

Designförändringar mellan PISA 2012 och PISA 2015 – en metodstudie

Publikationen finns att ladda ner som
kostnadsfri pdf från Skolverkets webbplats:
www.skolverket.se/publikationer

ISBN: 978-91-7559-326-5

Skolverket, Stockholm 2018

Inledning.....	2
Sveriges resultat i PISA.....	3
Förändringar i PISA:s design och genomförande	3
Förändringar av mätmodellen.....	3
PISA-provens utformning och övergång till datorbaserade prov	4
Studiens resultat.....	6
Förändringar av mätmodellen påverkade inte resultaten	6
Förbättrad uthållighet förklarar inte uppgången	6
Förbättrad provmotivation kan ha påverkat resultaten.....	7
Förändringar i elevernas demografiska sammansättning påverkade inte resultaten	8
Skolverkets slutsatser	8

Inledning

Allt sedan den första PISA-undersökningen 2000 har utvecklingen av de svenska elevernas kunskapsresultat visat på en nedåtgående trend. Men 2015 bröts den negativa resultattrenden: de svenska elevernas resultat visade för första gången en förbättring jämfört med föregående undersökning. Frågan är om detta trendbrott speglar ett trendbrott i utvecklingen av elevernas kunskaper.

I samband med PISA 2015 genomfördes några viktiga förändringar i PISA-undersökningens design och genomförande. Bland annat förändrades PISA-provens utformning och den statistiska mätmodell som används för att beräkna elevernas kunskapsresultat. Dessutom genomfördes proven på dator för första gången. Förändringarna kan ha påverkat elevernas resultat och därmed också jämförelser med tidigare års undersökningar.

Skolverket har med anledning av detta gett Christian Monseur i uppdrag att undersöka i vilken utsträckning de resultatförbättringar som svenska elever uppvisat 2015 jämfört med 2012 kan förklaras av dessa förändringar i PISA:s design och genomförande. Christian Monseur är forskare vid Liege universitet och medlem i PISA:s *technical advisory group*.

Vi vet också att elevers motivation att göra sitt bästa på prov påverkar resultaten. Christian Monseur fick därför också i uppdrag att studera hur elevernas motivation att delta i PISA kan ha påverkat deras prestationer. Den här sammanfattningen återger de viktigaste resultaten från Christian Monseurs analysuppdrag. För en mer detaljerad beskrivning av resultaten se Monseur 2017.

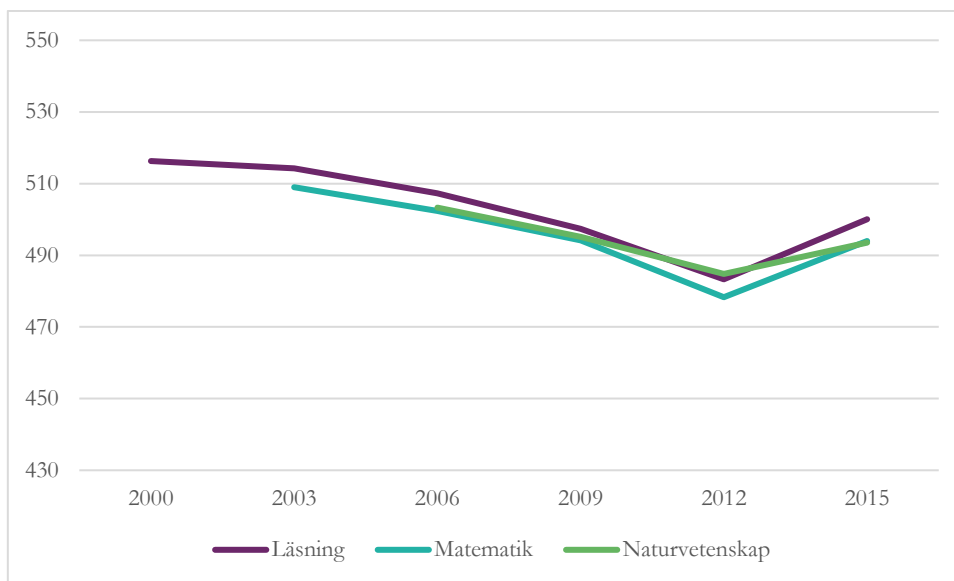
Skolverket har även anlitat den svenska forskaren Hanna Eklöf vid Umeå universitet att analysera elevernas provmotivation och bedöma hur eventuella förändringar kan ha påverkat resultaten i PISA. Detta diskuterades i Skolverkets rapport om PISA 2015 och kommer därför inte behandlas i den här sammanfattningen.

Sveriges resultat i PISA

Allt sedan PISA:s första mätning år 2000 har Sveriges resultat visat på en negativ kunskapsutveckling över tid. Resultaten från PISA 2015 visade på en förbättring jämfört med PISA 2012, och det såg ut som att den negativa kunskapstrenden nu hade brutits.

De svenska 15-åringarnas resultat hade förbättrats i alla tre kunskapsområden som PISA mäter. I läsförståelse förbättrades resultaten med 17 poäng, i matematik med 16 poäng och i naturvetenskap med 9 poäng. Förbättringen i naturvetenskap var däremot inte statistiskt signifikant.

Figur 1. Sveriges resultat i PISA mellan 2000 och 2015.



Förändringar i PISA:s design och genomförande

OECD arbetar kontinuerligt med att förbättra PISA-undersökningen och inför 2015 års undersökning genomfördes några förändringar vad det gäller mätmodellen och designen av PISA-proven. Förändringarna kan ha påverkat resultaten i 2015 års undersökning och därmed också jämförelser med tidigare års undersökningar.

Förändringar av mätmodellen

Mätmodellen har förändrats på två sätt. Det ena är att man övergått från en 1-parametermodell (även kallad Raschmodell) till en 2-parametermodell. Det andra är att man inom ramen för denna modell valt att hantera de provuppgifter som eleverna inte hunnit nå annorlunda jämfört med tidigare.

Övergång till 2-parametermodell

I likhet med de flesta andra storskaliga internationella studier, som till exempel TIMSS och PIRLS, används en statistisk modell för att kunna skatta elevernas olika kunskapsnivåer på en gemensam skala. Den statistiska modellen utgår från det man brukar kalla den moderna testteorin och det som kallas *Item Response Theory* (IRT). Från 2000 fram till och med 2012 användes den ”klassiska” Raschmodellen, som är en IRT-modell med bara en parameter, för att bestämma provens förmågeskalor. Den här enda parametern återger hur svår eller lätt en provuppgift är utifrån den kunskapsförmåga som är av intresse.

Raschmodellen har fått kritik för att den inte i tillräckligt stor utsträckning förmått förmådde hantera komplexiteten i datan från PISA-provet. Inför 2015 års undersökning valde man därför att överge Raschmodellen till förmån för en 2-parametermodell. Den andra parametern, utöver den första som återger provuppgiftens svårighetsgrad, utgörs av så kallad differentiering. Differentieringen fångar hur mycket eller lite information en provuppgift ger för att kunna särskilja olika förmågor inom ett visst kunskapsintervall.

Ny hantering av provuppgifter som eleven inte hunnit nå

När en elev inte hunnit besvara alla provhäftets provuppgifter kallas de provuppgifter som eleven inte hunnit fram till, och alltså inte gett något svar på, för *not reached*. Mellan åren 2000 och 2012 betraktades dessa *not reached* som om eleven hade gett ett felaktigt svar. Detta ändrades i och med 2015 års undersökning till att i stället betraktas som om eleven inte hade tagit del av provuppgifterna. I PISA 2015 kodas alltså inte längre *not reached* som felaktigt svar.

PISA-provens utformning och övergång till datorbaserade prov

Utöver de redan nämnda förändringarna genomfördes även några förändringar av för eleven mer konkret art. Man förändrade PISA-provens utformning samtidigt som man gick över från pappersbaserade till datorbaserade prov.

Förändringar i provens utformning

I varje PISA-undersökning är det ett av tre kunskapsområden (läsförståelse, matematik och naturvetenskap) som är huvudområde (*major domain*). Övriga kunskapsområden kallas biområden (*minor domains*), eftersom de inte undersöks lika grundligt som huvudområdet.

Fram till och med 2012 års undersökning har detta framför allt inneburit att elevernas prov innehållit betydligt färre provuppgifter från de kunskapsområden som varit biområden jämfört med det som varit huvudområde. Konsekvensen har varit att trendskattningarna för de två kunskapsområden som varit biområden har varit betydligt mer osäkra än för det kunskapsområde som varit huvudområde.

För att öka säkerheten i biområdets trendskattningar, infördes fler provuppgifter i de två kunskapsområden som utgör biområden i PISA 2015 jämfört med tidigare år. Allt för att ge en större innehållslig täckning som ökar mätsäkerheten och i

förlängningen validiteten. Samtidigt som man genomförde denna förändring ville man inte öka belastningen för eleverna som skulle skriva proven. Omfattningen på de prov som används i PISA har därför hållits konstant mellan undersökningarna.

Antalet provuppgifter i PISA är alldeles för många för att det ska vara möjligt för eleverna att besvara alla. Därför använder PISA en så kallad roterad design, vilket även andra storskaliga internationella studier gör. Det innebär att man har tagit fram flera olika prov som vart och ett utgörs av fyra delar, eller kluster som man kallar det i PISA. Varje del består av provuppgifter från bara ett av studiens kunskapsområden. Proven är sedan sammansatta så att de utgörs av kluster från mer än ett av studiens kunskapsområden samtidigt som dessa kluster förekommer i mer än ett av studiens prov. Konsekvensen av detta är att varje enskild elev bara besvarar en mindre andel av alla provuppgifter och att eleverna på en viss skola i stor utsträckning gör olika prov. Proven är sammansatta av provuppgifter som i viss utsträckning överlappar varandra mellan de olika proven.

När man utökade antalet provuppgifter från biområdet, ändrade man också så att andelen elever som besvarade provuppgifter från mer än två kunskapsområden minskade jämfört med 2012. Det vill säga andelen elever som besvarade prov med kluster från mer än två kunskapsområden var betydligt mindre 2015 jämfört med 2012. I 2015 års undersökningen var det bara 12 procent av eleverna som besvarade provuppgifter från mer än två kunskapsområden. Motsvarande andel var cirka 50 procent 2012.

Från pappersbaserade till datorbaserade prov

Inför 2015 års undersökning övergav man de traditionella pappersproven till förmån för datorbaserade motsvarigheter. Datorbaserade prov ger nya och utökade möjligheter att mäta elevernas kunskaper och förmågor.

En provuppgift som genomförs på dator behöver inte nödvändigtvis mäta samma förmågor och kunskaper som när samma provuppgift genomförs med papper och penna. Om så är fallet har övergången till datorbaserat prov gett upphov till så kallade *mode effects*, något som i så fall skulle kunna ha påverkat Sveriges resultat i endera riktningen.

OECD har försökt kontrollera för detta genom att i PISA 2015:s förstudie låta elever göra samma provuppgifter på både papper och dator. De provuppgifter som visade markanta skillnader mellan papper- och datorgenomförande togs sedan inte med i undersökningens huvudstudie.

Men även om man har säkrat att provuppgifterna mäter samma förmågor och kunskaper oavsett om de genomförs på papper eller dator, kan *mode effects* uppstå om elevernas uthållighet och motivation påverkas av övergången från pappersprov till datorbaserade prov.

Studiens resultat

Eftersom resultaten för Sverige i PISA 2015 visar på ett trendbrott är det särskilt viktigt att utreda och klargöra om de förändringar som har beskrivits ovan kan ha påverkat Sveriges resultat. Christian Monseurs uppdrag var att studera och estimerade de eventuella effekter som dessa metodologiska förändringar kan ha gett upphov till. Nästa avsnitt återger Christian Monseurs resultat.

Förändringar av mätmodellen påverkade inte resultaten

I det första av Christian Monseurs arbeten har han fokuserat de två grundläggande förändringarna i PISA:s provdesign vad det gäller övergången till en 2-parametermodell samt behandlingen av provuppgifter som elever inte hunnit med (*not reached*). För ändamålet har sex olika modeller estimerats för att i möjligaste mån utröna om förändringarna kan ha påverkat de svenska elevernas resultat eller inte. De sex modellerna utgör olika kombinationer av Rasch- respektive 2-parametermodellen samt hanterandet av *not reached* som felsvar respektive eleven har inte erhållit provuppgiften.

I analyserna har Monseur fokuserat på matematik som var huvudämne 2012 och naturvetenskap som var huvudämne 2015 och han konstaterar att den ökning i resultat vi kunnat se för Sveriges del inte kan förklaras av dessa två förändringar. Visserligen ser det ut som att Sveriges resultat har förbättrats något av att *not reached* inte längre betraktas som felaktiga provsvar. Men då övergången till 2-parametermodellen ser ut att ha missgynnat de svenska resultaten lite grand blir den sammantagna bilden att dessa metodologiska förändringar har haft en försumbar inverkan på de svenska elevernas resultatutveckling mellan 2012 och 2015.

Förbättrad uthållighet förklarar inte uppgången

För att se hur förändringarna av proven har påverkat elevernas uthållighet har Christian Monseur studerat hur eleverna presterat i de tidigare delarna av proven och jämfört dem med hur de har presterat i de senare delarna av proven. Mer konkret har han jämfört lösningsfrekvenser för provuppgifter när de förekommer i inledningen av proven med när samma provuppgifter kommer närmare slutet av proven. Detta har varit möjligt i och med den så kallade roterade designen som medför att en given provuppgift förekommer på olika platser i de olika proven. Så beroende på vilket prov en elev har blivit slumpmässigt tilldelad kan en provuppgift förekomma tidigt samtidigt som samma provuppgift förekommer sent i en annan elevs prov.

I analysen har de provuppgifter som ingått i både 2012 och 2015 års undersökningar använts. Resultaten visar att de svenska eleverna generellt sett har en lägre uthållighet än övriga OECD-länder. Det vill säga, det är större skillnad i andel korrekt lösta provuppgifter i början jämfört med slutet av proven jämfört med de andra OECD-länderna. Man kan också se att uthålligheten har förbättrats från 2012 till 2015 både för Sverige och för OECD-länderna i stort.

Analysen visar också att sannolikheten för att besvara en provuppgift korrekt samvarierar med hur många kunskapsområden man som elev utsatts för vid provtillfället. De elever som besvarat ett provhäfte med provuppgifter från bara två kunskapsområden har i något större utsträckning besvarat provuppgifterna korrekt jämfört med de elever som besvarat provhäften med provuppgifter från tre kunskapsområden.

Man kan tänka sig att både övergången till datorbaserade prov och de förändringar som genomförts i provens design till viss del har påverkat elevernas uthållighet i positiv riktning. Men uppgången i läsförståelse och matematik, det vill säga de kunskapsområden där de svenska eleverna presterat signifikant bättre jämfört med 2012, kan inte ha berott på en ökad uthållighet bland svenska elever menar Christian Monseur. Skälet är att de svenska elevernas förbättringar i uthållighet varit lägre i läsförståelse och matematik jämfört med övriga OECD-länder, samtidigt som de flesta OECD-länders resultat försämrats i läsförståelse och matematik.

Förbättrad provmotivation kan ha påverkat resultaten

Christian Monseur bedömer att de svenska eleverna har varit mer motiverade 2015. Han konstaterar att om eleverna hade varit lika motiverade 2015 som 2012 hade de presterat sämre än vad de nu gjorde.

Sverige, Norge och Danmark ställde 2012 och 2015 sex frågor till eleverna om deras provmotivation i anslutning till PISA-proven. De sex frågorna har använts för att konstruera en provmotivationsskala som sedan använts vid analysen. Resultatet visar att motivationen ser ut att ha ökat för alla de tre länderna. Störst ökning ser vi för Sverige. Från att i genomsnitt ha legat på en lägre motivationsnivå 2012 visar de svenska eleverna 2015 en högre genomsnittlig motivation än de norska och danska eleverna.

De elever som har valt att besvara provmotivationsfrågorna har i genomsnitt presterat betydligt högre än de som inte svarat på frågorna. Den här skillnaden varierar dels i storlek mellan Sverige, Norge och Danmark, dels hur den har förändrats mellan 2012 och 2015. I Norge har differensen minskat något, medan den i Danmark i stället har ökat något. I Sverige har differensen minskat drastiskt. Christian Monseur menar att det finns två tecken på att de svenska eleverna kan ha varit mer motiverade 2015: dels bortfallet på dessa frågor, dels att skillnaden har minskat i prestation mellan de som besvarat jämfört med de som inte besvarat.

Vidare visar analysen att om eleverna gett uttryck för provmotivation eller inte samvarierar med elevernas prestationer i PISA. Elever med högre provmotivation har presterat bättre resultat. Eftersom de svenska eleverna uppvisade en högre provmotivation 2015 jämfört med 2012, konstaterar Monseur att om eleverna hade varit lika motiverade 2015 som 2012 hade de presterat sämre än vad de nu gjorde.

Förändringar i elevernas demografiska sammansättning påverkade inte resultaten

De förbättrade resultaten kan också ha påverkats av oväntade förändringar i de deltagande elevernas demografiska sammansättning. Särskilt fokus har därför lagts på olika socio-kulturella faktorer som har en dokumenterad stor påverkan på elevers skolprestationer.

För ändamålet har Christian Monseur använt sig av variabler som föräldrarnas utbildningsnivå, migrationsbakgrund, antalet böcker i hemmet och ekonomiskt välstånd för att nämna några. Han har dels studerat hur det svenska elevurvalet för respektive PISA-undersökning sett ut i relation till dessa variabler, dels om förändringarna i sammansättning har haft någon inverkan på de svenska resultaten 2015 jämfört med 2012. Monseur konstaterar att elevernas demografiska sammansättning inte har påverkat trendestimatet, det vill säga hur 2015 års resultat förhåller sig till 2012 års resultat.

Skolverkets slutsatser

De förändringar som OECD genomfört av PISA-studiens mätmodell och provdesign har för de svenska elevernas resultat i stort varit försumbara. Det är även osannolikt att de svenska elevernas demografiska sammansättning 2015 jämfört med 2012 har påverkat resultaten.

De svenska elevernas uthållighet och provmotivation tycks ha ökat mellan 2012 och 2015. Övergången till datorbaserade prov kan ha påverkat detta. Den ökade uthålligheten förklarar däremot inte den svenska resultatuppgången medan den ökade provmotivationen förmodligen har varit en bidragande faktor till de förbättrade resultaten. Det är dock svårt att precisera hur stort detta bidrag har varit eftersom det i analyserna inte varit möjligt att studera orsak och verkan bakom elevernas prestationer och provmotivation. Det finns goda skäl till att anta att provmotivation påverkar resultaten, men det finns också goda skäl till att anta att provmotivation påverkas av kunskapsnivå. Så om de svenska elevernas kunskapsnivå har höjts från 2012 till 2015 är det också rimligt att de svenska eleverna uppvisar en högre provmotivation 2015.