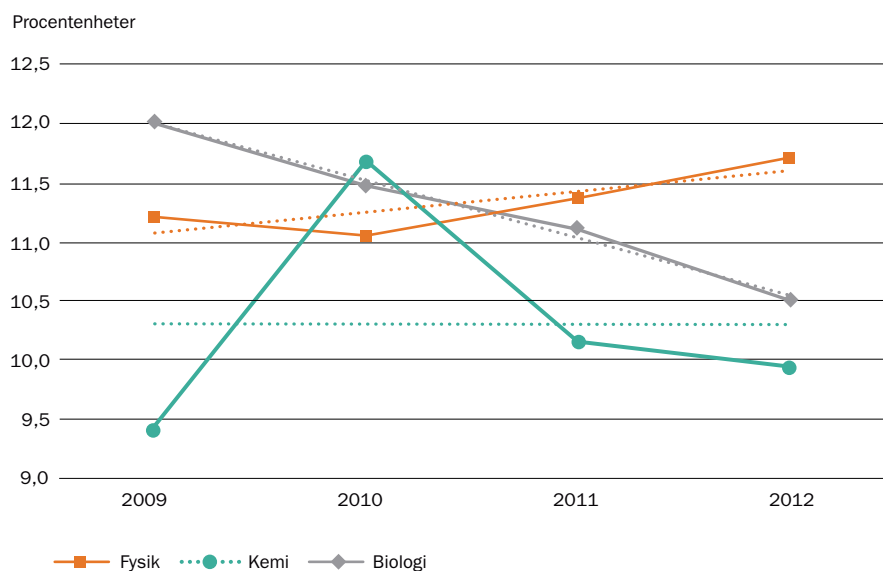
## NO-ämnena

Nationella prov i de tre NO-ämnena har genomförts sedan 2009, där 2009 var en försöksomgång med obligatoriskt deltagande för skolorna. Det finns därmed provresultat redovisade för åren 2009–2012 med provbetyg enligt MVG-skalan. Fyra år är kort för att etablera en tydlig trend och vart och ett av proven genomförs endast av en tredjedel av årskullen. Dock tyder resultaten för de redovisade proven 1998–2002, då endast ett stickprov på 10 000 provresultat insamlades att man torde kunna bortse från samlingsfelet när det gäller NO-ämnena. Trots den korta tiden kan det vara värt mödan att se hur utvecklingen ser ut de fyra år proven funnits, men resultaten får tolkas varsamt.

### Genomsnittlig betygspoäng

Figur 13 nedan visar den genomsnittliga betygspoängen för de tre ämnena som ingår i NO.

**Figur 13.** Genomsnittlig betygspoäng på nationella prov i NO-ämnena 2009–2012



Figuren visar en ganska spretig bild.<sup>41</sup> För biologi har betygspoängen sjunkit kontinuerligt över de fyra åren, för fysik har den först sjunkit något och sedan stigit, medan betygspoängen i kemi varierat kraftigt, främst genom 2010 års resultat. Liksom tidigare är det svårt att utifrån skillnader i genomsnittlig betygspoäng dra några mer konkreta slutsatser om hur stor andel av eleverna som påverkas av de årliga variationerna, så låt oss därför liksom tidigare dela upp resultaten på betygsnivå.

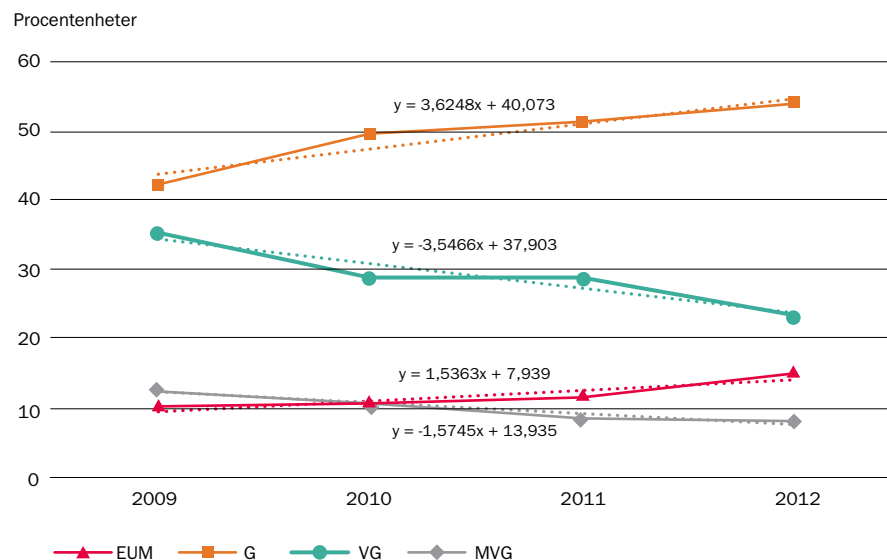
<sup>41</sup> Kan dock jämföras med de fyra första åren i figur 2, som också ger en rätt spretig bild.

## Stabilitet på betygsnivå

### Biologi

För biologi har fördelningarna av provbetyg sett ut enligt figur 14.

**Figur 14.** Procentuell fördelning av provbetyg olika år med trendlinje och ekvation. Biologi åk 9



Figuren visar att andelen elever med G och EUM ökar med 3,6 resp. 1,5 procentenheter per år (i genomsnitt) medan andelen med VG och MVG minskar med 3,5 resp. 1,6 procentenheter per år. Detta avspeglar sig också i sjunkande GBP (figur 13). Tabell 9 visar rapporterade och modellbaserade värden.

Av tabellen framgår att GBP för observerade och modellerade värden stämmer väl överens (vilket också är i linje med figur 13). Skillnaderna mellan aktuella värden och modellvärden blir små. Däremot sjunker GBP systematiskt över de fyra åren, vilket förstås kan kräva sin förklaring, men det är inte en fråga för den typ av stabilisering som är aktuell i den här studien.

**Tabell 9.** Aktuell andel elever med olika provbetyg i vänstra tabellen. Modellbaserade värden i den högra. Biologi åk 9.

Rapporterade värden						Modellbaserade värden							
År	EUM	G	VG	MVG	GBP	k =	m =	Löpnr	EUM	G	VG	MVG	GBP
2009	10,1	42,1	35,1	12,6	12,0	1,54	7,94	1	9,5	43,7	34,4	12,4	12,0
2010	10,5	49,6	29,0	10,8	11,5	3,62	40,07	2	11,0	47,3	30,8	10,8	11,5
2011	11,7	51,2	28,7	8,5	11,1	-3,55	37,9	3	12,6	50,9	27,3	9,2	11,0
2012	14,8	53,7	23,4	8,1	10,5	-1,57	13,94	4	14,1	54,6	23,7	7,7	10,5
Medel	11,8	49,1	29,0	10,0	11,3			Medel	11,8	49,1	29,0	10,0	11,3
Std	2,1	5,0	4,8	2,1	0,6			Std	2,0	4,7	4,6	2,0	0,6

**Tabell 10.** Skillnad i procentandelar av provdeltagare som i relation till modellvärden fått ”för låga” (negativa värden) eller ”för höga” (positiva värden) provbetyg. ”Sum(ABS)” anger summan av absolutvärdena. Den vänstra tabellen anger procentenheter. Den högra anger antalet avvikelser om man antar att årskursen är 100 000 elever (med värden avrundade till 100-tal, elever 50-tal). Biologi åk 9.

År	EUM	G	VG	MVG	Sum(ABS)
2009	0,62	-1,59	0,75	0,23	3,19
2010	-0,52	2,29	-1,80	0,00	4,61
2011	-0,86	0,26	1,40	-0,76	3,29
2012	0,72	-0,90	-0,31	0,47	2,39
Summa	-0,04	0,06	0,05	-0,06	13,48
Medel	-0,01	0,01	0,01	-0,02	3,37
Std	0,80	1,70	1,40	0,53	0,92

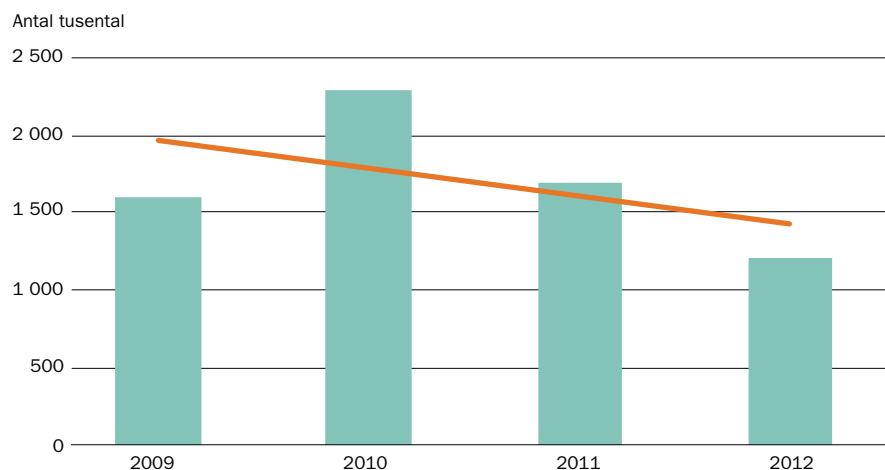
År	EUM	G	VG	MVG	Sum(ABS)	Antal elever
2009	600	-1600	800	200	3200	1600
2010	-500	2300	-1800	0	4600	2300
2011	-900	300	1400	-800	3400	1700
2012	700	-900	-300	500	2400	1200
Sum(ABS)	2700	5100	4300	1500	13600	6800
Medel(ABS)	700	1300	1100	400	3400	1700
Std	800	1700	1400	500	900	400

Den största variationen finns för ”betyget” G, vilket framgår av såväl ”Medel(ABS)” som ”Std” (standardavvikelsen).

Trots god överensstämmelse mellan genomsnittlig betygspoäng och förhållandevis liten årlig variation, blir den genomsnittliga avvikelsen mätt (på standardpopulationen 100 000 elever) cirka 3 400. Uttryckt i antalet elever får i genomsnitt årligen 1700 elever ”fel” provbetyg på grund av att betygsgränserna inte stämmer med den antagna modellen.<sup>42</sup>

Stabiliteten över de fyra aktuella åren visas i figur 15.

**Figur 15.** Antal elever (i en antagen årlig population på 100 000) som fått ett i förhållande till modellvärdet avvikande provbetyg. Biologi åk 9.



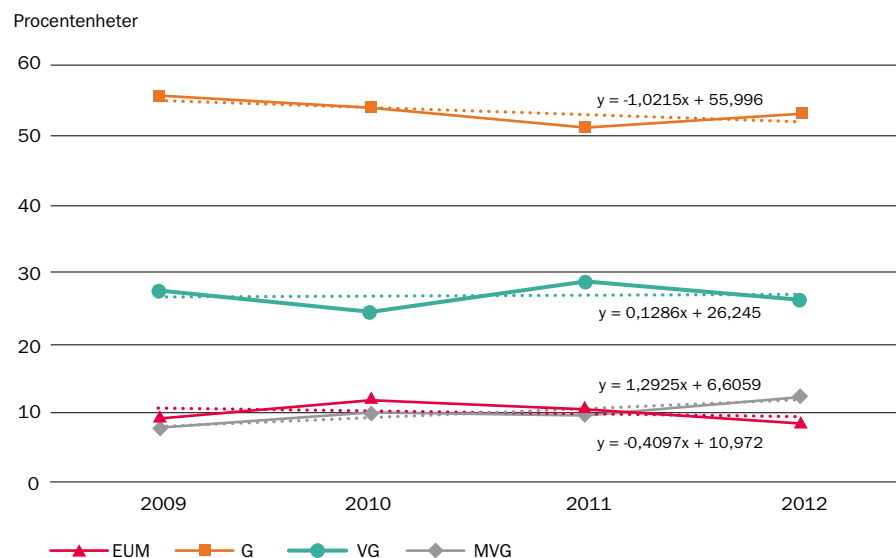
I relation till modellvärdet visar biologi en förhållandevis liten variation. Mer bekymmersamt är kanske den kraftigt sjunkande trenden i GBP (se figur 13).

<sup>42</sup> I själva verket är det endast en tredjedel av en årskull elever som gör respektive NO-prov. Därför anges andelen elever i procent av dem som gjort proven i figurerna 15–17. Någon principiell skillnad föreligger dock inte mellan procent och ”per 100 000”. Procent utgår från andel av 100 och används även om det aktuella antalet inte råkar vara 100. I princip kan man lika gärna utgå från andel av 100 000 även om det aktuella antalet inte är 100 000. Vitsen med att använda procent (eller 100 000) är förstas att underlätta jämförelser).

## Fysik

Även för fysikens del är de årliga variationerna i betygsandelar små och förändringen över tid är även den förhållandevis liten (liten lutning på trendlinjerna) som figur 16 visar.

**Figur 16.** Procentuell fördelning av provbetyg olika år med trendlinje och ekvation. Fysik åk 9



Även GBP för redovisade betyg och modellbetyg stämmer väl överens som framgår av tabell 11. Här finns inte heller någon tydlig trend för GBP utan den ligger liksom betygsandelarna på i stort sett samma nivå hela perioden.

I relation till biologi är den årliga variationen något större och i genomsnitt 4 400 elever (av en standardpopulation på 100 000) får ett provbetyg som avviker från modellens betyg (tabell 12).

**Tabell 11.** Aktuell andel elever med olika provbetyg i vänstra tabellen. Modellbaserade värden i den högra. Fysik åk 9.

Rapporterade värden					
År	EUM	G	VG	MVG	GBP
2009	9,2	55,7	27,5	7,6	11,2
2010	11,9	53,8	24,3	10,1	11,0
2011	10,3	51,1	29,0	9,6	11,4
2012	8,4	53,2	26,3	12,1	11,7
Medel	9,9	53,4	26,8	9,8	11,3
Std	1,5	1,9	2,0	1,8	0,3

k =	-0,41	-1,02	0,13	1,29	
m =	10,97	56	26,48	6,61	
Löpnr	EUM	G	VG	MVG	GBP
1	10,6	55,0	26,6	7,9	11,1
2	10,2	54,0	26,7	9,2	11,2
3	9,7	52,9	26,9	10,5	11,4
4	9,3	51,9	27,0	11,8	11,6
Medel	9,9	53,5	26,8	9,8	11,3
Std	0,5	1,3	0,2	1,7	0,2

**Tabell 12.** Skillnad i procentandelar av provdeltagare som i relation till modellvärden fått "för låga" (negativa värden) eller "för höga" (positiva värden) provbetyg. "Sum(ABS)" anger summan av absolutvärdena. Den vänstra tabellen anger procentenheter. Den högra anger antalet avvikelser om man antar att årskursen är 100 000 elever (med värden avrundade till 100-tal, elever 50-tal). Fysik åk 9.

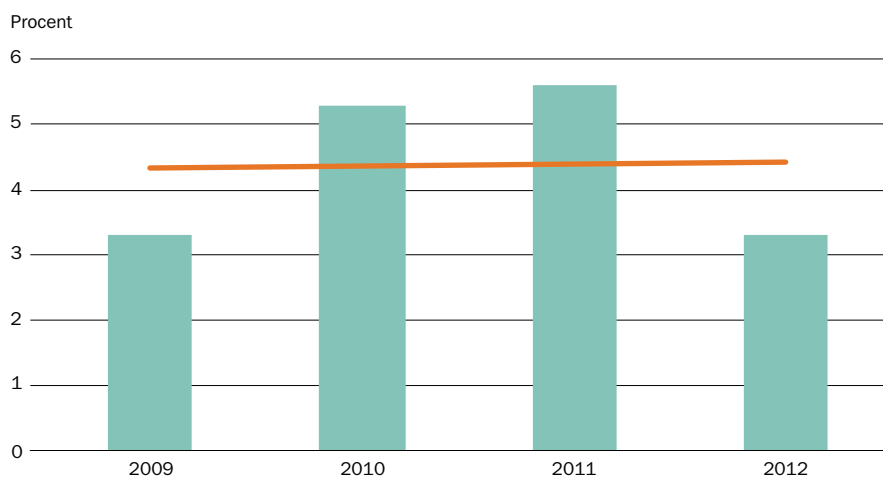
År	EUM	G	VG	MVG	Sum(ABS)
2009	-1,36	0,72	0,89	-0,30	3,27
2010	1,75	-0,16	-2,44	0,91	5,26
2011	0,60	-1,88	2,17	-0,92	5,56
2012	-0,97	1,29	-0,65	0,32	3,23
Summa	0,01	-0,03	-0,03	0,01	17,32
Medel	0,00	-0,01	-0,01	0,00	4,33
Std	1,44	1,38	1,99	0,79	1,25

År	EUM	G	VG	MVG	Sum(ABS)	Antal elever
2009	-1 400	700	900	-300	3 300	1 650
2010	1 800	-200	-2 400	900	5 300	2 650
2011	600	-1 900	2 200	-900	5 600	2 800
2012	-1 000	1 300	-700	300	3 300	1 650
Sum(ABS)	4 800	4 100	6 200	2 400	17 500	8 750
Medel(ABS)	1 200	1 000	1 600	600	4 400	2 200
Std	1 400	1 400	2 000	800	1 300	550

Man kan också notera att den största årliga variationen för fysikens del finns för betyget VG.

**Figur 17.** Antal elever (i en antagen årlig population på 100 000) som fått ett i förhållande till modellvärdet avvikande provbetyg. Fysik åk 9.

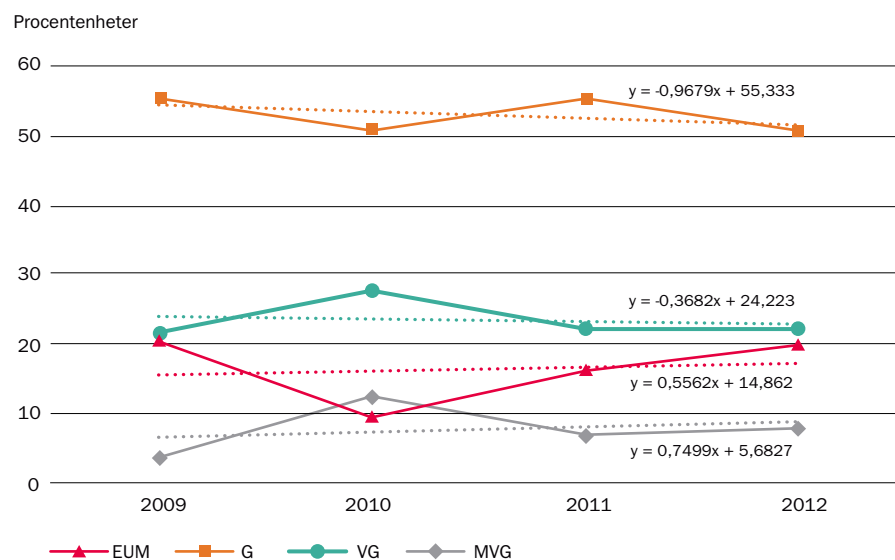


I fysik är det i genomsnitt knappt 2 200 elever som fått ett provbetyg som avviker från trendvärdet (standardavvikelse 550). Någon trend i variationen kan inte noteras för de fyra aktuella åren (fig. 17), däremot viss variation i stabiliteten. För fysikens del stiger GBP (fig. 13 och tabell 11).

## Kemi

Kemibetygen visar i enlighet med den tidigare noterade variationen i GBP den största variationen i andelen elever med olika betyg.

**Figur 18.** Procentuell fördelning av provbetyg olika år med trendlinje och ekvation. Kemi åk 9



Som framgår av tabell 13 (och tidigare figur 13) är modellens GBP konstant under perioden, medan de ingående betygen och rapporterad GBP uppvisar betydande variation.

**Tabell 13.** Aktuell andel elever med olika provbetyg i vänstra tabellen. Modellbaserade värden i den högra. Kemi åk 9.

Rapporterade värden					
År	EUM	G	VG	MVG	GBP
2009	20,0	55,2	21,4	3,5	9,4
2010	9,3	50,8	27,7	12,2	11,7
2011	16,1	55,1	22,1	6,7	10,2
2012	19,6	50,5	22,1	7,8	9,9
Medel	16,3	52,9	23,3	7,6	10,3
Std	4,9	2,6	2,9	3,6	1,0

Modellbaserade värden					
k =	0,556	-0,968	-0,368	0,75	
m =	14,86	55,33	24,22	5,68	
Löpnr	EUM	G	VG	MVG	GBP
1	15,4	54,4	23,9	6,4	10,3
2	16,0	53,4	23,5	7,2	10,3
3	16,5	52,4	23,1	7,9	10,3
4	17,1	51,5	22,7	8,7	10,3
Medel	16,3	52,9	23,3	7,6	10,3
Std	0,7	1,2	0,5	1,0	0,0

**Tabell 14.** Skillnad i procentandelar av provdeltagare som i relation till modellvärden fått "för låga" (negativa värden) eller "för höga" (positiva värden) provbetyg. "Sum(ABS)" anger summan av absolutvärdena. Den vänstra tabellen anger procentenheter. Den högra anger antalet avvikelser om man antar att årskursen är 100 000 elever (med värden avrundade till 100-tal, elever 50-tal). Kemi åk 9.

År	EUM	G	VG	MVG	Sum(ABS)
2009	4,58	0,84	-2,45	-2,93	10,80
2010	-6,67	-2,59	4,22	5,02	18,50
2011	-0,40	2,69	-1,06	-1,23	5,38
2012	2,49	-0,92	-0,69	-0,85	4,96
Summa	0,01	0,01	0,01	0,01	39,64
Medel	0,00	0,00	0,00	0,00	9,91
Std	4,90	2,27	2,91	3,47	6,32

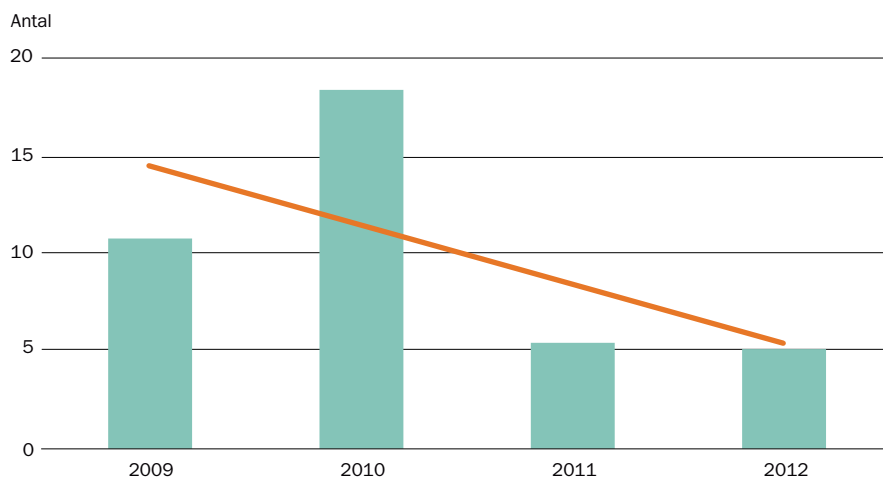
År	EUM	G	VG	MVG	Sum(ABS)	Antal elever
2009	4 600	800	-2 500	-2 900	10 800	5 400
2010	-6 700	-2 600	4 200	5 000	18 500	9 250
2011	-400	2 700	-1 100	-1 200	5 400	2 700
2012	2 500	-900	-700	-800	4 900	2 450
Sum(ABS)	14 200	7 000	8 500	9 900	39 600	19 800
Medel(ABS)	3 600	1 800	2 100	2 500	9 900	4 950
Std	4 900	2 300	2 900	3 500	6300	2750

I relation till biologi kan man notera att en betydligt större andel elever inte når målen (EUM) i kemi enligt provens betygsgräns. Även MVG visar stor variation enligt tabell 13. Framför allt är det resultaten 2010 som sticker ut. Enligt tabell 14 var då avvikelserna 18,5 procent i relation till modellvärdet. Samtidigt får man ha i åtanke att den aktuella trendlinjen endast baseras på fyra mätpunkter och därmed får ett kraftigt avvikande resultat förhållandevis stor betydelse.

Knappt 5 000 elever i genomsnitt har fått ett avvikande provbetyg om det aktuella stickprov som gjort provet skulle vara representativt för hela populationen.

Stabiliteten över tid för provbetygen i kemi visas i figur 19 nedan.

**Figur 19.** Antal elever (i en antagen årlig population på 100 000) som fått ett i förhållande till modellvärdet avvikande provbetyg. Kemi åk 9.



De två första åren avvek kraftigt från modellvärdena, men avvikelserna har minskat de två sista åren. Som redan nämnts ska inte alltför stor vikt läggas på variationen i NO-ämnenas provbetyg eftersom tidsserierna endast sträcker sig över fyra år.

## Sammanfattning, slutsatser och diskussionspunkter

Syftet med den här studien är främst att granska de nationella provbetygens stabilitet – i betydelsen frånvaro av tillfälliga betygsvängningar – under den tid nationella prov funnits inom ramen för 1994 års kursplaner för grundskolan, det vill säga för åren 1998–2012. Som referens används de trendlinjer som betygen eller den genomsnittliga betygspoängen genererar. Däremot syftar stabilitet i det här sammanhanget inte på någon extern referens från andra typer av studier, lärarbetyg eller annat. Inte heller den mer långsiktiga stabilitet, eller brist på stabilitet, som trendlinjerna i sig kan vara indikationer på tas upp till diskussion.

Genomsnittlig betygspoäng (GBP) används i många sammanhang för att redovisa sammanfattande resultat. Det kan gälla provbetyg såväl som kursbetyg och ämnesbetyg. I många sammanhang är GBP ett praktiskt mått för att göra jämförelser, men det kan också vara svårt att se vad skillnader i GBP representerar i till exempel antal eller andel elever. Därmed kan det också vara svårt att avgöra om de skillnader man noterar är försumbara eller om de bör leda till åtgärder. GBP kan således vara ett praktiskt mått för översiktlig bedömning av till exempel provbetygs stabilitet över tid, men inte särskilt användbart för grundligare analyser. Av dessa skäl ges i rapporten endast en tämligen kort redovisning av variationer i provbetygens genomsnittliga betygspoäng, men en grundligare granskning av variationer över tid i andelen elever med olika provbetyg.

En grundläggande utgångspunkt är att välkonstruerade prov inte bör uppvisa annat än tämligen marginella resultatvariationer mellan närliggande år, i synnerhet då det handlar om prov som genomförs av stora grupper och ämnen som studerats under lång tid. De prov som här är aktuella genomförs av hela årskullar, cirka 100 000, elever för de traditionella provämnena svenska, engelska och matematik och en tredjedel av årskullen för övriga prov (NO-ämnena). Några samplingsfel av betydelse torde därmed inte vara aktuella.

En felkälla som medför viss osäkerhet i beräkningarna är de relativt oprecisa data i form av andel elever med olika betyg som används. Vissa år anges andelarna med hela procent andra år med en decimal. Detta innebär att avrundningsfelen kan få relativt stor betydelse. En mer fördjupad studie bör baseras på mikrodata.

Det finns sammanfattningsvis inte skäl att tro att det i något särskilt avseende sker så snabba förändringar att det får mer än mycket marginell påverkan på elevresultaten mellan närliggande år. På längre sikt sker förstås förändringar, men i det här sammanhanget görs som nämnts inga jämförelser mot externa referenser utan provstabiliteten mäts enbart i relation till provseriens egna resultat. Studien har ingen hypotesprövande ansats och gör inte anspråk på att presentera signifikansprövade resultat utan ska ses som en explorativ granskning med syfte att ge underlag för bedömning av vilka krav på stabilitet som är rimliga att ställa på nationella prov samt för att bedöma vilka systematiska felmarginaler som kan föreligga när det gäller de angivna gränserna för olika betyg på nationella prov.

Den fråga som kvarstår blir hur stora årliga avvikelser i provbetyg som är acceptabla för att de nationella proven ska anses uppfylla rimliga krav på stabilitet.

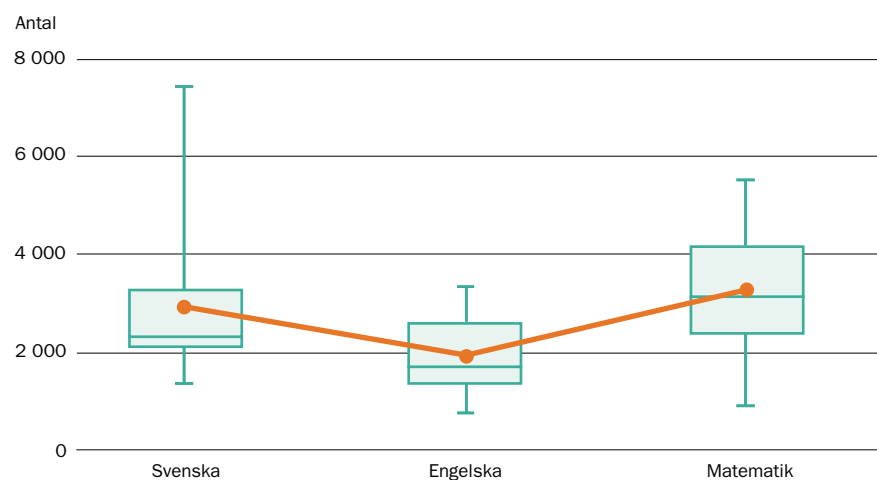


## Hur stora är de genomsnittliga avvikelserna?

Av redovisningen har framgått att provbetygen varierar över tid för alla ämnen. Om man sammanställer avvikelserna i form av boxplot-diagram för hela perioden ser den ut som i figur 20.

Den röda linjen sammanbinder *medelvärdena* medan den vågräta linjen mitt i boxen anger *medianvärdet*. Lådans botten anger den 1:a kvartilen (25:e percentilen) och lådans topp den 3:e (75:e percentilen). Själva lådan markerar således de mittersta 50 procenten av värdena. De vertikala strecken anger högsta respektive lägsta erhållna värde.<sup>43</sup>

**Figur 20.** Illustration av genomsnittligt (medelvärde och median) antal elever med avvikande betyg för hela perioden samt variationen i avvikelse olika år.



Man kan notera att medelvärdet i allmänhet ligger något högre än medianvärdet. Det beror på att extremvärden får större påverkan på medelvärdet än på medianen, vilket framgår tydligt för provet i svenska som hade ett extremt värde 2010. Medianen minskar betydelsen av extremvärden och kan ur det perspektivet ses som ett bättre mått på den genomsnittliga avvikelsen.

Av figuren framgår att provbetygen i engelska har den högsta stabiliteten med en genomsnittlig (median) avvikelse på cirka 1 800 elever och en variation (50 procent av värdena) mellan 1 300 och 2 600. Matematik har den högsta genomsnittliga avvikelsen med 3 200 elever och en variation (50 procent) mellan 2 300 och 4 200 elever.

<sup>43</sup> Egentligen anger de den 5:e respektive 95:e percentilen, men eftersom endast 12 värden ingår för svenska och engelska och 14 för matematik ligger alla värden inom intervallet 5:e till 95:e percentilen och strecken anger därför min- respektive maxvärden.

## Vilken stabilitet kan anses tillräcklig?

Något bra och vedertaget mått på stabilitet är svårt att hitta och det ligger nära till hands att snarare tala om brist på stabilitet i form av avvikelser.<sup>44</sup>

Frågan om vilken stabilitet eller avvikelse som ska anses acceptabel har inget självklart svar. Om man är nöjd med den bild som framträder i den här studien kan man kanske säga att de över perioden erhållna genomsnittsvärdena kan ses som etablerade mått på acceptabel stabilitet. En följdfråga blir då om det är rimligt att ha olika kvalitetskriterier för stabilitet mellan olika årskurser och mellan olika ämnen. Och i så fall varför?

En möjlig väg kunde vara att ange en undre gräns för stabiliteten t.ex. att en årlig avvikelse större än 4 procentenheter inte uppfyller kvalitetskriteriet, dvs. stabiliteten så som den används här ska vara minst 96 procent. Uttryckt som antal av standardpopulationen 100 000 skulle det betyda att högst 2000 elever får ha ett avvikande provbetyg. Det skulle i de aktuella fallen betyda att engelska (1950 elever enl. tabell 6) och matematik (2000 elever enl. tabell 8) skulle klara kvalitetskravet 2012, medan svenska inte skulle göra det (4000 elever enl. tabell 4). I vilket fall förefaller det rimligt att någon form av kriterium eller mål sätts upp för betygsfördelningarnas variation i relation till en mer långsiktig trend. Om de uppsatta målen inte uppfylls ett år bör orsakerna analyseras. Detta skulle kunna bidra till att stabiliteten förbättras efterhand. Hittills har utvecklingen snarast varit den motsatta. Det är ett förhållande som behöver diskuteras av de närmast berörda. Hur kan detta förklaras?

## Konsekvenser för den nya sexgradiga betygsskalan

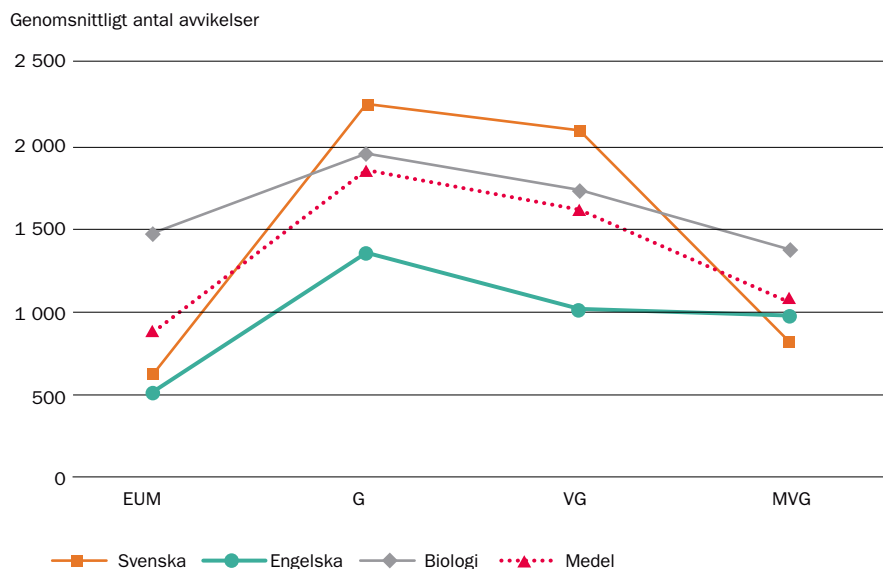
Den analys som gjorts i den här rapporten baseras på en fyrgradig betygsskala, vilket innebär att det finns tre betygsgränser. Det är också dessa betygsgränserns placering som i huvudsak avgör hur stor andel av eleverna som tilldelas olika betyg. I den nya betygsskalan som gäller från 2011 finns sex betygssteg. Det innebär att det finns *fem* betygsgränser. Eftersom det är vid betygsgränserna avvikelserna sker betyder det också att den totala avvikelsen kommer att öka för de nya provbetygen genom att två betygssteg tillkommer. Frågan är hur stor betydelse detta får för den totala absoluta avvikelsen.

Avvikelserna är som framgått varierande för olika betyg. Om man sammanställer medelavvikelser för 1998–2012 för olika betyg uppdelade på ämnen får man en bild som i figur 21.

---

<sup>44</sup> Man kunde naturligtvis tala om stabilitet som 100 procent –felprocent eller exempelvis för engelska 100 000 - 3 500 elever = 96 500 elever, eller en genomsnittlig stabilitet på 96,5 procent. Det är måhända en smaksak, men också lite av om man ska se glaset som halvfyllt eller halvtomt. Det vill säga vilket perspektiv man prioriterar. Om man som här redovisar avvikelsen blir det den som betonas.

**Figur 21.** Genomsnittligt antal avvikelser 1998–2012.



Av figur 21 framgår att betygen G och VG visar större avvikelse än EUM och MVG räknat i procentenheter eller antal. Detta förefaller inte märkligt eftersom de är de vanligaste betygen. Figur 22 visar ett exempel på hur poängfördelning och betygsfördelning kan se ut för ett prov.

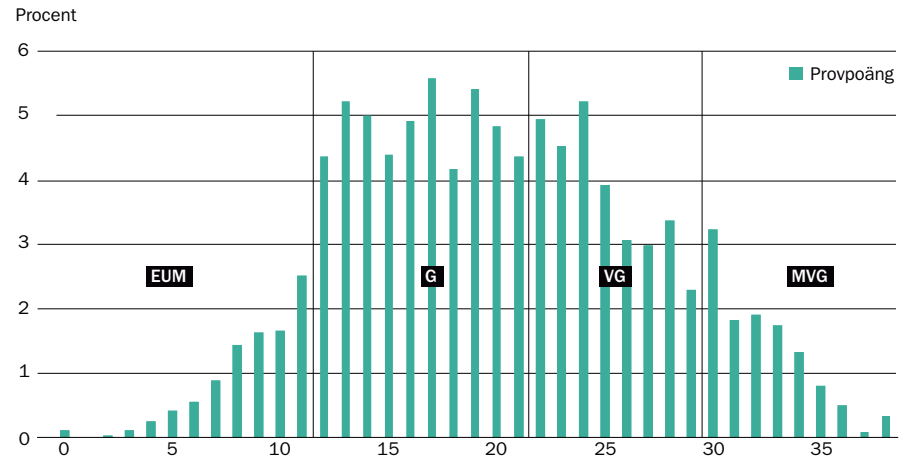
Figuren illustrerar hur en förflyttning av en poänggräns ett steg kan innebära att andelen elever som får ett visst betyg kan förändras med flera procentenheter. I synnerhet om betygsgränserna ligger på den del av skalan där det finns många elever får ett stegs förflyttning av betygsgränsen påtaglig inverkan. För betygen EUM och MVG finns endast en betygsgräns, medan G och VG har två gränser som båda inverkar på andelen elever som får dessa betyg. Vilken effekt denna dubbla påverkan får kan förstås variera på olika sätt beroende på om gränserna är sådana att de båda ökar eller minskar ett intervall eller om de samvarierar på något annat sätt.

Lite parentetiskt kan man också notera att en förflyttning av betygsgränsen ett steg i allmänhet innebär att någon betygsgrupp ökar påtagligt (och någon annan minskar). I figur 22 har cirka 9 procent av eleverna provbetyget EUM. Antag att trendlinjen anger att värdet bör vara 10,5 procent. Det aktuella värdet är alltså för lågt. En flyttning av gränsen för provbetyget G från 12 till 13 poäng för att öka andelen EUM innebär ett tillskott till gruppen EUM på cirka 4,3 procentenheter (se figur 22) till 13,3 procentenhet (9+4,3). Detta leder till ett överskott på 2,8 procentenheter (13,3-10,5), dvs. en ännu större avvikelse. De stora procentuella förändringar som ett poängsteg innebär medför således att det ofta inte är möjligt att komma särskilt nära trendvärdet. I det här exemplet är således en avvikelse neråt på 1,5 procentenheter det minst dåliga alternativet.<sup>45</sup>

<sup>45</sup> I själva verket bör man granska alla betygsgrupper samtidigt, eftersom alla betygsgrupper är kopplade till varandra genom att summan ska bli 100 procentenheter.

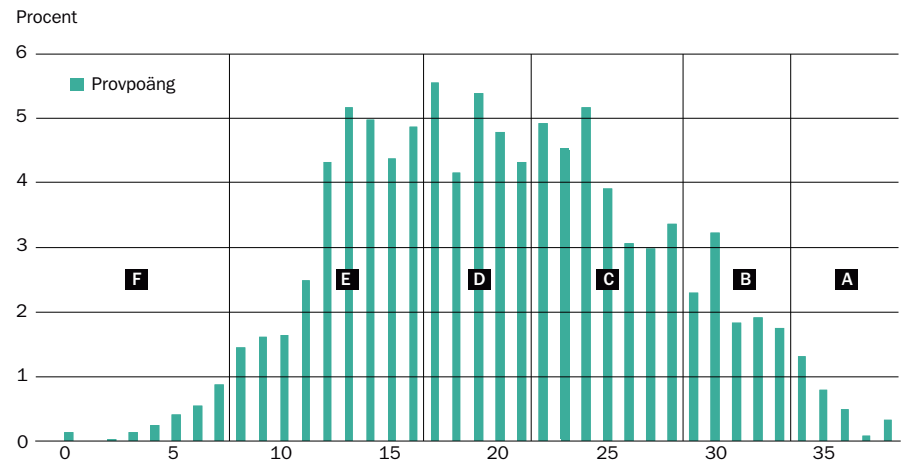
Det här är en aspekt att ha i åtanke inför beslut om kvalitetskriterier för kravgränsernas stabilitet.

**Figur 22.** Poängfördelning på ett prov med betygsgränser, betygsskala 94.



Om betygsskalan utökas med ytterligare två betygssteg finns fem betygsgränser som är mer eller mindre osäkra. Om man utgår från figur 22<sup>46</sup> och det hypotetiska fallet i figur 23 kan man se att avvikelserna för ytterbetygen antagligen minskar något i den nya skalan eftersom gränserna antas bli flyttade ut mot kanterna.<sup>47</sup> De två nya betygsgränserna däremot medför sannolikt en avvikelse i nivå med den som gällt för G och VG eftersom de ligger mer mot mitten av skalan.

**Figur 23.** Poängfördelning med betygsskala 2011 (fingerad).

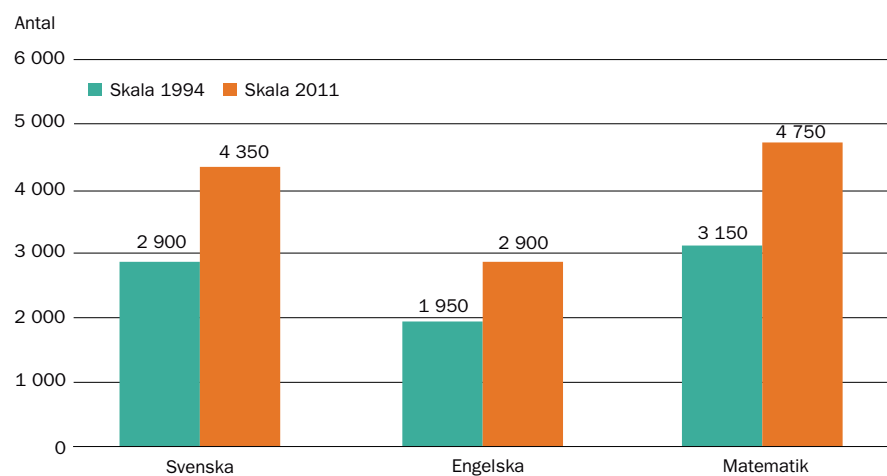


<sup>46</sup> Verkliga betygsgränser.

<sup>47</sup> Kanske kan man tänka sig att gränsen för E kommer att läggas vid samma poäng som gränsen för G om kraven för E ska anses vara lika med kraven för G. I så fall ska de två nya betygsgränserna läggas in ovanför G-gränsen. I illustrationen antas dock att hela skalan blir ny, vilket inte har någon betydelse för det principiella resonemanget. Det blir detsamma även om G-gränsen antas intakt.

Skulle man utifrån detta försöka skatta den totala avvikelsen och utgår från att de nya betygen får en absolut avvikelse i linje med genomsnittet för tidigare betyg så innebär det att avvikelserna i den nya skala gäller sex betygssteg mot fyra i den gamla. Det innebär en ökning från fyra till sex betygs-kategorier, det vill säga en ökning med 50 procent, vilket torde medföra att även avvikelsen eller felklassificeringen ökar med cirka 50 procent. Figur 24 sammanfattar utfallet uttryckt i antalet elever som får fel betyg.

**Figur 24.** Genomsnittligt antal elever per år som fått från den långsiktiga trenden 1998–2012 avvikande provbetyg enligt 1994 års betygsskala samt en skattning av hur många elever som skulle få avvikande betyg med 2011 års utökade betygsskala om stabiliteten per betyg antas ligga på samma nivå i båda fallen (medelvärden per år för 100 000 elever).



Figur 24 visar hur det genomsnittliga antalet (eller andelen per 100 000) elever som får avvikande betyg ökar med den nya betygsskalan om man antar att stabiliteten ligger på samma nivå som genomsnittet för 1998–2011. Som vi sett har den genomsnittliga stabiliteten legat på ungefär samma nivå under den aktuella perioden och det finns knappast skäl tro att den kommer att ändras nämnvärt om inte särskilda åtgärder vidtas.

Den ökade differentiering som den nya betygsskalan medger sker således till priset av minskad tillförlitlighet. Mellan 3 och 5 procent av eleverna kan antas få ett provbetyg som avviker från det betyg de borde få enligt den långsiktiga trenden för provbetyg.

## Olika feltypers betydelse

I inledningen nämndes tre olika typer av fel som påverkar provresultaten och därmed provbetygen. Grovt kan de anges enligt följande:

1. *Slumpfel på individnivå*, möjliga att skatta utifrån reliabilitetsmått baserade på provpoäng.
2. *Bedömningsfel* beroende på olika bedömares skilda tolkningar av mer komplexa svar. Normalt ej skattningsbara men kan uttryckas och beräknas som interbedömarreliabilitet i särskilda studier. Slumpmässiga på gruppnivå (beroende på vilken bedömare en klass råkar få).
3. *Stabilitetsbrister* kopplade till provens betygsanvisningar (kravgränser). Slumpmässiga(?) på systemnivå mellan olika års provversioner.

Nedan ges några korta kommentarer om respektive fel och deras betydelse.

### Slumpmässiga fel på individnivå

Alla tre feltyper verkar i allmänhet samtidigt eftersom flertalet prov innehåller olika delprov som kan vara item- och poängbaserade, essäbaserade, muntliga, laborativa etc. Bedömningen av elevprestationerna kan vara mer eller mindre tolkande och den slutliga betygssättningen baseras på olika former av anvisningar; poänggränser, exempeluppsatser etc. En väsentlig fråga är vilken verkan de olika feltyperna *sammantaget* har på det totala felet i provresultatet. Den första typen av fel kan estimeras för itembaserade prov med hjälp av provpoängens standardavvikelse och provets reliabilitet (vanligtvis angiven som koefficient alfa).<sup>48</sup>

Om vi som exempel beräknar mätfelet  $SEM^{49}$  för exemplet i figur 22 och 23 fås

$$SEM = \sigma * \sqrt{(1 - r)}$$

där standardavvikelsen  $\sigma = 7,1$  och reliabiliteten  $r = 0,88$  detta ger  $SEM = 2,46 \approx 2,5$  poäng. Detta betyder att för en provdeltagare som fått exempelvis 25 poäng på provet kan man säga att sannolikheten är 68 procent att provdeltagarens "sanna poäng" ligger i intervallet  $25 \pm 2,5$  poäng.<sup>50</sup> Vill man vara säker till 95 procent kan man säga att den sanna poängen för eleven med 25 poäng ligger i intervallet  $25 \pm 5$  poäng,<sup>51</sup> dvs. mellan 20 och 30 poäng (figur 25).

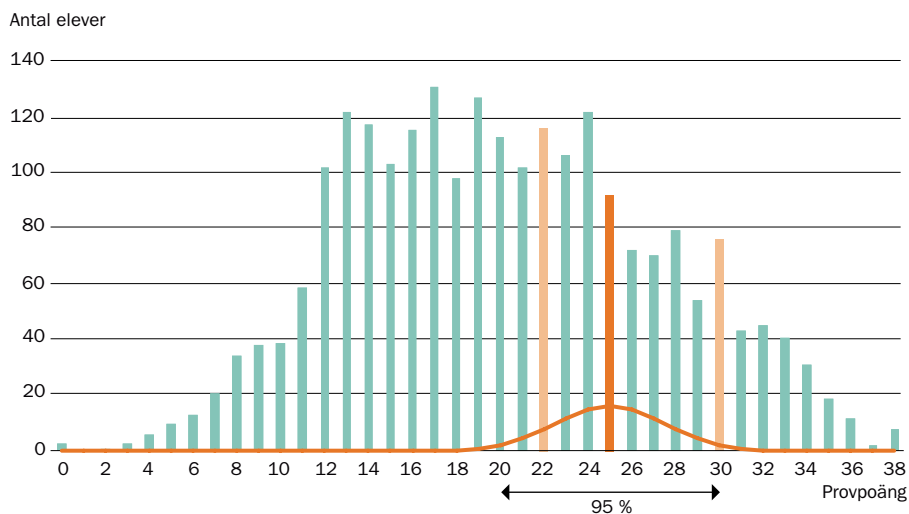
48 Ibland benämnd Cronbachs alfa, ett mått på provets "interna konsistens", vilket kan sägas ange i vilken utsträckning uppgifterna mäter samma förmåga. Värdet kan ligga mellan 0 och 1 där 0 innebär att poängen är helt slumpmässigt fördelad och 1 att den är helt konsistent.

49 Standard Error of Measurement

50 "Sann poäng" är i sammanhanget ett statistiskt begrepp, inte ett logiskt. Se t.ex. Crocker & Algina (1986) eller någon annan text om klassisk testteori.

51  $25 \pm 1,96 * 2,5 \approx 25 \pm 2 * 2,5 = 25 \pm 5$

**Figur 25** Fördelning av sann poäng för en elev med den observerade poängen 25 (den dubbelriktade pilen anger 95 % konfidensintervall)

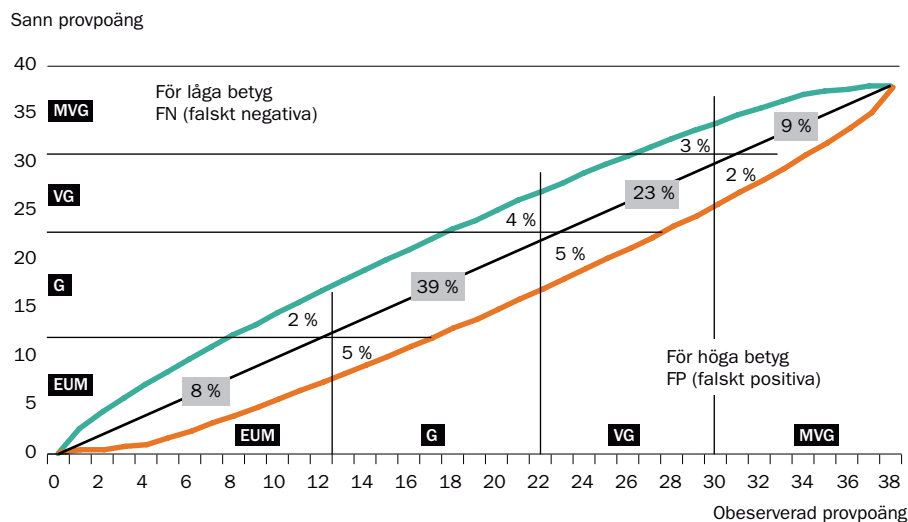


De gula staplarna i figuren anger den undre gränsen för betyget VG respektive MVG. Som framgår av figuren finns en viss sannolikhet för att den elev som har 25 poäng på provet har en sann poäng som inte når upp till VG-gränsen, liksom en viss (mindre) sannolikhet att det sanna provbetyget i stället är MVG. För en elev som träffar den undre gränsen för VG och har 22 poäng på provet ökar sannolikheten för att det sanna betyget är G påtagligt och är nästan lika stor som sannolikheten för att det sanna betyget är VG. För en sådan elev kan man endast säga att sannolikheten är 95 procent att den sanna provpoängen ligger i intervallet 17 till 27 poäng.

Om man gör en mer ingående felanalys och beräknar felmarginalen för varje provpoäng och jämför med gällande betygsgränser erhåller man nedanstående resultat (figur 26).<sup>52</sup>

52 I det här fallet används ett villkorat mått på SEM ("Conditional SEM"), vilket innebär att SEM varierar med provpoängen, samt Keats korrektion. Bilden är approximativ eftersom regressionslinjen den sanna poängen har en något annan lutning. Se Skolverket (2013) för en närmare beskrivning av de olika begreppen. Approximationerna ger dock tillräckligt god bild av felens storleksordning, vilket rä det viktiga i det här sammanhanget.

**Figur 26.** Samband mellan observerad och sann provpoäng samt 95 procentigt konfidsintervall för den sanna poängen (den blå respektive röda kurvan). De blå vertikala och horisontella strecken anger betygsgränserna. FN anger andelen elever som fått för lågt provbetyg



Figur 26 visar sambandet mellan den sanna och den observerade provpoängen samt hur stor andel av eleverna som enligt modellen fått korrekta betyg enligt sin sanna poäng (dvs. samma betyg enligt den sanna poängskalan och den observerade), för låga betyg (FN, dvs. den sanna poängen över gränsen, den observerade under gränsen för betyget) eller för höga betyg (FP, tvärtom mot FN).

Räknar man samman procentandelarna finner man att 79 procent fått korrekta betyg (rutorna efter diagonalen), 9 procent för låga betyg (FN) och 12 procent för höga betyg (FP). För en normpopulation enligt tidigare skulle således uppskattningsvis 21 000 elever få ett felaktigt provbetyg. Vilka dessa elever är kan dock inte bestämmas, varken innan provet görs eller efter att det är genomfört. Det enda man kan säga är att enligt de erhållna måtten på reliabilitet och spridning får 21 procent av eleverna fel provbetyg.

Figur 26 innehåller tre betygsgränser. Med den nya betygs-skalan 2011 tillkommer ytterligare två. Det betyder att ytterligare tvåområden med falskt positiva respektive falskt negativa provbetyg tillkommer. Själva poängfördelningen beror endast på provets reliabilitet och provpoängens fördelning. Sambandet mellan observerad poäng och sann poäng har inget samband med hur betygsgränserna sätts. Att det tillkommer två betygsgränser påverkar därmed inte det slumpmässiga felet i poäng på individnivå (SEM).<sup>53</sup>

De nya betygsgränserna innebär således att andelen elever som får för högt eller för lågt provbetyg ökar ytterligare. Om vi antar att bilden blir ungefär samma som i figur 26 för de två nya betygsgränserna tillkommer ytterligare 5–6 procent falskt negativa betyg (FN) och 7–8 procent falskt positiva (FP). Det vill säga sammantaget innefattas cirka 12–14 procent ytterligare av eleverna i gruppen som får falska provbetyg. Lägg detta till de tidigare 21 procenten med för höga eller låga provbetyg blir resultatet att *mellan 33 och 35 procent av eleverna får för högt eller lågt provbetyg*. En tredjedel av eleverna kan således antas

<sup>53</sup> Däremot kan ändrade betygsgränser ändra andelen elever som får fel betyg.



få felaktiga betyg enbart på grund av det slumpmässiga mätfelet, men det är som nämnts inte möjligt att utifrån provresultatet avgöra vilka dessa elever är. Det kan endast en lärare med god kännedom om elevernas övriga prestationer göra till större eller mindre del.

Man kunde kanske tro att det aktuella provet har särskilt låg reliabilitet, men så är inte fallet. En reliabilitet på 0,88 är klart godkänt enligt de normer som brukar gälla för den aktuella typen av prov.<sup>54</sup>

### Bedömningsfel på gruppnivå (lärarnivå)

Den andra typen av fel är inte möjlig att närmare gå in på i det här sammanhanget. Skolinspektionens kontrollrättningar av nationella prov ger dock underlag för viss bedömning av storleken på denna typ av fel<sup>55</sup>. Felen varierar stort med typen av prov. Prov baserade på item med fasta svarsalternativ eller korta svar har färre bedömningsavvikelser, medan prov med omfattande uppgifter av typen uppsatser kan ha stora avvikelser mellan olika bedömare. Vissa av de prov som ingår i den här studien, främst delprov i svenska och engelska, består av delprov med olika typer av uppgifter, såväl itembaserade som av essätyp. Här är det förstås svårt att skatta de olika typerna av fel. En möjlig uppgift för en kommande studie skulle kunna vara att mer i detalj granska och analysera de enskilda delproven och deras betyg där sådana finns. Om man har flera bedömare är det möjligt att identifiera de elevarbeten som det råder samstämmighet om och de där det råder oenighet mellan bedömare. I det här fallet är det till skillnad mot slumpfelet på individnivå möjligt om än omständligt att i efterskott identifiera vilka elevarbeten som bedömts olika.

Skolinspektionens kontrollrättningar tyder på att osäkerheten i lärares bedömningar ökar när antalet betygsssteg ökar (figur 27). Fler betygsssteg innebär således att andelen osäkra bedömningar ökar även på gruppnivå, det vill säga när det gäller de bedömningar som lärare eller andra bedömare (av grupper) gör av olika elevarbeten. Skolinspektionens resultat i figur 27 tyder på detta. För gymnasieskolans del ser man en klar ökning i omgång 4 då den nya betygsskalan tillämpas för första gången<sup>56</sup>. En skattning av figur 27 indikerar en ungefärlig ökning på 50 procent, vilket stämmer väl med den tidigare skattningen av stabilitetsfelets ökning när skalan ökas från fyra till sex steg (fig. 24). Motsvarande ökning finns inte i skolinspektionens granskning av grundskolan som använde den gamla fyrgradiga skalan i omgång 4.

Att försöka bedöma tillförlitligheten i felavvikelsens storlek utifrån Skolinspektionens kontrollrättning är svårt eftersom det saknas uppgifter om hur reliabla bedömningarna är. Låt oss mellan tummen och pekfingret uppskatta att avvikelserna i genomsnitt är hälften av vad som anges i figur 27. Om vi väljer det prov som har minsta avvikelsen och antar att omrättarna har rätt i två tredjedelar av fallen och ursprungsrättarna i en tredjedel skulle felskattningarna

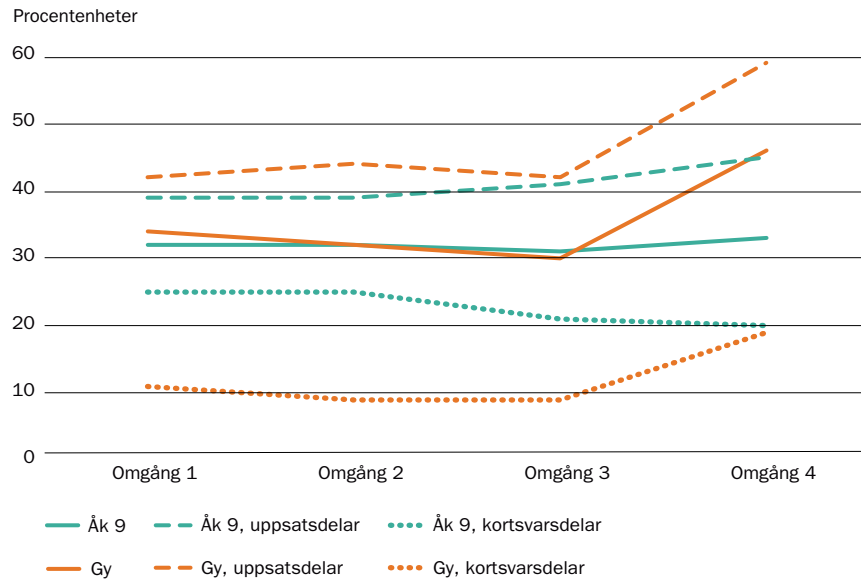
54 Wells & Wollack (2003) anger t.ex. följande: "Professionally developed high-stakes standardized tests should have internal consistency coefficients of at least .90. Lower-stakes standardized tests should have internal consistencies of at least .80 or .85. For a classroom exam, it is desirable to have a reliability coefficient of .70 or higher." De nationella proven får räknas som "lower stakes" eftersom de inte avgör betygen utan är ett stöd för lärarens bedömning.

55 Skolinspektionen (2013). Kontrollrättning...

56 Ur Skolinspektionen (2013)

landa på cirka 20 procent för den fyrgradiga skalan, vilket utifrån figur 27 knappast kan ses som en överskattning. En utökning med två betygsteg skulle då i enlighet med vad som tycks ha skett för gymnasieskolan innebära en avvikelse på cirka 30 procent.<sup>57</sup>

**Figur 27.** Ur Skolinspektionens rapport "Olikheterna är för stora", sidan 15.



### Gränssättningsfel på nationell nivå

Detta är den typ av fel som huvuddelen av den här rapporten handlar om. Även här är det möjligt att i efterskott identifiera de elever som fått "fel" provbetyg.<sup>58</sup> Dock kan detta ske först när den nationella statistiken redovisas och då är betygen sedan länge satta.

Om man jämför poängfördelningen för provet i figur 22 ser man att för vissa betyg kan en förflyttning av betygsgränsen innebära en ökning eller minskning av andelen elever i betygs kategorin med upp mot 5 procentenheter (eller 5 000 elever om vi räknar på en standardpopulation på 100 000 elever). Ur det perspektivet förefaller metoderna för att fastställa betygsgränser fungera tämligen väl i jämförelse med de svängningar som noterats för andelen elever som får respektive provbetyg olika år. Samtidigt bör man då hålla i minnet att andelen elever med visst betyg kan vara beroende av två betygsgränser som kan samvariera på olika sätt.

### Samverkan mellan olika typer av fel

Frågan om hur de olika felen interagerar med varandra får tills vidare kvarstå obesvarad. Slumpfelen (SEM och till viss del bedömningsfelen) har betydelse

<sup>57</sup> De bör observeras att de prov som i Skolinspektionens omrättningar ger största skillnaderna är de delprov som gäller längre texter av typ uppsats. Dessa utgör dock endast en mindre del av provbetyget. Övriga delprov är av itemtyp och visar mindre avvikelser vid omrättningarna, men innehåller å andra sidan slumpfel (SEM) som inte framträder vid en omrättning.

<sup>58</sup> Dock krävs förstås en hel del arbete med att granska elevernas poäng och jämföra dessa med de (eventuellt) justerade betygsgrensarna.

för vilka poäng de enskilda provdeltagarna får. De fel som diskuteras i den här rapporten gäller däremot de resulterande provbetygen oberoende av hur olika provtyper – itembaserade prov, essäprov och eventuellt andra former av prov (muntliga, laborativa etc.) – kombinerats och bedömts, och oberoende av vilken roll slumpen spelat vid bedömningen av de enskilda elevernas prov. Det rimliga är att utgå från att de olika felen ibland förstärker varandra och ibland motverkar varandra. Det man kan vara rätt säker på är dock att den typ av stabilitetsbrister på systemnivå som behandlats i den här studien påverkar alla som har provpoäng i närheten av betygsgränserna oberoende av vilken roll slumpen spelar i övrigt.

Det samlade fel som blir resultatet av de olika delfelen kan inte närmare behandlas inom ramen för den här rapporten. Enligt Suen<sup>59</sup> tillhandahåller den klassiska testteorin inga metoder för att sammanställa olika källor till felvariation till ett sammanfattande mått på felets storlek. Det krävs betydligt grundligare analyser och mer sofistikerade metoder för detta, till exempel kunde användning av ”generalizability theory” vara en möjlig ansats.<sup>60</sup>

Ställer man samman resultaten för de olika feltyperna får man följande tabell.

**Tabell 15.** Skattning av storlek på felavvikelse (andel lever med avvikande betyg) för olika typer av fel i ett nationellt prov.

Betygsskala	Feltyp		
	SEM (individnivå)	Bedömningsfel (grupp-nivå)	Stabilitetsfel (systemnivå)
Skala 1994	20 procent	20 procent	2–3 procent
Skala 2011	30 procent	30 procent	3–5 procent

Om man alltså jämför de olika feltyperna framstår stabilitetsfelet<sup>61</sup> som det klart minsta i sammanhanget. I relation till övriga fel framstår stabilitetsbristerna därmed inte som en fråga som särskilt behöver prioriteras. Dock bör det betonas att de olika felen verkar på olika aggregerade nivåer. Den *enskilda provdeltagaren* berörs av alla tre. Om resultaten endast ska anges som medelvärden på aggregerad nivå antas huvuddelen av felen på individnivå ta ut varandra.<sup>62</sup> Felen på bedömningsnivå kvarstår i den utsträckning de inte kan betraktas som slumpmässiga och Skolinspektionens kontrollrättningar indikerar att så inte är fallet. Felet på systemnivå kvarstår vid aggregering och får sin största betydelse för till exempel bedömning av kunskapsutvecklingen på nationell nivå enligt betyg på nationella prov, men också för andra jämförelser där betyg från olika år ingår.

Sammanfattningsvis kan man i fråga om övergången till en mer fingeraderad betygsskala säga att *upplösningen* i betygsskalan *har ökat* genom införandet av två nya betygssteg, men det har skett till priset av en *sämr precision*. Det vill säga *osäkerheten* om en elevs erhållna provbetyg överensstämmer med elevens sanna provbetyg har ökat när antalet betygssteg har ökat. Vilka praktiska konsekvenser detta skulle få vid till exempel meritvärdesberäkningar och betygsbase-erade urval är en fråga som behöver utredas närmare.

59 Suen (2007)

60 Se t.ex. Brennan(1992)

61 Ibland kallat kravgränsfelet.

62 Grundantagande i den klassiska testteorin. (Se t.ex. Crocker & Algina (1985))

## De använda modellernas relevans

”All models are wrong, but some are useful” skriver den engelska statistikern George E. P. Box. Det gäller förstås även de modeller som använts här. De linjära trender som använts som underlag förefaller passa väl till de data som används, men i långa loppet är de förstås inte hållbara. Om till exempel andelen elever med VG visar en stigande trend kan inte den fortsätta hur länge som helst för då skulle den till sist uppgå till 100 procent, det vill säga inga andra betyg skulle finnas, alla elever skulle få provbetyget VG. En sådan utveckling är knappast trolig.

Samtidigt innebär modellen att ju längre tidsserien blir desto mindre blir inverkan av ett nytt års värde. Frågan är om det är rimligt. Måhända vore det bättre att exempelvis utgå från de senaste tio åren och ha någon form av ”glidande” trendlinje för att bättre kunna följa den förändring över tid som torde vara ofrånkomlig – och nödvändig.

Regressionslinjer används vanligen för att ange samband mellan oberoende och beroende variabler. Något sådant samband gäller inte här utan här handlar det om tidsserier där man inte kan estimerade kommande värden utifrån den erhållna trendlinjen. Tidsserierna och den anpassade trendlinjen beskriver endast vad som varit och kan som sådana endast ge underlag för retrospektiva studier. Som underlag för en explorativ granskning av de provbetyg som getts under de 15 år som betygs-skalan från 1994 gällt kan de därmed vara användbara.

Även de modeller som används för att bedöma den enskilda provpoängens tillförlitlighet och felmarginaler (SEM) bygger på antaganden. Dessa finns redovisade i Skolverket (2013).

Slutligen några frågor för vidare utredning och bedömning.

### Frågor som behöver utredas

- Är det rimligt att betygsfördelningarna antas följa någon form av trend över tid?
- Bör en norm för acceptabel avvikelse från trend anges? Hur stor i så fall?
- Bör samma norm gälla för olika ämnen?
- Bör andelen elever med samma provbetyg vara lika för olika ämnen?
- Vilken betydelse har det för provbetygens variation om proven prövar olika mål (kunskapskrav, centrala innehåll) olika år?
- Varför ökar inte provbetygens stabilitet över tid?
- Bör man förvänta sig att stabiliteten ska öka efterhand?
- Hur påverkar fler betygssteg betygens och meritvärdernas tillförlitlighet?
- Vad bör rapporteras om provens tillförlitlighet:
  - På individnivå? (SEM)
  - På grupp-nivå? (Klass/lärare, skola, kommun, SIs kontrollrättningar)
  - På nationell nivå? (Årliga variationer, stabilitet)

## Referenser

- Brennan, R. L. (1992). *An NCME Instructional Module on Generalizability theory*. NCME, Educational Measurement: Issues and Practice. <http://ncme.org/linkserverid/6675416A-1320-5CAE-6EABBE83D800FB23/showMeta/0/>
- Crocker, L. & Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Holt, Rinehart and Winston.
- Dorans, N. J., Moses, T., & Eignor, D. (2010). *Principles and practices of test score equating* (ETS Research Report No. RR-10-29). Princeton, NJ: ETS. <https://www.ets.org/Media/Research/pdf/RR-10-29.pdf>
- Kolen, M. J., & Brennan, R. L. (2004). *Test equating, scaling, and linking: Methods and practices* (2nd ed.). New York, NY: Springer Verlag.
- Livingston, S. A. (2004). *Equating test scores (without IRT)*. Princeton, NJ: ETS. <http://www.ets.org/Media/Research/pdf/LIVINGSTON.pdf>
- Ryan, J. and Brockman, F. (2009). *A practitioners introduction to Equating*. Washington: Council of Chief State School Officers State Collaborative on Assessment and Student Standards. [http://www.ccsso.org/Documents/2008/A\\_Practitioners\\_Introduction\\_To\\_2008.pdf](http://www.ccsso.org/Documents/2008/A_Practitioners_Introduction_To_2008.pdf)
- Skolverket (2004). *Det nationella provsystemet i den målstyrda skolan*. Stockholm: Fritzes kundservice.
- Skolverket (2007). *Provbetyg – Slutbetyg – Likvärdig bedömning? Rapport 300*. Stockholm: Skolverket. [http://www.skolverket.se/omskolverket/publikation/sok?\\_xurl\\_=http%3A%2F%2Fwww5.skolverket.se%2Fwtpub%2Fws%2Fs%2Fkolbok%2Fwpubext%2Ftrycksak%2Fblob%2Fpdf1725.pdf%3Fk%3D1725](http://www.skolverket.se/omskolverket/publikation/sok?_xurl_=http%3A%2F%2Fwww5.skolverket.se%2Fwtpub%2Fws%2Fs%2Fkolbok%2Fwpubext%2Ftrycksak%2Fblob%2Fpdf1725.pdf%3Fk%3D1725)
- Skolverket (2012). *Redovisning av uppdrag om avvikelser mellan provresultat och betyg i grundskolans årskurs 9* (Dnr 75-2012:311).
- Skolverket (2015). *Om nationella prov. Provpöängens tillförlitlighet*. Dnr 2015:5
- Skolinspektionen (2013). *Olikheterna är för stora*. Stockholm: Skolinspektionen. <http://www.skolinspektionen.se/Documents/omrattning/omrattning-nationella-prov-2013.pdf>
- Suen, H, K & Lei, P-W. (2007). *Classical versus Generalizability theory of measurement*. Educational Measurement 4. <http://suen.educ.psu.edu/~hsuen/pubs/Gtheory.pdf>
- Wells, C. S. & Wollack, J. A. (2003). *An Instructor's Guide to Understanding Test Reliability* Testing & Evaluation Services, Madison, WI: University of Wisconsin. <http://testing.wisc.edu/Reliability.pdf>

## Bilaga 1

Nedan anges redovisningen av provresultat i svenska och engelska för samliga elever som fått provresultat rapporterade. I särskilda tabeller redovisas resultat för fristående skolor (finns ej med här).

	Totalt						Flickor						Pojkar					
	Antal elever	Andel (%) av elever som gjort delprovet			Ej nått målen	Antal elever	Andel (%) av elever som gjort delprovet			Ej nått målen	Antal elever	Andel (%) av elever som gjort delprovet			Ej nått målen			
		G	VG	MVG			G	VG	MVG			G	VG	MVG				
<b>Svenska Delprov</b>	<b>Antal elever</b>	<b>G</b>	<b>VG</b>	<b>MVG</b>	<b>Ej nått målen</b>	<b>Antal elever</b>	<b>G</b>	<b>VG</b>	<b>MVG</b>	<b>Ej nått målen</b>	<b>Antal elever</b>	<b>G</b>	<b>VG</b>	<b>MVG</b>	<b>Ej nått målen</b>			
<b>Sammanvägt provresultat 1)</b>	<b>91 448</b>	<b>87 668</b>	<b>37,7</b>	<b>50,1</b>	<b>9,6</b>	<b>2,6</b>	<b>45 079</b>	<b>43 467</b>	<b>25,8</b>	<b>58,4</b>	<b>14,6</b>	<b>1,2</b>	<b>46 369</b>	<b>44 201</b>	<b>49,3</b>	<b>41,9</b>	<b>4,7</b>	<b>4,0</b>
därrav elever med svensk bakgrund	80 369	77 350	36,9	50,6	10,0	2,5	39 403	38 119	24,7	59,0	15,2	1,1	40 966	39 231	48,8	42,4	4,9	3,9
utländsk bakgrund	11 079	10 318	43,2	46,3	6,8	3,7	5 676	5 348	33,4	54,1	10,1	2,4	5 403	4 970	53,8	37,9	3,4	5,0
Totalt exklusive Nynvandrade elever <sup>1</sup>	91 004	87 502	37,7	50,1	9,6	2,6	44 888	43 391	25,8	58,4	14,6	1,2	46 116	44 111	49,3	41,9	4,7	4,0
Nynvandrade elever	444	166	41,0	42,2	6,0	10,8	191	76	23,7	56,6	13,2	6,6	253	90	55,6	30,0	..	14,4
<b>Läsförståelse (delprov A)</b>	<b>91 448</b>	<b>89 166</b>	<b>26,3</b>	<b>54,2</b>	<b>12,1</b>	<b>7,4</b>	<b>45 079</b>	<b>44 123</b>	<b>20,0</b>	<b>57,9</b>	<b>17,5</b>	<b>4,7</b>	<b>46 369</b>	<b>45 043</b>	<b>32,4</b>	<b>50,6</b>	<b>6,8</b>	<b>10,2</b>
därrav elever med svensk bakgrund	80 369	78 659	25,5	54,9	12,5	7,0	39 403	38 683	19,1	58,6	18,1	4,2	40 966	39 976	31,7	51,4	7,1	9,8
utländsk bakgrund	11 079	10 507	31,7	48,9	8,9	10,5	5 676	5 440	26,0	53,0	12,7	8,3	5 403	5 067	37,8	44,5	4,9	12,8
Totalt exkl. nynvandrade elever <sup>1</sup>	91 004	88 986	26,2	54,2	12,1	7,4	44 888	44 040	20,0	57,9	17,5	4,7	46 116	44 946	32,4	50,7	6,8	10,1
Nynvandrade elever	444	180	32,8	40,6	9,4	17,2	191	83	24,1	47,0	16,9	12,0	253	97	40,2	35,1	3,1	21,6
<b>Muntlig förmåga (delprov B)</b>	<b>91 448</b>	<b>88 375</b>	<b>35,5</b>	<b>44,6</b>	<b>18,0</b>	<b>2,0</b>	<b>45 079</b>	<b>43 795</b>	<b>25,6</b>	<b>48,7</b>	<b>24,4</b>	<b>1,3</b>	<b>46 369</b>	<b>44 580</b>	<b>45,1</b>	<b>40,5</b>	<b>11,7</b>	<b>2,7</b>
därrav elever med svensk bakgrund	80 369	77 986	35,3	44,5	18,3	1,9	39 403	38 406	25,2	48,7	24,9	1,2	40 966	39 580	45,0	40,5	11,9	2,6
utländsk bakgrund	11 079	10 389	37,1	45,0	15,6	2,3	5 676	5 389	28,7	49,1	20,6	1,6	5 403	5 000	46,2	40,6	10,1	3,1
Totalt exkl. nynvandrade elever <sup>1</sup>	91 004	88 206	35,5	44,6	18,0	1,9	44 888	43 717	25,6	48,7	24,4	1,2	46 116	44 489	45,1	40,5	11,7	2,6
Nynvandrade elever	444	169	47,3	35,5	10,7	6,5	191	78	37,2	43,6	14,1	5,1	253	91	56,0	28,6	7,7	7,7
<b>Skriftlig förmåga (delprov C)</b>	<b>91 448</b>	<b>88 836</b>	<b>46,1</b>	<b>37,1</b>	<b>10,6</b>	<b>6,2</b>	<b>45 079</b>	<b>44 022</b>	<b>35,4</b>	<b>45,9</b>	<b>15,8</b>	<b>2,8</b>	<b>46 369</b>	<b>44 814</b>	<b>56,6</b>	<b>28,4</b>	<b>5,5</b>	<b>9,5</b>
därrav elever med svensk bakgrund	80 369	78 365	45,6	37,4	11,0	6,1	39 403	38 602	34,6	46,4	16,4	2,6	40 966	39 763	56,3	28,6	5,7	9,4
utländsk bakgrund	11 079	10 471	49,5	35,2	8,2	7,1	5 676	5 420	41,1	42,5	11,8	4,6	5 403	5 051	58,6	27,3	4,4	9,8
Totalt exklusive nynvandrade elever <sup>1</sup>	91 004	88 660	46,1	37,1	10,6	6,2	44 888	43 942	35,4	45,9	15,8	2,8	46 116	44 718	56,6	28,4	5,6	9,4
Nynvandrade elever	444	176	46,6	29,0	6,3	18,2	191	80	35,0	40,0	13,8	11,3	253	96	56,3	19,8	..	24,0

Elever som läser svenska som andraspråk gör samma prov som elever som läser svenska, men bedöms enligt andra anvisningar. Resultaten för elever som läser svenska som andraspråk redovisas i en särskild tabell (gäller ca 9 000 elever 2012 och redovisas inte här eftersom de inte ingår i studien. (Fr.o.m. 2013 räknas de två grupperna samman i samma tabell)).

	Totalt					Flickor					Pojkar							
	Antal elever som gjort delprovet	G	VG	MVG	Ej nått målen	Antal elever som gjort delprovet	G	VG	MVG	Ej nått målen	Antal elever som gjort delprovet	G	VG	MVG	Ej nått målen			
<b>Engelska Delprov</b>	<b>Antal elever som gjort delprovet</b>	<b>G</b>	<b>VG</b>	<b>MVG</b>	<b>Ej nått målen</b>	<b>Antal elever som gjort delprovet</b>	<b>G</b>	<b>VG</b>	<b>MVG</b>	<b>Ej nått målen</b>	<b>Antal elever som gjort delprovet</b>	<b>G</b>	<b>VG</b>	<b>MVG</b>	<b>Ej nått målen</b>			
<b>Sammanvägt provresultat<sup>†</sup></b>	<b>100 521</b>	<b>93 901</b>	<b>30,1</b>	<b>46,3</b>	<b>20,3</b>	<b>49 183</b>	<b>46 275</b>	<b>29,3</b>	<b>46,1</b>	<b>21,6</b>	<b>51 338</b>	<b>47 626</b>	<b>30,9</b>	<b>46,5</b>	<b>18,9</b>	<b>3,7</b>		
därrav elever med																		
svensk bakgrund	80 869	76 829	28,8	47,6	21,2	39 629	37 836	28,2	47,3	22,4	2,0	41 240	38 993	29,4	47,9	20,1	2,7	
utländsk bakgrund	19 652	17 072	36,0	40,4	15,8	9 554	8 439	34,3	40,5	17,9	7,2	10 098	8 633	37,6	40,3	13,8	8,3	
Totalt exklusive Njinvandrade elever <sup>†</sup>	97 764	92 427	29,9	46,7	20,4	2,9	47 985	45 615	29,1	46,4	21,8	2,6	49 779	46 812	30,7	47,0	19,1	3,2
Njinvandrade elever	2 757	1 474	42,5	20,7	9,2	27,6	1 198	660	43,5	22,7	9,8	23,9	1 559	814	41,8	19,0	8,6	30,6
<b>Muntlig förmåga (delprov A)</b>	<b>100 521</b>	<b>95 134</b>	<b>33,0</b>	<b>39,9</b>	<b>23,8</b>	<b>3,3</b>	<b>49 183</b>	<b>46 832</b>	<b>31,4</b>	<b>40,5</b>	<b>25,0</b>	<b>3,1</b>	<b>51 338</b>	<b>48 302</b>	<b>34,5</b>	<b>39,4</b>	<b>22,6</b>	<b>3,5</b>
därrav elever med																		
svensk bakgrund	80 869	77 772	32,5	41,0	24,0	2,5	39 629	38 264	31,0	41,7	24,9	2,3	41 240	39 508	34,0	40,4	23,1	2,6
utländsk bakgrund	19 652	17 362	35,1	34,8	23,0	7,0	9 554	8 568	33,1	34,8	25,4	6,7	10 098	8 794	37,1	34,9	20,7	7,4
Totalt exklusive Njinvandrade elever <sup>†</sup>	97 764	93 616	32,9	40,2	24,0	3,0	47 985	46 153	31,3	40,7	25,2	2,8	49 779	47 463	34,4	39,7	22,8	3,1
Njinvandrade elever	2 757	1 518	42,0	20,2	13,7	24,0	1 198	679	40,5	22,1	15,5	21,9	1 559	839	43,3	18,7	12,3	25,7
<b>Receptiv förmåga (delprov B)</b>	<b>100 521</b>	<b>95 618</b>	<b>29,8</b>	<b>44,9</b>	<b>20,1</b>	<b>5,2</b>	<b>49 183</b>	<b>47 043</b>	<b>30,0</b>	<b>44,2</b>	<b>20,6</b>	<b>5,1</b>	<b>51 338</b>	<b>48 575</b>	<b>29,6</b>	<b>45,5</b>	<b>19,5</b>	<b>5,3</b>
därrav elever med																		
svensk bakgrund	80 869	78 031	28,3	46,3	21,7	3,7	39 629	38 375	28,7	45,5	22,2	3,6	41 240	39 656	27,8	47,0	21,3	3,8
utländsk bakgrund	19 652	17 587	36,7	38,7	12,7	11,9	9 554	8 668	35,9	38,6	13,7	11,8	10 098	8 919	37,5	38,8	11,7	12,0
Totalt exklusive Njinvandrade elever <sup>†</sup>	97 764	94 012	29,7	45,4	20,3	4,7	47 985	46 333	29,9	44,6	20,8	4,6	49 779	47 679	29,5	46,1	19,8	4,7
Njinvandrade elever	2 757	1 606	38,4	17,4	6,7	37,5	1 198	710	40,8	18,5	6,6	34,1	1 559	896	36,5	16,6	6,7	40,2
<b>Skriftlig förmåga (delprov C)</b>	<b>100 521</b>	<b>95 242</b>	<b>38,4</b>	<b>38,4</b>	<b>18,5</b>	<b>4,7</b>	<b>49 183</b>	<b>46 915</b>	<b>35,2</b>	<b>40,7</b>	<b>20,7</b>	<b>3,4</b>	<b>51 338</b>	<b>48 327</b>	<b>41,4</b>	<b>36,1</b>	<b>16,4</b>	<b>6,0</b>
därrav elever med																		
svensk bakgrund	80 869	77 796	38,0	39,4	18,9	3,7	39 629	38 308	35,0	41,8	20,8	2,4	41 240	39 488	41,0	37,1	17,0	4,9
utländsk bakgrund	19 652	17 446	40,0	33,8	17,1	9,1	9 554	8 607	36,4	35,9	20,1	7,6	10 098	8 839	43,5	31,7	14,1	10,7
Totalt exklusive Njinvandrade elever <sup>†</sup>	97 764	93 691	38,3	38,7	18,7	4,3	47 985	46 228	35,1	41,0	20,8	3,1	49 779	47 463	41,4	36,5	16,6	5,5
Njinvandrade elever	2 757	1 551	42,4	18,6	10,4	28,7	1 198	687	43,8	19,9	13,2	23,0	1 559	864	41,2	17,5	8,1	33,2

#### Förklaringar och definitioner

Samtliga uppgifter avser elever i år 9 vårterminen 2012.

\* Antal elever

Antal elever som har avslutat årskurs 9, dvs. elever med slutbetyg från såväl det mål- och kunskapsrelaterade betygssystemet som andra bedömningssystem och elever som lämnat grundskolan utan betyg i något ämne.

\* Antal elever som gjort delprovet/provet

Antal elever som har gjort delprovet/provet och avslutat årskurs 9, dvs. elever med slutbetyg från såväl det mål- och kunskapsrelaterade betygssystemet som andra bedömningssystem och elever som lämnat grundskolan utan betyg i något ämne.

\* Elever med svensk bakgrund

I redovisningsgruppen "elever med svensk bakgrund" ingår följande:

- 1) Elever födda i Sverige med minst en förälder född i Sverige
- 2) Elever med okänd bakgrund (417 st.)

Observera att ovanstående elever kan vara berättigade till undervisning i modersmål och svenska som andraspråk. Se vidare 2 kap. 9–16 § grundskoleförordningen.

\* Elever med utländsk bakgrund

Elever födda utomlands samt elever födda i Sverige med båda föräldrarna födda utomlands.

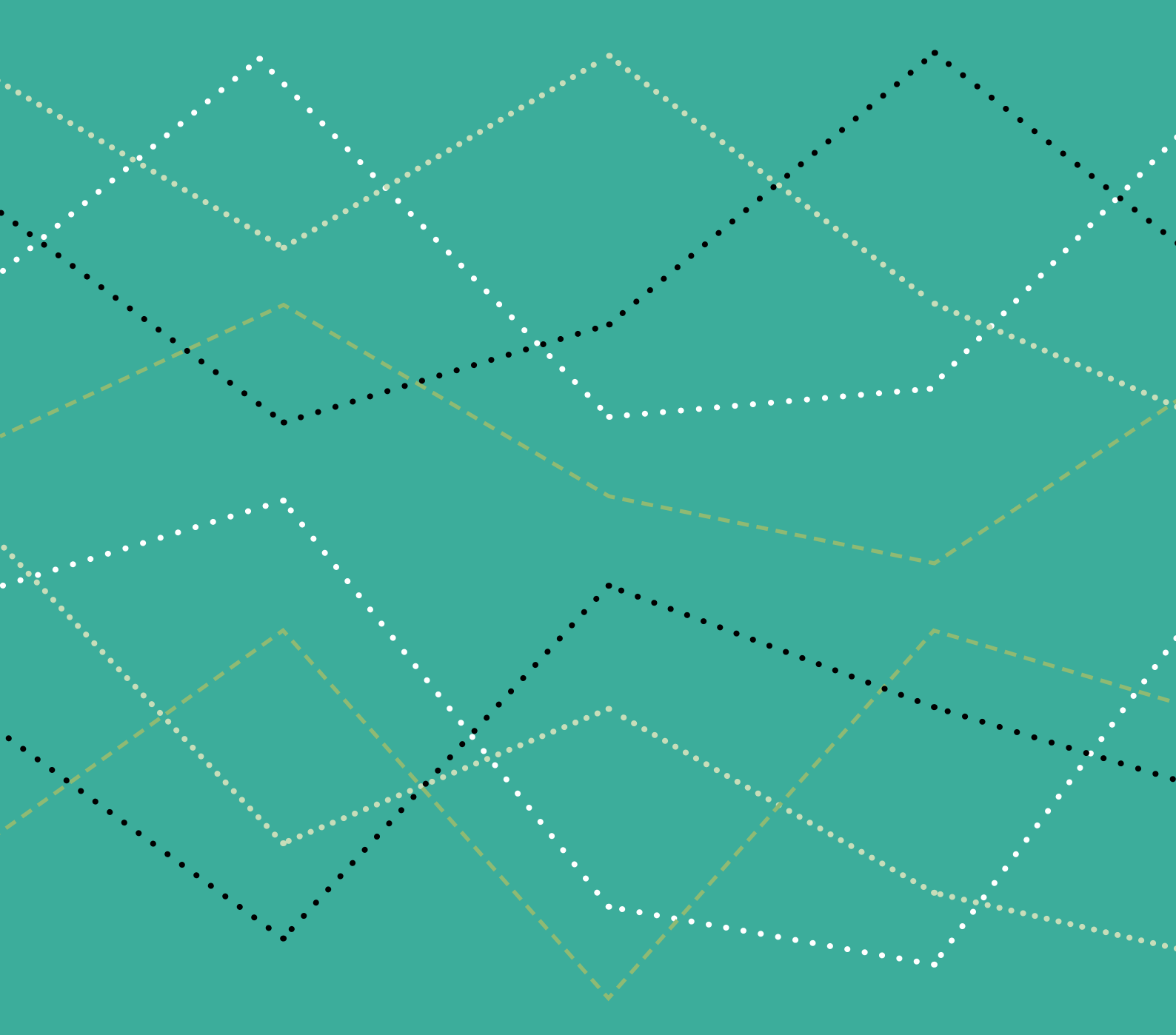
- 1) I svenska/svenska som andraspråk och engelska bedöms och betygsätts varje delprov separat. Resultaten på de olika delarna vägs sedan samman till ett sammanvägt provresultat. För att få ett sammanvägt provresultat ska eleven ha deltagit i samtliga delprov.
- 2) I matematik ges inga delprovsbetyg. För att få ett provbetyg i matematik ska eleven ha deltagit i samtliga delprov.
- 3) Fristående skolor inkluderar internationella skolor samt riksinternatskolor.
- 4) Nyinvandrade elever har kommit till Sverige de senaste 4 åren. De har inte bott i Sverige eller gått i svenska skola tidigare.

Ytterligare information om uppgifterna finns i kvalitetsdeklarationen.

Närmast föregående publicering: Grundskolan – Betyg och prov – Riksnivå – Läsår 2008/09 – Tabell 8 A-B Skolverkets rapport nr 359.







*Skolverket*

[www.skolverket.se](http://www.skolverket.se)