



EN FÖRDJUPNING

Ålderseffekt och årskurseffekt

En sammanfattande bild och
grundläggande metodgenomgång

Skolverket

Publikationen kan laddas ner kostnadsfritt från
Skolverkets webbplats www.skolverket.se/publikationer

ISBN: 978-91-86529-21-5

Grafisk produktion: AB Typoform
Stockholm 2010

Ålderseffekt och årskurseffekt

En sammanfattande bild och
grundläggande metodgenomgång

Innehåll

Förord	5
Inledning	6
Vad säger forskningen?	8
Metod	9
Resultat	10
PIRLS 2001, Sverige åk 3 och 4	10
PIRLS 2006 Island och Norge, årskurs 4 och 5	16
TIMSS 1995, Sverige, årskurs 6, 7 och 8	18
TIMSS Advanced 2008, Sverige, årskurs 11–12	19
ICCS 2009, Sverige, åk 8–9	20
Problem med modellen och utvidgningar av modellen	21
Skalning	22
Multikollinjäritet	22
Standardiserade koefficienter	22
Flernivåansats	22
Avslutande diskussion	24
Referenser	26
Appendix 1	27
Appendix 2	29

Förord

Målgruppen i internationella studier kan definieras på olika sätt. Ett sätt är att utgå från årskurs, som IEA-studierna (International Association for Evaluation of Educational Achievement) gör och fokusera på till exempel elever i årskurs 4 eller 8. Ett annat sätt är att utgå ifrån ålder som i OECDs PISA-studie (Programme for International Student Assessment), där 15-åringar utgör målgruppen. I Sverige sker skolstarten vid 7 års ålder, medan eleverna i många andra länder börjar vid 6 års ålder eller t o m tidigare. Det betyder att Sverige i vissa internationella studier deltar med elever som är något äldre än genomsnittet för andra deltagande länder (IEA-studierna) men har gått lika många år i skolan. I andra studier (PISA) deltar Sverige med elever som är lika gamla, 15 år, men som har ett skolår mindre än elever i de flesta andra länder.

Detta får konsekvenser för resultatens jämförbarhet och det är därför viktigt att försöka få en uppfattning om hur stor effekten av årskurs respektive ålder är, vilket är syftet med föreliggande rapport.

I föreliggande studie har effekterna av ålder respektive årskurs undersökts i en läsförståelsestudie från 2001 med svenska elever, en läsförståelsestudie från 2006 där data från Norge och Island använts, en studie av elevers kunskaper i matematik och fysik i gymnasieskolan 2008 med svenska elever, samt en studie av demokrati, samhälle och medborgarskap 2009, med svenska elever.

Analyserna visar att ålders- respektive årskurseffekterna på elevers resultat är beroende av flera faktorer, såsom vilka åldrar och årskurser som studeras, vilka ämnen det gäller och vilket land som studeras. Det betyder att det inte på ett enkelt sätt går att korrigera för ålders- eller årskurseffekt genom att lägga till eller dra ifrån en konstant. Effekterna varierar sålunda avsevärt mellan åldrar/årskurser, ämnen och länder.

Anders Auer har genomfört analyserna och skrivit rapporten.

Stockholm augusti 2010

Sverker Härd
Enhetschef

Anita Wester
Projektledare

Inledning

Syftet med denna rapport är att ge Skolverkets medarbetare en snabb och enkel överblick om problematiken gällande elevers ålder och årskurs i internationella kunskapsmätningar. Ett ytterligare syfte är att redovisa hur man kan gå tillväga för att skatta effekten av ålder respektive årskurs på provresultat och vilken typ av data som krävs för att sådana analyser skall kunna vara möjliga att genomföra. Att kunna bedöma den relativa effekten av ålder respektive årskurs är väsentligt för att kunna få en uppfattning om hur mycket dessa faktorer kan tänkas påverka Sveriges relativa position i olika internationella kunskapsundersökningar.

Rapporten innehåller en sammanfattning och presentation av viss tidigare publicerad forskning inom området samt replikationer av några tidigare studier i syfte att förklara metoden. Forskningsöversikten kan ses som en introduktion till området och gör inte anspråk på att vara heltäckande.

I detta kapitel ges en introduktion till problemområdet och i följande kapitel en kortfattad forskningsöverblick. I metodkapitlet beskrivs hur man går till väga för att skatta effekten av ålder respektive årskurs. I resultatkapitlet redovisas resultaten från skattningar gjorda på svenska PIRLS¹ 2001 data (replikering av Gustafsson 2009), data från Norge och Island från PIRLS 2006, svenska data från TIMSS 1995, årskurs 6, 7 och 8 (replikering av Cliffordson 2008), svenska data från TIMSS advanced 2008, svenska elever från årskurs 2 och 3 på gymnasiet samt data från ICCS 2009, svenska elever i årskurs 8 och 9. Sedan följer ett kapitel där problem med metodansatsen diskuteras samt på vilka sätt metoden kan förfinas och utvecklas utifrån ytterligare redan publicerad forskning. I det sista kapitlet förs en avslutande diskussion om hur resultaten kan värderas. Appendix innehåller mer detaljerad information om respektive replikering.

I alla internationella kunskapsmätningar är det viktigt att försöka göra resultaten mellan elever i olika länder så jämförbara som möjligt, till exempel genom att alla elever, oavsett land, skriver samma prov (även om inte alla elever inom samma land nödvändigtvis besvarar samma frågor) och har lika mycket tid till sitt förfogande etc. Ett ytterligare sätt att maximera jämförbarheten är att fixera antingen årskursen som eleverna går i eller elevernas ålder. I TIMSS och PIRLS har man valt att fixera årskursen, vilket till exempel innebär att alla elever som deltar i TIMSS går i årskurs 4 eller 8 (ett prov för åk 4 och ett för åk 8). Men eftersom skolstarten är olika i olika länder innebär en sådan ansats att elever i olika länder kan ha olika genomsnittlig ålder. För Sveriges del innebär det att svenska elever som deltar i till exempel TIMSS åk 4 är ett halvår äldre än den internationella genomsnittsåldern.² I PISA har man istället valt att fixera åldern så att alla som deltar i PISA är 15 år. Med Sveriges relativt sena skolstart innebär detta att cirka 95 procent av alla svenska elever som deltar i PISA går i åk 9. I många andra länder befinner sig de flesta elever i genomsnitt i åk 10 och i några länder till och med i åk 11.³

1 I dokumentet används de allmänt accepterade förkortningarna för de olika internationella studierna. För en närmare presentation av de internationella undersökningarna samt förklaring av vad förkortningarna står för, se Skolverket 2004.

2 Den svenska genomsnittsåldern var 10,8 år i TIMSS 2007 medan genomsnittsåldern för EU/OECD var 10,3 år, en skillnad på ett halvår.

3 Exempel på sådana länder är Australien, Kanada och Storbritannien.

Konsekvensen av att fixera årskursen (som fallet är i TIMSS och PIRLS) blir för Sveriges del att svenska elever får en relativ fördel då de givet samma antal år i skolan är något äldre och därmed sannolikt mognare. När istället åldern fixeras, som är fallet med PISA, blir istället konsekvensen att svenska elever inte har hunnit med att lära sig lika mycket som i länder där de flesta eleverna som deltar i PISA går i åk 10. Detta ger svenska PISA-elever en relativ nackdel.⁴

Ovanstående problematik gör att jämförelser mellan två länder med olika ålder för skolstart blir något orättvis, oavsett om det gäller TIMSS, PIRLS eller PISA. Länder med sen skolstart gynnas i TIMSS och PIRLS medan länder med tidig skolstart gynnas i PISA. Detta leder till frågan om hur stor effekten är på elevers akademiska förmåga, mätt som provresultat, av ett års ytterligare skolgång samt effekten av ett ytterligare levnadsår. Att kunna få en uppfattning om den relativa storleken på dessa effekter är betydelsefullt då man skulle kunna ta viss hänsyn till detta när olika länder jämförs eller när till exempel Sveriges relativa position i PISA jämförs med motsvarande relativa position i till exempel TIMSS.

⁴ Det är dock värt att notera att Finland, precis som Sverige, har en relativt sen skolstart och att de flesta finska eleverna också befinner sig i åk 9 när de deltar i PISA. Trots detta är Finland det land som klarat sig allra bäst i de tre tidigare PISA-undersökningarna.

Vad säger forskningen?

För den delen av forskningen som fokuserar jämförelser av ålderseffekt och årskurseffekt, är den dominerande metoden den så kallade ”*between grade regression discontinuity design*”. Denna rapport koncentrerar sig därför endast på denna metod. Cahan & Davis, 1987 och Cahan & Cohen, 1989 anses ligga bakom denna metod som kommer att förklaras mer i detalj i nästa kapitel. Cahan & Cohen skattade årskurseffekten till ungefär dubbelt så stor som ålderseffekten och även om skattningarna varierar beroende på vilka data som används så kan det mycket grovt sägas att årskurseffekten framstår som större än ålderseffekten. Men undantag finns. Martin, Mullis & Foy presenterade under IEA konferensen i Taipei 2008 analyser av ålders- och årskurseffekt baserade på data från PIRLS 2006 (läsförståelse) för Norge och Island och fann att för båda länderna var ålderseffekten dubbelt så stor som årskurseffekten gällande årskurs 4 och 5.

Cliffordson presenterade också resultat gällande svenska TIMSS 1995 data (matematik och naturvetenskap), årskurs 6, 7 och 8 på samma konferens.⁵ Hon finner att årskurseffekten är ungefär dubbelt så stor som ålderseffekten även om årskurseffekten är mindre mellan årskurs 7 och 8 jämfört med mellan årskurs 6 och 7. Cliffordsons resultat, liksom Mullis et al. kommer att presenteras och replikeras i resultatkapitlet. Även Luyten (2006) som analyserat flera länder från TIMSS 1995 (men inte Sverige), och gällande årskurs 3–4 i matematik och naturvetenskap, finner också att årskurseffekten generellt är starkare än ålderseffekten även om proportionerna varierade mellan länderna.⁶

Gustafsson (2009) har också presenterat resultat utifrån svenska PIRLS 2001 data i läsförståelse, där Sverige deltog med elever både i åk 3 och 4. Även Gustafsson finner att årskurseffekten är ungefär dubbelt så stor som ålderseffekten.

Även om mycket forskning pekar på att årskurseffekten är större än ålderseffekten är det troligt att detta förhållande inte nödvändigtvis är konstant över alla skolår eller för den delen konstant mellan olika förmågor/skolämnen. Dessutom är det också tydligt (se Luyten 2006) att effekterna kan variera mellan olika länder. Att effekterna verkar variera mellan olika länder kommer att tas upp senare då metoden beskrivits och bättre kan förstås. I kapitlet som behandlar problem och utvidgningar nämns också ytterligare forskning på detta område men som kan ses som utvidgningar av grundmodellen.

⁵ Detta forskningsbidrag är nu publicerad, se Cliffordson, 2010.

⁶ De åtta länder som Luyten analyserar är Cypern, England, Grekland, Island, Japan, Norge, Skottland och Singapore.

Metod

Den metod som enligt forskningen är den dominerande när det gäller att separera *ålderseffekt* och *skolårseffekt* är, som nämnts, den så kallade ”*Between grade regression discontinuity design*”. Metoden går ut på att skatta en regressionslinje för prestation baserat på födelsemånaden inom ett givet år. Genom att skatta denna regressionslinje för två intilliggande år med en dikotom (0 eller 1) variabel för årskurs kan den differens som uppstår mellan åren (effekten av den dikotoma variabeln) tolkas som årskurseffekten medan riktningskoefficienten på regressionslinjen kan tolkas som ålderseffekten.

De viktigaste underliggande antaganden som ska vara uppfyllda för att metoden skall fungera är:

1. att eleverna startar skolan efter en strikt datumregel, till exempel som i Sverige där alla barn som är födda samma kalenderår i regel startar skolan samma år.
2. att sambandet mellan prestation och ålder är approximativt linjärt och därmed lika stort för de två intilliggande årskurserna.

Om barn i allt för stor utsträckning startar skolan efter hur mogna eller ”begåvade” de är suddas sambandet mellan ålder och prestation till viss del ut då många duktiga elever startar tidigare och därmed framstår som unga i sin årskurs medan de som är mindre ”begåvade” startar senare (eller går om) framstår som äldre i sin årskurs. För att kunna studera ålders- och årskurseffekter på ett optimalt sätt skulle således alla barn allokeras till årskurser strikt efter sitt födelseår/datum.

I fallet Sverige finns det naturligtvis möjligheter att barn startar tidigare eller senare men då det är relativt ovanligt (ungefär 95–96 procent startar utifrån sitt kalenderår) så påverkar inte dessa elever, som generellt är ovanligt starka/svaga, skattningarna på något avgörande sätt.⁷

Antagandet om att sambandet är linjärt har testats av Cliffordson (2008) och hon finner utifrån TIMSS 1995 data för Sverige att antagandet håller. I ett teoretiskt perspektiv finns det anledning att ifrågasätta antagandet om ett linjärt samband mellan prestation och ålder. Speciellt om man studerar kunskapsutvecklingen över ett antal år. I detta sammanhang avses dock att sambandet kan ses som approximativt linjärt över två intilliggande skolår.

Antagandet om att minst 95 procent av en årskull följer en given datumregel håller endast för ett visst antal länder, däribland Sverige.⁸ Martin, Mullis and Foy (Martin et al 2008) går igenom alla länder som deltog i PIRLS 2006 och visar att av de 40 länder och provinser som deltog så uppfyller 10 detta krav.⁹

Ett ytterligare krav som måste vara uppfyllt för att ett land skall kunna analyseras i termer av ålders- och årskurseffekt är naturligtvis att data måste finnas för två intilliggande årskurser. Detta begränsar därför antalet länder ytterligare och när det gäller PIRLS 2006 var det bara Island och Norge som deltog med både årskurs 4 och 5. I PIRLS 2001 deltog Sverige med både årskurs 3 och 4.

7 En tumregel som Luyten (2006) föreslår är att minst 95 % av eleverna följer regeln att starta skolan efter sin ålder.

8 Även vid strikt åldersbestämd skolstart uppgår ålderskillnaden till 1 år inom kohorten. Detta hantearas i de kommande analyserna.

9 De länder som enligt PIRLS 2006 har minst 95 % av eleverna samlade enligt en strikt datumregel är i fallande ordning: Island, Norge, England, Polen, Taiwan, British Columbia (Kanada), Singapore, Ontario (Kanada), Slovenien och Sverige.

Resultat

PIRLS 2001, Sverige åk 3 och 4

För att åskådliggöra konceptet med metoden har samtliga elever i årskurs 3 och 4 som deltog i PIRLS 2001 grupperats efter sitt födelsedatum (månad). De elever som är ”underåriga” respektive ”överåriga” har förts till egna grupper. I figur 1 och 2 nedan visas elevernas genomsnittliga prestation på PIRLS-provet uppdelat efter födelsemånad. Yngre elever är till vänster på den horisontella axeln och äldre är till höger.

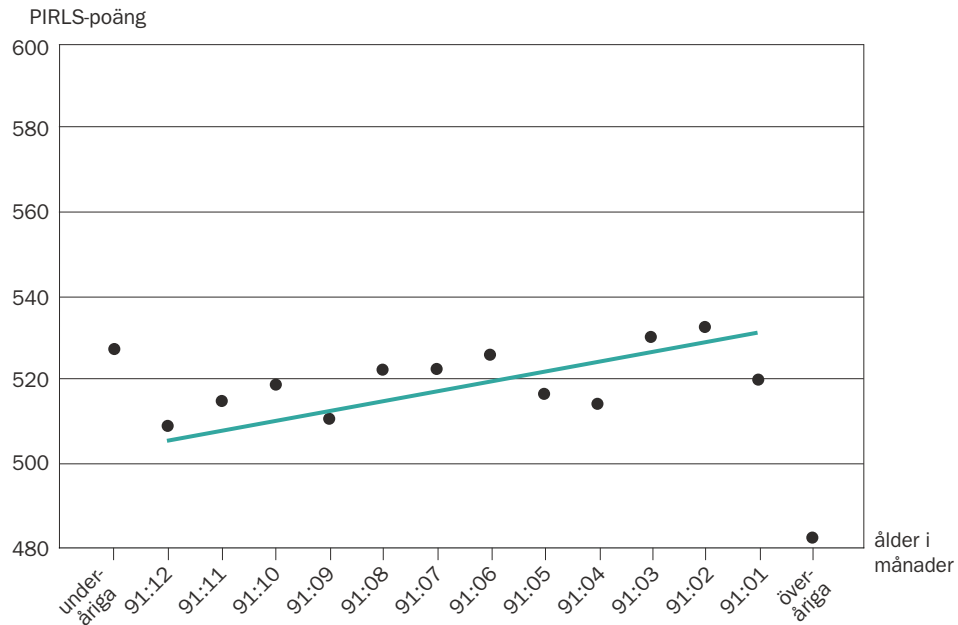
I figurerna syns att inom varje årskull finns ett samband mellan ålder i månader och genomsnittlig prestation, äldre elever tenderar att prestera bättre i genomsnitt. Men detta samband bryts helt av de underåriga respektive överåriga då de underåriga presterar betydligt bättre än andra unga i sin årskurs och överåriga presterar betydligt sämre än samtliga övriga grupper i årskursen.

Dessa figurer visar att om underåriga och överåriga inte exkluderas när ålders-effekten inom en årskurs skall skattas så kommer sambandet mellan ålder och resultat att mer eller mindre suddas ut eller underskattas. En första åtgärd är därför att i följande analys utesluta underåriga och överåriga elever, vilket också är praxis i andra studier. Att exkludera dessa elever är också helt logiskt då de inte tillhör årskohorten som är den egentliga analyspopulationen som vi vill studera.

Observera att även om överåriga och underåriga exkluderas i de fortsatta analyserna så är de återstående eleverna också påverkade av att det saknas elever som tillhör kohorten men som går i en högre eller lägre årskurs. De elever som tillhör kohorten men som går i en högre årskurs (underåriga) tenderar att vara födda tidigare på året och de elever som går i en lägre årskurs (överåriga) tenderar att vara födda senare på året. Detta i kombination med att underåriga elever tenderar att prestera något bättre än övriga elever födda samma månad och att överåriga elever tenderar att prestera något sämre än övriga födda samma månad innebär att dessa elever inte är representativa för resten av kohorten. Konsekvensen blir att det genomsnittliga resultatet för de normalåriga eleverna i de tidiga födelsemånaderna underskattas något och att det genomsnittliga resultatet för eleverna födda de senare månaderna överskattas jämfört med om alla elever från samma årskohort befunnits sig i samma årskurs.

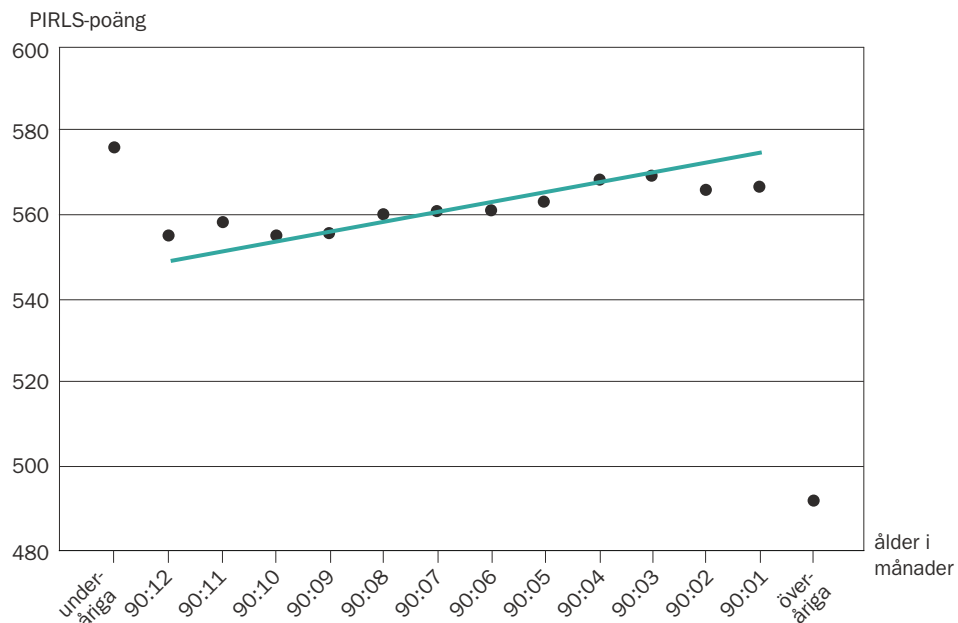
Detta åskådliggörs i figur 3. De normalåriga födda 1991 (ljusgrön ellips) går i åk 3 (rektangel). I åk 3 går även elever födda 1992 och 1990, vilket syns i åk 3-rektangeln av att det finns ett grönt respektive svart segment. Dessa elever exkluderas när metoden appliceras. Men elever som är födda 1991 men som av någon anledning startat skolan tidigare eller senare finns i åk 2 respektive åk 4-rektanglarna (ljusgrönt segment).

PIRLS 2001 åk 3, alla

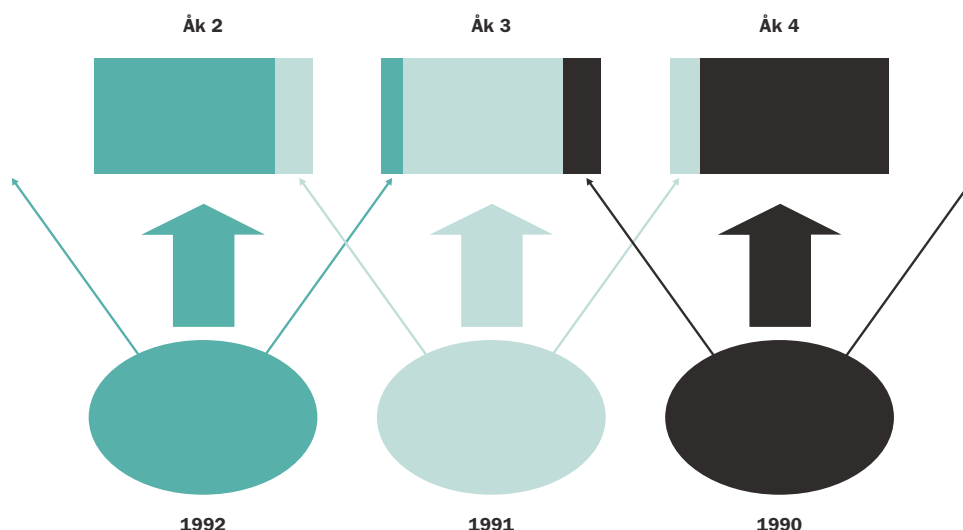


Figur 1. Sambandet mellan genomsnittligt resultat och ålder i månader enligt PIRLS 2001, åk 3, Sverige. Varje punkt representerar genomsnittet för de elever som är födda samma månad, förutom under- respektive överåriga. Regressionslinjen är bara approximativ och är endast baserad på de normalåriga eleverna, det vill säga utan hänsyn till ändpunkterna "underåriga" respektive "överåriga". Observera att den vertikala skalan inte startar på noll.

PIRLS 2001 åk 4, alla



Figur 2. Sambandet mellan genomsnittligt resultat och ålder i månader enligt PIRLS 2001, åk 4, Sverige. Varje punkt representerar genomsnittet för de elever som är födda samma månad, förutom under- respektive överåriga. Regressionslinjen är bara approximativ och är endast baserad på de normalåriga eleverna, det vill säga utan hänsyn till ändpunkterna "underåriga" respektive "överåriga".



Figur 3. Födelsekohorternas fördelning på olika årskurser. Cirkelarna och färgerna representerar alla elever som är födda samma år medan rektanglarna representerar alla elever som tillhör samma årskurs.

Konsekvensen av detta blir att de samband av ålderseffekten som vi skattar med de normalåriga eleverna kommer att underskattas något. Av den anledningen är det viktigt att inte andelen under- respektive överåriga är alltför stor, därav tumregeln att de inte bör överskrida 5 procent av alla elever.¹⁰

I tabell 1 visas genomsnittliga resultat för respektive årskurs uppdelat efter om eleverna är normalåriga eller under- respektive överåriga. Där syns också att prestationsnivån generellt är högre för de normalåriga åk 4-eleverna jämfört med de normalåriga åk 3-eleverna, ungefär 42 poäng.

Tabell 1. Deskriptiva resultat för PIRLS 2001-elever i årskurs 3 och 4

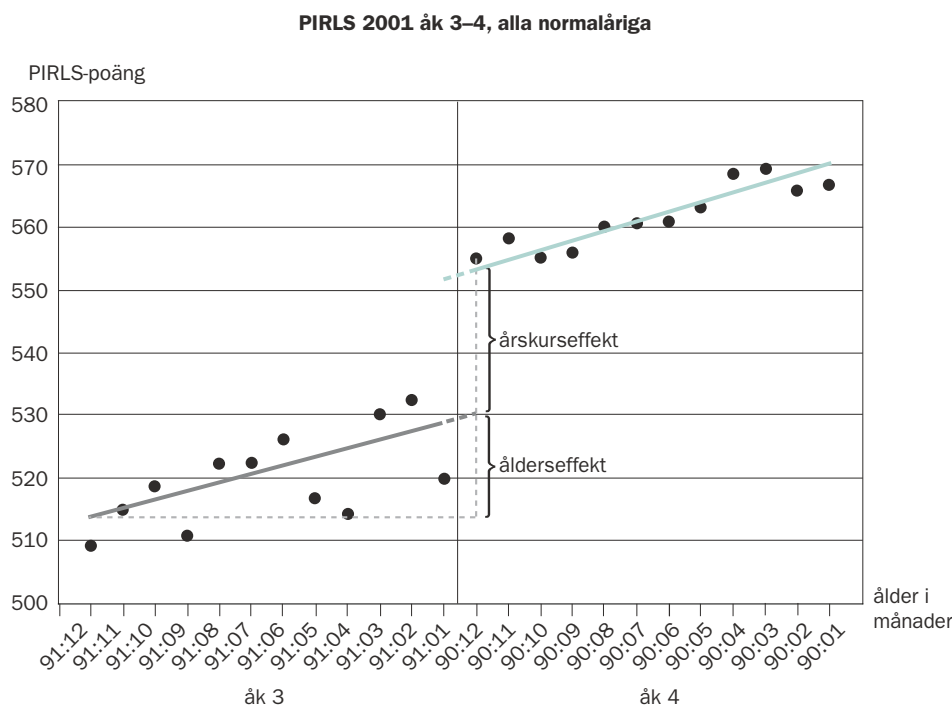
Årskurs	Elevtyp	Antal elever	Summa houwgt*	Andel %	Resultat PVread**	Standardavvikelse
3	Underåriga	181	139	2,6	527,3	71,3
	Normalåriga	4 987	5 040	95,6	520,2	76,9
	Överåriga	103	92	1,7	482,0	88,0
4	Underåriga	170	117	1,9	576,0	63,7
	Normalåriga	5 778	5 830	96,5	561,9	64,9
	Överåriga	95	94	1,6	491,6	79,8

* houwgt = houseweight och är en vikt som finns tillgänglig i de internationella databaserna. Houseweight motsvarar den vanligaste vikten "total student weight" med skillnaden att den inte "blåser" upp antalet observationer till populationsstorleken (som total student weight gör) utan multipliceras med en konstant så att den ursprungliga urvalsstorleken behålls.

** PVread = resultatmättet i läsförståelse som fås i form av plausibla värden. Utifrån elevernas svar på provet skattas deras förmåga i form av en sannolikhetsfördelning varifrån 5 plausibla värden dras.

¹⁰ Observera att med en tumregel på 5 procent menas andelen elever som exkluderas från respektive årskurs. Dessa elever tillhör inte kohorten (elever födda samma år). Däremot finns andra elever som tillhör kohorten men som går i andra årskurser. Optimalt är det hur stor andel dessa senare nämnda elever utgör av hela kohorten som är det sanna bortfallet eller exkluderingsgraden men dessa känner man sällan till om man inte har uppgifter från ytterligare årskurser över och under den aktuella årskursen. Därför används andelen elever som exkluderats från årskursen som en approximation på andelen elever som egentligen skulle tillhöra kohorten men som går i andra årskurser.

Att prestationsnivån generellt är högre för årskurs 4-eleverna syns även i figur 4 där alla *normalåriga* elever från både årskurs 3 och 4 förts ihop.



Figur 4. Årskurs 3 och 4-elevers genomsnittliga prestation efter födelsemånad (endast normalåriga elever). Observera att regressionslinjerna i figuren inte är korrekta skattade linjer utan endast där för att åskådliggöra konceptet.

Figur 4 visar tydligt att det finns en nivåskillnad mellan årskurs 3 och 4-elever i genomsnitt. Men om vi bara skulle beräkna den genomsnittliga skillnaden mellan de två årskullarna som vi skulle kunna benämna *totaleffekten* och som uppgår till cirka 42 poäng, så vet vi inte hur mycket av denna som består av *ålderseffekt* respektive *årskurseffekt*.¹¹

För att separera ålders- och årskurseffekt skattas en regressionslinje med samma lutning för båda årskullarna men där interceptet (skärningen med y-axeln i figuren eller konstanten i regressionsekvationen) tillåts vara annorlunda för de två årskurserna. Pedagogiskt sett kan årskurseffekten åskådliggöras genom att förlänga regressionslinjen för åk 3 till första månaden för årskurs 4 och sedan beräkna det vertikala avståndet mellan linjerna (se figur). Matematiskt kan årskurseffekten beräknas som skillnaden i interceptet för de två regressionslinjerna. Statistiskt sett skattas årskurseffekten genom att i regressionsmodellen införa en dikotom variabel för årskurs och där effekten av den dikotoma variabeln utgör årskurseffekten.

¹¹ Totaleffekten motsvarar den genomsnittliga skillnaden mellan de två "normalåriga" årskullarna, vilket innebär att den äldre årskullen i genomsnitt är ett kalenderår äldre och dessutom har gått i skolan exakt ett år längre. Därmed innefattar totaleffekten både effekten av att eleverna i den äldre årskursen har gått i skolan ett år längre (årskurseffekten) och dessutom effekten av att de i genomsnitt har levt ett år längre (ålderseffekten).

Ålderseffekten åskådliggörs i figuren genom riktningskoefficienten på regressionslinjen, vilken beräknas genom att beräkna förändringen i poäng (den vertikala förändringen av linjen) över ett helt år, det vill säga 12 månader (se figuren).

I verkligheten beräknas inte ålders- och årskurseffekt utifrån medelvärdena inom respektive födelsemånad utan direkt i en regressionsmodell på elevnivå, det vill säga med data på elevnivå för elevens ålder (oftast i år och månader) samt årskurs. Regressionsmodellen ser då ut så här:

$$PVread = b_0 + b_1 * (ålder) + b_2 * (årskurs)$$

Där $PVread$ = resultat i läsförståelse enligt PIRLS (som uttrycks som ett plausibelt värde)

$ålder$ = variabel för ålder uttryckt i månader och här omkodat så att yngsta eleven har värdet 1.

$årskurs$ = dikotom variabel för årskurs där elever från den lägre årskursen har värdet 0 och elever från den högre årskursen har värdet 1.

b_0 = konstant (intercept för den lägre årskursen) som ska skattas i modellen.

b_1 = ålderseffekt uttryckt som effekt per månad. Denna effekt multipliceras med 12 för att få ålderseffekt för ett helt år.

b_2 = årskurseffekt (uttryckt som effekt per läsår)

Observera att alla under- respektive överåriga i både årskurs 3 och 4 har exkluderats innan modellen skattas. Innan den sammanslagna modellen skattas bör separata enkla regressioner skattas för de två årskurserna vad gäller ålder och resultat. Sedan jämförs de skattade regressionskoefficienterna (effekten av ålder inom respektive år) med varandra och testas för om de skiljer sig signifikant. Detta har gjorts och effekten av ålder inom respektive årskurs skiljer sig inte signifikant vilket innebär att modellen kan skattas med en gemensam effekt av ålder för båda årskurserna.¹²

Skattad regressionsmodell¹³:

$$PVread = 495,8 + 1,30 * (ålder) + 26,3 * (årskurs) \quad (\text{ålderseffekt uttryckt per månad})$$

$$PVread = 495,8 + 15,6 * (ålder) + 26,3 * (årskurs) \quad (\text{ålderseffekt uttryckt per år})$$

Både ålderseffekten och årskurseffekten är statistiskt signifikanta med t-värden på 5,8 respektive 7,2.¹⁴

12 Två separata enkla regressioner skattades, en för varje årskurs. Effekten av ålder (uttryckt per månad) uppmättes till 1,24 för åk 3 och 1,34 för åk 4. Standardfelet för respektive effekt var 0,34 respektive 0,28. Standardfelet för differensen 0,1 approximeras till 0,44. Skillnaden är därmed långt ifrån signifikant.

13 Programvaran är IDB Analyzer, vilken dels tar hänsyn till klustereffekter eller med annat språk att elever från samma skolor inte är helt oberoende av varandra, dels att prestationen uttrycks i form av plausibla värden. I modellen har vikten houseweight använts.

14 Som en tumregel indikerar t-värden på 2 eller mer att koefficienten är statistiskt signifikant.

Tabell 1 visar att totaleffekten, som utgör den genomsnittliga skillnaden i resultat för de två årskurserna (normalåriga) är 41,7 poäng (se tabell 1). Resultaten från regressionskattningen ovan visar att denna skillnad utgörs av en ålderseffekt på 15,6 poäng och en årskurseffekt på 26,3 poäng.¹⁵ Detta innebär att nettoeffekten av att vara ett helt år äldre, det vill säga när antalet år i skolan konstanthållits, i genomsnitt är ungefär 16 poäng. Den genomsnittliga nettoeffekten av att gå ett extra år i skolan, det vill säga efter att elevens ålder konstanthållits, är ungefär 26 poäng. Således är årskurseffekten betydligt större än ålderseffekten, den utgör ungefär 63 procent av den totala effekten.

Den första koefficienten i den skattade regressionsekvationen, b_0 , som skattats till 495,8 kan tolkas som det förväntade värdet för en elev som har värdet 0 i ålder och 0 i värdet på den dikotoma variabeln. För detta specifika fall innebär detta det förväntade värdet för en elev som är född i januari 1992, det vill säga en månad senare än de yngsta eleverna i datamaterialet och går i årskurs 3. Men i praktiken har den inget tolkningsvärde förutom att ingå i ekvationen för att bestämma det förväntade värdet för en given elev.

Dessa skattningar av ålders- respektive årskurseffekt är identiska med de skattningar som Gustafsson (se Djurfeldt & Barmark 2009) erhåller. Däremot skiljer sig t-värdena marginellt, vilket sannolikt beror på att olika program använts: M-plus respektive IDB Analyzer.¹⁶

Som en utvidgning har även separata regressionsekvationer skattats för flickor och pojkar. Pojkar uppvisar en något större effekt av ålder men något mindre effekt av årskurs jämfört med flickorna. Skillnaderna är dock relativt små och inte statistiskt signifikanta.¹⁷

15 Differensen på 0,2 poäng beror antagligen på att statistikprogrammet (IDB Analyzer) använder olika procedurer för regressionskattningar och medelvärdeskattningar. Men det är bara en gissning.

16 Gustafssons t-värde för ålderseffekten är 5,6, att jämföra med 5,8 enligt denna replikering. För årskurseffekten erhåller Gustafsson t-värdet 8,2, att jämföra med 7,2 enligt replikeringen.

17 För flickor uppmättes ålderseffekten till 15,2 respektive 18,2 för pojkar. Årskurseffekten uppmättes till 27,4 för flickor respektive 23,5 för pojkar.

PIRLS 2006 Island och Norge, årskurs 4 och 5

I tabell 2 presenteras vissa deskriptiva data för de elever i årskurs 4 och 5 som deltog i PIRLS 2006 i Norge och på Island.

Tabell 2 Deskriptiv bild av data för Island och Norge i PIRLS 2006.

Årskurs	Elevtyp	Antal elever	Summa houwgt	Andel inom årskursen	Resultat PVread	Standard- avvikelse
Island						
4	Normalåriga	3 658	3 657	99,6 %	510,5	68,0
5	Normalåriga	1 375	1 371	99,6 %	549,1	63,5
Norge						
4	Normalåriga	3 812	3 815	99,3 %	498,1	66,6
5	Normalåriga	1 796	1 793	99,3 %	540,4	64,6

Skattade regressionsmodeller¹⁸:

Island

$$PVread = 495,3 + 28,0 * (ålder) + 10,3 * (årskurs) \quad (\text{ålderseffekt uttryckt i år})$$

t-värden för skattade koefficienter: 8,4 respektive 1,8

Norge

$$PVread = 481,2 + 30,1 * (ålder) + 12,1 * (årskurs) \quad (\text{ålderseffekt uttryckt i år})$$

t-värden för skattade koefficienter: 4,6 respektive 1,96

De resultat som Martin et al. redovisar är identiska.

Tabell 3. Jämförelse av resultat för Sverige, Island och Norge

Studie	Land	Ämne och årskurser	Ålderseffekt	Årskurseffekt	Täckning av födelsekohort
PIRLS 2001	Sverige	Read 3-4	15,6 (2,6)	26,3 (3,6)	96 %
PIRLS 2006	Island	Read 4-5	28,0 (3,5)	10,3 (5,3)	99,6 %
PIRLS 2006	Norge	Read 4-5	30,2 (6,1)	12,1 (7,2)	99,3 %

Standardfel inom parentes

I tabell 3 sammanfattas resultaten för Sverige, Island och Norge. Det slående mönstret är att proportionerna mellan ålders- och årskurseffekt är mer eller mindre omvända. Både Island och Norge uppvisar betydligt större ålderseffekt och samtidigt betydligt mindre årskurseffekt jämfört med Sverige. Totaleffekten (summan av ålders- och årskurseffekt) är dock ungefär lika stor för varje land.

¹⁸ Programvaran är IDB Analyzer, vilken dels tar hänsyn till klustereffekter eller med annat språk att elever från samma skolor inte är helt oberoende av varandra, dels att prestationen ges i form av plausibla värden. I modellen har vikten totwgt använts.

Tyvärr så är det mer än en variabel som skiljer sig mellan Sverige å ena sidan och Island och Norge å den andra. Dels gäller resultaten för Sverige årskurs 3–4 medan de gäller 4–5 för Island och Norge. Men man skulle kanske ha gissat att ålders-effekten borde vara som störst i tidiga år för att sedan få mindre och mindre betydelse? Eller så innebär övergången från ”låg” till ”mellanstadiet” en mer markant nivåhöjning. En annan sak som skiljer sig är tidpunkten för studien. Resultaten för Sverige gäller 2001 medan för Island och Norge gäller det 2006. Det är svårt att bedöma på vilket sätt detta skulle kunna ha påverkat skattningarna.

En sista faktor som skiljer sig och som skulle kunna förklara en del av de påvisade skillnaderna är täckningsgraden, det vill säga hur stor andel av födelsekohorten som är normalåriga. Här ligger Sverige på 96 procent medan Island och Norge ligger mycket nära 100 procent.¹⁹ I kapitel 2 argumenterades för att desto fler elever som exkluderas (eller mindre täckning), desto mer underskattad kan man förvänta sig att ålderseffekten är, trots att underåriga och överåriga exkluderas. Det är rimligt att anta att de exkluderade eleverna påverkar resultaten i någon mån, men att de knappast förändrar resultaten på ett signifikant sätt.

19 Egentligen är dessa siffror bara en slags approximation av täckningen då andelarna representerar andelen normalåriga i årskursen. Detta är inte samma sak som andelen av födelsekohorten men kan ses som en approximation. Den eventuella underskattningen av ålderseffekten beror inte på att överåriga och underåriga exkluderas från årskursen (de är födda ett annat år) utan att det saknas över- och underåriga från födelsekohorten som går i andra årskurser.

TIMSS 1995, Sverige, årskurs 6, 7 och 8

I TIMSS 1995 deltog Sverige med hela tre årskurser, 6, 7 och 8. Nedan presenteras resultaten för de skattningar som gjorts enligt exakt samma princip som tidigare i dokumentet och är således ett försök till replikering av Cliffordson. Eftersom resultaten från TIMSS 1995 har skalats om i efterhand har effekterna skattats för data baserad på både originalskalet och den nya skalan.²⁰

Tabell 4a. TIMSS 1995, Sverige, originalskalet

Studie	Ämne och årskurs	Ålderseffekt	Årskurseffekt	Kommentar
Matematik				
TIMSS 1995	Ma åk 6–7	13,1 (3,8)	28,4 (5,8)	Originalskala
TIMSS 1995	Ma åk 7–8	16,8 (4,4)	17,5 (6,2)	Originalskala
Naturvetenskap				
TIMSS 1995	Sci åk 6–7	11,5 (3,8)	34,4 (5,9)	Originalskala
TIMSS 1995	Sci åk 7–8	21,1 (5,6)	13,6 (6,8)	Originalskala

Standardfel inom parentes

Tabell 4b. TIMSS 1995, Sverige, nyskala

Studie	Ämne och årskurs	Ålderseffekt	Årskurseffekt	Kommentar
Matematik				
TIMSS 1995	Ma åk 6–7	13,7 (3,2)	24,0 (5,4)	Nyskala
TIMSS 1995	Ma åk 7–8	18,0 (3,8)	8,4 (5,9)	Nyskala
Naturvetenskap				
TIMSS 1995	Sci åk 6–7	13,0 (4,1)	26,7 (6,5)	Nyskala
TIMSS 1995	Sci åk 7–8	20,9 (4,4)	8,9 (5,8)	Nyskala

Standardfel inom parentes

Ett intressant mönster är att för åk 6–7 är årskurseffekten större än ålderseffekten men för åk 7–8 är de lika eller så är ålderseffekten större. En tänkbar förklaring till detta resultat skulle kunna vara att det sker ett större hopp akademiskt mellan åk 6 till 7 jämfört med åk 7–8. Men samtidigt resonerades det tidigare om att ålderseffekten borde avta med senare år, vilket inte tycks vara fallet enligt dessa skattningar. Det skall dock påpekas att skillnaderna mellan skattningarna ligger inom osäkerhetsintervallet och är inte signifikanta.

Ett annat intressant resultat är att skattningarna inte är helt oberoende av hur resultaten är skalade. Vad detta kan bero på är inte känt och ligger utanför ambitionen med detta dokument men det är uppenbarligen något som bör redas ut och är ett område för framtida forskning.

Jämfört med Cliffordson är den generella bilden av resultaten lika men inte identisk. Se appendix där en detaljerad genomgång av replikeringen av Cliffordson är redovisad.

²⁰ Cliffordson har endast använt originalskalet i sina analyser. Skillnaden i skalning innebär inte att de nominella värdena ligger på helt olika nivåer utan att den statistiska metod som använts för att skatta eleverns prestation utifrån svaren på provfrågorna skiljer sig. Originalskalan bygger på en så kallad 1-parametermodell medan den nya skalan bygger på en så kallad 3-parametermodell. Se IEA 2004.

TIMSS Advanced 2008, Sverige, årskurs 11–12

I TIMSS Advanced 2008 gjorde Sverige ett tilläggsurval av elever från åk 11 (gymnasiets årskurs 2), förutom det konventionella åk 12-urvalet (gymnasiets årskurs 3). Observera dock att det finns uppenbara problem med att försöka skatta ålders- och årskurseffekter i detta material då eleverna inte nödvändigtvis har läst samma kurser även om de går i samma årskurs. Men det är fortfarande intressant att se hur skattningarna slår och därmed finns ändå en viss poäng med att analysera resultaten med denna metod.

Redan vid en första anblick på de deskriptiva resultaten syns att det är något konstigt med resultaten. (se tabell 5).

Tabell 5. Genomsnittliga resultat i matematik för elever i årskurs 11 respektive 12 som deltog i TIMSS Advanced 2008.

Årskurs	Antal elever i dataset	PVma	Standardfel
Åk 11	1 058	411,4	6,2
Åk 12	2 043	414,0	5,3

Skillnaden mellan resultaten i matematik mellan åk 11 och 12 eleverna är marginell. Eftersom totaleffekten är obefintlig kommer antagligen inte heller ålders- respektive årskurseffekterna att bli annat än marginella och knappast signifikanta.

$$PVma = 421 - 3,8 * (\text{ålder}) + 6,4 * (\text{årskurs}) \quad \begin{array}{l} (\text{ålderseffekt uttryckt per år}) \\ (10,3) \quad (13,5) \quad \quad \quad (\text{standardfel inom parentes}) \end{array}$$

Varken effekten av ålder eller årskurs är signifikant. Orsaken till avsaknad av effekt är sannolikt det faktum att en ganska betydande andel (cirka 40 procent) av åk 12-eleverna inte läste matematik sista läsåret och snarare har tappat kunskaper än ackumulerat nya kunskaper.²¹ Detta beror i sin tur delvis på att Matematik E inte är en obligatorisk kurs. Tyvärr saknas uppgift om vilka av åk 11-eleverna som avser att läsa Matematik E i år 12. Med de uppgifterna hade analysen kunnat göras på endast de elever i åk 11 och 12 som antingen läste matematik E eller avsåg att läsa matematik E i åk 12. Då hade också en glömskeeffekt kunnat skattas genom att jämföra resultat för endast de elever som läste upp till matematik D. Denna analys tillför således tyvärr ingen kunskap om det egentliga förhållandet mellan ålders- och årskurseffekt i de högre skolåren.

²¹ Se Skolverket 2009

ICCS 2009, Sverige, åk 8–9

I den internationella studien ICCS 2009, som mäter elevers kunskaper i demokrati och medborgarfrågor, gjordes ett tilläggsurval så att elever från både årskurs 8 och 9 deltog i undersökningen. De deskriptiva resultaten redovisas i tabell 6.

Tabell 6. Genomsnittliga resultat i demokrati och medborgarfrågor för elever i årskurs 8 respektive 9 som deltog i ICCS 2009, Sverige.

Årskurs	Antal elever i dataset (normalåriga)	Exkluderade (under- och överåriga)	PVciv (resultat i poäng)	Standardfel
Åk 8	3 183	6,4 %	541	3,0
Åk 9	3 250	6,0 %	578	3,6

Den genomsnittliga skillnaden i resultat mellan årskurs 8 och 9 är 37 poäng. Observera att andelen exkluderade och som utgör ett mått på bortfallet uppgår till ungefär 6 procent vilket är något över tumregeln om 5 procent.

Skattad regression:

$$PVciv = 311 + 15,6 * (\text{ålder}) + 21,3 * (\text{årskurs}) \quad (\text{ålderseffekt uttryckt per år})^{22}$$

(5,6) (6,7) (standardfel inom parentes)

Den totala effekten på 37 poäng kan delas upp i en ålderseffekt på cirka 16 poäng och en årskurseffekt på cirka 21 poäng. Dessa resultat ligger således i linje med övriga resultat från Sverige gällande andra kunskaper, det vill säga att årskurseffekten är något större än ålderseffekten.

22 Variabeln ålder är här inte standardiserad så att de yngsta eleverna har åldern 0 utan här är åldern kodad som den riktiga åldern i år. Konstanten på 311 poäng är därmed modellens skattning av vad en nyfödd person skulle prestera men saknar naturligtvis rimlig tolkningsgrund.

Problem med modellen och utvidgningar av modellen

Viktning

Ett problem som uppmärksammats i samband med replikeringarna av dessa studier är att forskarna oftast använt sig av houseweight, vilken är en av de vikter som finns tillgängliga i dataseten från IEA-studierna TIMSS och PIRLS.²³ Med användandet av houseweight (houwgt) konstanthålls urvalsstorleken, vilket har den fördelen att standardfel i SPSS inte blåses upp. Om totwgt istället används blåses urvalet upp till populationsstorleken och då kommer SPSS att betrakta urvalet som mycket större än vad det är och därmed underskatta standardfelen.

Men nu bör man inte använda sig av SPSS direkt i alla fall då dessa studiers urvaldesigner är så komplicerade att användande av SPSS rätt av ändå kommer att ge fel standardfelskattningar. Egentligen skall inte skattningen av en populations medelvärde påverkas av om man använder houwgt eller totwgt. Problemet uppstår emellertid när vi använder oss av elever från två årskurser (som vi naturligtvis gör i dessa sammanhang) och då urvalen för respektive årskurs är av olika storlek. Eftersom antalet elever i två intilliggande årskurser kan förväntas vara ungefär lika stora bör de också ges ungefär samma vikt när regressionsparametrarna skattas. Detta är också fallet om totwgt används, även om urvalen inte är lika stora. Med totwgt blåses urvalen upp till populationsstorleken det vill säga storleken på den population so urvalet är avsett att representera. Men med houwgt behålls urvalsstorleken och om då urvalet är relativt mindre för en årskurs kommer dess data att ges mindre relativ vikt i regressionen.

Därmed bör man använda totwgt. I dokumentet har däremot den vikt använts som forskarna har använt, oftast houwgt, då ett syfte har varit att försöka replikera resultaten från de forskare som genomfört analyserna i första hand. För Island och Norge i PIRLS 2006 har regressionerna däremot i denna studie körts med totwgt eftersom forskarna som ursprungligen genomförde dessa analyser har använt totwgt, vilket ger lite annorlunda resultat jämfört med om houwgt används, men inte på något avgörande sätt.²⁴ Urvalsstorlekarna i Norge var 3 812 för åk 4 och 1 796 för åk 5. För Island var motsvarande urvalsstorlekar 3 658 respektive 1 371, således betydande asymmetri i urval i förhållande till population. Att resultaten ändå är relativt stabila kan tolkas som att ålderseffekten är ungefär lika stor för både årskurserna (det vill säga regressionen av ålder inom respektive årskurs), så att om den ena årskursen ges något större vikt inte har någon egentlig betydelse. Men rent teoretiskt bör det, som rapportförfattaren ser det, vara mer korrekt att använda totwgt.

²³ Undantaget är Mullis studie med PIRLS 2006 data från Island och Norge där totwgt används.

²⁴ För Island minskade ålderseffekten med cirka 1 poäng medan årskurseffekten minskade med cirka 1 poäng. För Norge ökade istället ålderseffekten med cirka 2 poäng medan årskurseffekten minskade med cirka 2 poäng.

Skalning

För TIMSS 1995 är det uppenbart att skalningen av prestationsvariabeln (de plausibla värdena) har en viss betydelse för skattningarna. Detta skulle kunna utredas lite närmare vad det kan bero på och är ett område för framtida forskning. En möjlig ansats vore att upprepa analyserna för flera länder från TIMSS 1995 med olika skalning, förslagsvis med utgångspunkt i Luytens (2006) studie, för att se om avvikelserna är generella.

Multikollinjäritet

Multikollinjäritet innebär att två förklarande variabler är högt korrelerade med varandra. Detta kan leda till svårigheter att separera effekterna från respektive variabel. En indikation på att multikollinjäritet är ett problem är om modellen visar på en relativt god passning (högt R^2) men samtidigt är inte de enskilda regressionskoefficienterna signifikanta. Det är ganska uppenbart att de två oberoende variablerna ålder och årskurs inte kan anses oberoende. För PIRLS 2001, Sverige, är korrelationen hela 0,87. Men i så gott som samtliga fall i detta dokument är skattningarna signifikanta med god marginal. Multikollinjäritet kan vara ett större problem när två oberoende variabler utgör mått på likartade variabler. Detta är dock inte fallet här och därmed bedöms inte multikollinjäritet vara ett problem i dessa modellskattningar.²⁵

Standardiserade koefficienter

Utifrån den empiriska observationen att de elever som inte är normalåriga tenderar att vara födda antingen tidigt eller sent på året, kommer åldersfördelningen för de normalåriga inom en årskurs inte att vara perfekt jämt fördelade. Detta skulle möjligtvis kunna påverka skattningarna och ett sätt att kompensera för detta skulle kunna vara att standardisera koefficienten för ålder.

Flernivåansats

Luyten (2006) (se forskningsöversikten inledningsvis) använder en flernivåmodell vilket innebär att han kan skatta hur årskurseffekten varierar mellan skolor inom respektive land. Han finner att variationen mellan skolor är substantiell men att variationen är olika stor i olika länder. Han finner också att i vissa länder är årskurseffekten större i skolor där eleverna har en mer fördelaktig socioekonomisk bakgrund. Det skulle naturligtvis vara intressant att applicera en liknande analys på svenska data.

Luyten har även tillsammans med några andra forskare (Luyten et. al., 2008) gjort en liknande flernivåanalys på data från PISA 2000 för England gällande läsförståelse, läsengagemang och läsattityd i årskurserna 10 och 11. De finner bara en signifikant årskurseffekt för läsförståelse och dessutom en ganska svag effekt, men ändå relativt större än ålderseffekten. De finner också, till skillnad från Luyten (2006), att årskurseffekten är svagare i skolor med mer fördelaktig socioekonomisk bakgrund. Det som gör denna studie lite extra intressant är att det är den enda

²⁵ Se sid 276–277 i Gustafsson (2009) för en diskussion.

studie (mig veterligen) som skattat ålders- och årskurseffekt i de högre skolåren. Eftersom analysen gäller just läsförståelse är det kanske inte helt överraskande att både årskurseffekt och ålderseffekt är relativt små, jämfört med i tidigare skolår. Det är däremot inte säkert att så är fallet för till exempel matematik.

Avslutande diskussion

Ovanstående replikeringar och redovisning visar hur det är möjligt att separera ålders- och årskurseffekt i ett datamaterial när det finns elever från två intilliggande årskurser. Det krävs dessutom att det finns uppgift om elevers födelsedatum samt att elever i inte alltför stor omfattning avviker vad gäller ålder vid skolstart.

Resultaten visar på det stora hela att årskurseffekten ofta är större än ålderseffekten men att det finns undantag. Dessutom varierar effektskattningarna mycket mellan olika studier och mellan olika länder.²⁶ Detta gör att det känns vanskligt att hävda att det finns ett generellt mått på årskurseffekt och ålderseffekt som kan användas för att justera resultat från internationella studier i syfte att göra resultatjämförelserna mer rättvisande, vilket skulle kunna ses som ett optimalt syfte med dessa analyser. Analyserna visar dock att det inte är så okomplicerat eftersom resultaten varierar så pass mycket som de gör.

I en bilaga till ett av Skolverkets regeringsuppdrag har Gustafsson (se Gustafsson 2009) för matematik använt ett standardiserat mått på ålder respektive årskurs som är 0,17 respektive 0,24, således är skolårseffekten större än åldereffekten. I naturvetenskap använder han 0,17 och 0,28 för ålders- respektive årskurseffekt. I läsförståelse har ålders- och årskurseffekten för yngre elever (åk 3–4) satts till 0,20 respektive 0,33 och för de äldre eleverna (åk 8–9) 0 respektive 0,20. Utifrån dessa effekter korrigeras sedan resultaten för respektive land. Konsekvensen blir för Sveriges del att resultaten blir något bättre i PISA-undersökningarna som bygger på åldersurval medan resultaten i TIMSS och PIRLS blir något sämre då dessa bygger på årskursurval. Denna kvalitativa förskjutning (det vill säga korrigeringsriktning) är helt enligt förväntningarna även om den kvantitativa storleken på korrigeringen står och faller med de antaganden om storleken på ålders- respektive årskurseffekten som Gustafsson gör.

Det är viktigt att påpeka att dessa korrigeringar inte påverkar trendresultat inom tex. PISA-studierna. Detta eftersom korrigeringsfaktorn är lika stor för alla PISA-undersökningar. Däremot påverkas tolkningen av Sveriges övergripande kunskaps-trend under tiden från 1960 och 1970-talet jämfört med idag: Detta beror på att de första IEA-studierna då använde åldersbaserat urval för att sedan övergå till årskursbaserat urval. Konsekvensen blir att Sveriges resultat i dessa tidiga undersökningar var bättre än vad som framgår om man inte korrigerar resultaten, i alla fall i ett tidsrelativt perspektiv.²⁷ Denna korrektion av resultaten kan således vara befogad i detta avgränsade fall då man gör jämförelser av Sveriges relativa resultat från tidigt 1960-tal och framåt till nutid eftersom jämförelsegrunden ändrats under senare decennier.

Utifrån den stora variation som har visats från olika studier görs bedömningen att det är vanskligt att genomföra sådana här korrigeringar i mer offentliga eller löpande sammanhang, som till exempel att rutinmässigt redovisa korrigeringar i samband med publiceringen av internationella resultat. Problemet är att även om det går att skatta ålders- respektive årskurseffekt för Sverige med någorlunda god säkerhet så måste samma effekt appliceras på alla länder för att resultaten skall

²⁶ Se framförallt Luyten (2006) som analyserar 8 olika länder med TIMSS 1995-data för åk 3–4.

²⁷ Korrigeringen går åt olika håll för de tidiga IEA-studierna jämfört med för de senare IEA-studierna.

kunna jämföras. Det är mycket som talar för att effekterna inte nödvändigtvis är av samma storlek i olika länder, delvis beroende på hur eleverna allokeras till olika årskurser vid skolstart eller senare under skolgången. Men eftersom det inte är möjligt att skatta dessa effekter för varje enskilt land riskerar dessa korrigeringar att bli orättvisa. Det är därmed uppenbart att det krävs mer forskning på området innan det är möjligt att använda dessa resultat i samband med internationella studier.

Referenser

- Cahan, S., Cohen, N. 1989. *Age versus schooling effects on intelligence development*. Child Development, 60, 1239-1249.
- Cahan, S. & Davies, D. 1987. *A Between-Grades-Level Approach to the Investigation of the Absolute Effects of Schooling on Achievement*, American Educational Research Journal, 24(1), 1-12.
- Cliffordson C, 2010. *Methodological issues in investigations of the relative effects of schooling and age on school performance: the between-grade regression discontinuity design applied to Swedish TIMSS 95 data*. Educational research and Evaluation, Vol. 16, No. 1, February 2010, pp. 39-52.
- Cliffordson C, Gustafsson J-E, 2008. *Effects of age and schooling on intellectual performance: Estimates obtained from analysis of continuous variation in age and length of schooling*. Intelligence 36 (2008), pp. 143-152.
- Gustafsson J-E, 2009. *Kunskaper och färdigheter I grundskolan under 40 år: En kritisk granskning av resultat från internationella jämförande studier*. Bilaga till Skolverkets svar på regeringsuppdrag: Skolverkets bild av utvecklingen av kunskapsresultaten i grundskolan och av elevers studiemiljö – redovisning av uppdrag att utarbeta ett sammanfattande underlag avseende utvecklingen av kunskapsresultaten i grundskolan (2009) Dnr 2008:3010.
- Gustafsson J-E, 2009. *Strukturella ekvationsmodeller*. kapitel 10 i Statistisk verktygslåda 2 – multivariat analys (Red: Djurfeldt G, Barmark M). Studentlitteratur, 2009.
- IEA 2004. TIMSS 2003 Technical Report.
- Luyten H, 2006. *An empirical assessment of the absolute effect of schooling: regression-discontinuity applied to TIMSS-95*. Oxford Review of Education, Vol. 32, No. 3, July 2006, pp. 397-429.
- Luyten H., Peschar J. & Coe R., 2008. *Effects on Schooling on Reading Performance, Reading Engagement and Reading Activities of 15-Year-Olds in England*. American Educational Research Journal, Vol. 45, No. 2, pp. 319-342.
- Luyten H, Veldkamp B, 2008 *Assessing the effect of schooling with cross-sectional data: Between grades differences addressed as a selection-bias problem*. Conference paper, presenterad på IEA-konferensen i Taipei september 2008.
- Martin M, Mullis I, Foy P, 2008. *Interrelationships among Reading Achievement, Grade level and Age in PIRLS 2006*. Konferenspaper, presenterad på IEA-konferensen i Taipei september 2008.
- Skolverket, 2004. *Internationella studier under 40 år – svenska resultat och erfarenheter*. Skolverkets aktuella analyser, Beställningsnr: 04:843. Skolverket.

Appendix 1

Årskurseffekt och ålderseffekt för PIRLS 2001, Sverige: dokumenterad databearbetning

Bearbetning

1. Utifrån de två dataseten [**ASGSWER1.sav**], innehållande åk4-elever och [**ASGSE3R1.sav**], innehållande åk 3-elever, har ett samlat dataset skapats med namn: [**ASGSWER1_arbetskopia med ak3-4.sav**].

Datasetet innehåller 11 315 elever uppdelade på 5 271 elever från åk 3 och 6 044 från åk 4.

2. Compute [**elevtyp**] = **itbirthy**
3. a. select if **idgrade** =3, transform elevtyp till: Accelerated = 3 (födda 92)
Normal = 2 (91)
Delayed = 1 (89+90)
b. select if **idgrade** = 4, transform elevtyp till: Accelerated = 3 (födda 91)
Normal = 2 (90)
Delayed = 1 (89)
4. Skapar variabel [**age**] = ålder i månader, där 1 = dec 92,48 = jan 89
Genom att: compute [**age**] = [**itbirthm**] * (-1) + 13 if [**itbirthy**] = 92
+ 25 if " = 91
+ 37 if " = 90
+ 49 if " = 89
5. Skapar variabel [**grade_dummy_34**] = 1 if **idgrade** = 4, och =0 if **idgrade** = 3

Resultat

A. deskriptiv

Replikering PIRLS 2001						
Årskurs	Elevtyp	Antal elever	Summa houwgt	Andel %	Resultat PVma	Standard-avvikelse
3	Accelerated	181	139	2,6	527,3	71,3
	Normal	4 987	5 040	95,6	520,2	76,9
	Delayed	103	92	1,7	482,0	88,0
4	Accelerated	170	117	1,9	576,0	63,7
	Normal	5 778	5 830	96,5	561,9	64,9
	Delayed	170	94	1,6	491,6	79,8

B. Regressioner

Utförda i IDB Analyzer med beroende variabel: [ASREA01-05] och förklarande variabler [age] och [grade_dummy_34]

$$\text{Modell: PVread} = 495,8 + 1,30 * (\text{age}_{\text{månad}}) + 26,31 * (\text{grade})$$

(t=5,8) (t=7,2)

vilket motsvarar: **15,6 * (age_{år})**

Enligt denna skattning är årskurseffekten (26,3) betydligt större än ålderseffekten (15,6). Resultaten är identiska med Gustafssons resultat (se Djurfeldt och Barmark, 2009).

Appendix 2

Replikering av Cliffordson "Effects of schooling and age on performance in mathematics and science: A between-grade regression discontinuity design applied to Swedish TIMSS 1995 data" (presenterad på IEA-konferensen i Taipei 2008).

I denna del av appendix replikeras en studie av Cliffordson där en "between grade regression discontinuity design" används.

Data

TIMSS 1995 data för Sverige

Original data set: [bsgswem1.sav] (elevfilen med bakgrundsvariabler och plausibla värden i ma och sci). Filen innehåller 8 855 elever tillhörande årskurs 6 (2 831 elever), 7 (4 075 elever) och 8 (1 949 elever), vilket stämmer med Cliffordson.

En första deskriptiv bild

STUDENT'S DATE OF BIRTH\YEAR

Tabell 1. Andel elever per födelseår

	Födelseår	Antal	Procent	Valid procent	Kumulativ procent
Valid	78	1	,0	,0	,0
	79	60	,7	,7	,7
	80	2 001	22,6	22,7	23,4
	81	3 995	45,1	45,3	68,7
	82	2 753	31,1	31,2	99,9
	83	11	,1	,1	100,0
	85	2	,0	,0	100,0
	Total		8 823	99,6	100,0
Bortfall	System	32	,4		
Total		8 855	100,0		

Variabel: itbirthy, oviktat

Som kan ses i tabell 1 saknas uppgift för 32 elever. Cliffordson uppger att den äldsta eleven är född i jan 1979 och den yngsta i dec 1983. I detta dataset finns ytterligare 3 elever (födda 1978 och 1985). Dessa tre elever har exkluderats i analyserna för att följa Cliffordson. Då återstår 8 820 elever.

I nästa steg har de 8 820 eleverna tilldelats ett värde mellan 1 och 60 som identifierar ålder (variabel **age**) i månader där 1 = dec 83 och 60 = jan 79 (på samma sätt som i Cliff). Variablerna **itbirthy** (födelseår) och **itbirthm** (födelsemånad) har använts för detta ändamål.

Sedan har utifrån variabeln **idgrade** (årskurs) i kombination med **itbirthy** (födelseår) elever klassificerats som ”accelerated”, ”normal” samt ”delayed” utifrån om de tillhör den normala årskullen för den årskursen eller om de har startat senare eller tidigare. Den variabel som skapats är **elevtyp** (där 1=accelerated, 2=normal, 3=delayed). Alla dessa databehandlingar har utförts med SPSS 15.0.0.

Genomsnittliga resultat i matematik

I tabell 2 presenteras genomsnittliga resultat efter årskurs samt elevtyp. Tabellen skall jämföras med table 1 i Cliffordson.

Tabell 2. Deskriptiv bild av elevmaterialet med genomsnittliga resultat i matematik efter årskurs och elevtyp

Årskurs	Elevtyp	Antal elever	Replikering		Resultat PVma	S.D.	Cliffordson	
			Summa houwgt	Andel %			Resultat PVma	S.D.
6	Accelerated	10	10	0,3	510,85	64,54	507,52	59,04
	Normal	2 724	2 783	96,7	479,56	75,97	479,12	75,82
	Delayed	82	85	3,0	421,01	76,86	422,38	80,97
7	Accelerated	30	21	0,7	560,30	70,06	553,87	66,50
	Normal	3 901	2 816	96,1	520,90	85,03	520,84	84,72
	Delayed	130	94	3,2	453,32	77,10	446,79	75,94
8	Accelerated	13	20	0,7	643,74	67,54	647,24	69,39
	Normal	1 872	2 898	96,4	555,52	89,99	555,86	89,25
	Delayed	58	89	3,0	479,33	93,85	475,10	103,44

Programvara: IDB analyzer, vikt: houseweight (houwgt), Plausibla värden från orginalskala för 1995.

Eftersom det finns så pass få elever som tillhör gruppen ”accelerated” eller ”delayed” så finns ingen större anledning att bekymra sig alltför mycket över avvikelser med Cliffordson för dessa grupper då enstaka individer kan påverka resultatet ganska mycket och då antalet elever (uttryckta i summa houwgt) inte heller riktigt stämmer för dessa grupper. Däremot verkar antalet normalelever i varje årskurs stämma exakt med Cliffordson (kan ej ses i denna tabell, måste jämföras med tabell 1 i Cliff), vilket får tas som ett bevis för att vi faktiskt använder exakt samma elever i datasetet. Det genomsnittliga resultatet i matematik för normalgruppen inom respektive årskurs stämmer också mer eller mindre exakt med Cliffordson även om det skiljer sig på decimalnivå.

Replikering av regressioner i matematik

En regressionsmodell har skattats baserat på normaleleverna i årskurs 6 och 7 och en modell baserat på normaleleverna i årskurs 7 och 8.²⁸

Regressionsmodell: $PV_{ma} = b_0 + b_1 * (AGE) + b_2 * (GRADE) + e$

Där **AGE** är ålder i månader och **GRADE** är årskurs i år.

Tabell 3. Skattning av ålders-, resp. årskurseffekt för åk 6–7 och 7–8 i matematik. Alla regressioner är baserade på "normalelever".

	Årskurs	Födelse- månad	N*	Koefficient b ₁			Koefficient b ₂				
				Konstant	AGE (månad)	s.e.	t-värde	AGE (år)**	GRADE (år)	s.e.	t-värde
Rep	6–7	jan–dec	6625	458,96	1,09	0,32	3,4	13,08	28,35	5,81	4,9
Cliff	6–7	Jan–dec	5599	Ej uppg	1,12	0,32	3,6	13,44	28,43	5,28	5,4
Rep	7–8	jan–dec	5773	477,94	1,40	0,37	3,8	16,8	17,51	6,19	2,8
Cliff	7–8	Jan–dec	5714	Ej uppg	1,34	0,39	3,5	16,07	18,67	6,99	2,7

*Cliffordson uppger N som det viktade antalet elever medan replikeringen syftar på det verkliga antalet eleverna.

**Denna koefficient är endast AGE (månad) multiplicerad med 12 för jämförbarhet med GRADE.

Genomsnittliga resultat i naturvetenskap

Tabell 4. Skattning av ålders-, resp. årskurseffekt för åk 6–7 och 7–8 i naturvetenskap. Alla regressioner är baserade på "normalelever". Obs. Här finns ingen tabell att jämföra med Cliffordson som endast uppger mönstret i resultaten i texten för naturvetenskap.

Replikering						
Årskurs	Elevtyp	Antal elever	Summa houwgt	Andel %	Resultat PVsci	S.D.
6	Accelerated	10	10	0,3	489,1	49,6
	Normal	2 724	2 783	96,7	490,2	84,8
	Delayed	82	85	3,0	427,4	91,5
7	Accelerated	30	21	0,7	574,0	69,8
	Normal	3 901	2 816	96,1	537,0	88,9
	Delayed	130	94	3,2	476,4	95,6
8	Accelerated	13	20	0,7	602,4	81,4
	Normal	1 872	2 898	96,4	572,1	92,9
	Delayed	58	89	3,0	489,3	96,1

²⁸ Regressionerna utfördes i IDB Analyzer efter att grupperingsvariablerna normal_67 respektive normal_78 skapades med värdet 1 om elev är dels normalelev och dels tillhör årskurs 6 eller 7 (respektive 7 eller 8), annars bortfall.

Replikering av regressioner i naturvetenskap

Regressionsmodell: $PVsci = b_0 + b_1 * (AGE) + b_2 * (GRADE) + e$

Tabell 5. Skattning av ålders-, resp. årskurseffekt för åk 6–7 och 7–8 i naturvetenskap. Alla regressioner är baserade på "normalelever".

	Årskurs	Födelse- månad	N*	Koefficient				Koefficient			
				Konstant	AGE (månad)	s.e.	t-värde	AGE (år)**	GRADE (år)	s.e.	t-värde
Rep	6–7	jan–dec	6625	473,1	0,96	0,32	3,0	11,5	34,4	5,91	5,8
Cliff	6–7	Jan–dec	5599	Ej uppg							
Rep	7–8	jan–dec	5773	483,2	1,76	0,47	3,8	21,1	13,6	6,79	2,0
Cliff	7–8	Jan–dec	5714	Ej uppg							

Här finns inga exakta siffror för Cliffordson men hon uppger att årskurseffekten var större i sci för åk 6–7 regressionen men mindre än ma för 7–8 regressionen, resultat som verifieras i tabell 5. Det är intressant att notera att årskurseffekten är hela tre gånger större än ålderseffekten för åk 6–7 medan årskurseffekten är mindre än ålderseffekten för åk 7–8. För ma var årskurseffekten dubbelt så stor som ålderseffekten för åk 6–7 och för åk 7–8 var motsvarande effekter ungefär lika stora.

En tänkbar förklaring till mönstret att årskurseffekten är större mellan åk 6 och åk 7 är att det sker en större förändring akademiskt när elever tar steget från åk 6 till åk 7 och ofta byter eleverna samtidigt skola.

Motsvarande analyser med omskalad data

När resultaten från TIMSS 1995 släpptes användes en enparametermodell för att skala resultaten i matematik och naturvetenskap. I senare internationella studier har man gått över till en treparametermodell. För att resultaten skall vara jämförbara har då resultaten från TIMSS 1995 skalats om med en treparametermodell.

I nedanstående tabeller har samma analyser och regressioner gjorts som i tidigare avsnitt men nu med de nya omskalade resultaten i matematik och naturvetenskap. Obs att allt är lika förutom de plausibla värdena i matematik respektive naturvetenskap.

Tabell 6. Deskriptiv bild av elevmaterialet med genomsnittliga resultat i matematik efter årskurs och elevtyp. OBS! Nyskalade plausibla värden (ej jämförbara med Cliffordson).

Replikering						
Årskurs	Elevtyp	Antal elever	Summa houwgt	Andel %	Resultat PVma	S.D.
6	Accelerated	10	10	0,3	502,8	54,8
	Normal	2 724	2 783	96,7	477,7	69,4
	Delayed	82	85	3,0	417,0	75,6
7	Accelerated	30	21	0,7	552,4	56,8
	Normal	3 901	2 816	96,1	515,2	69,0
	Delayed	130	94	3,2	451,7	68,5
8	Accelerated	13	20	0,7	605,4	49,1
	Normal	1 872	2 898	96,4	541,9	74,9
	Delayed	58	89	3,0	468,1	88,9

Programvara: IDB analyzer, vikt: houseweight (houwgt), Plausibla värden nyskalade för 1995.

Tabell 7. Skattning av ålders-, resp. årskurseffekt för åk 6–7 och 7–8 i matematik. Alla regressioner är baserade på "normalelever". OBS. Nyskalade plausibla värden. Ej jämförbara med Cliffordson.

	Årskurs	Födelse- månad	N*	Koefficient				Koefficient			
				Konstant	AGE (månad)	s.e.	t-värde	AGE (år)**	GRADE (år)	s.e.	t-värde
Rep	6–7	jan–dec	6625	456,2	1,14	0,27	4,3	13,7	24,0	5,45	4,4
Rep	7–8	jan–dec	5773	469,3	1,50	0,32	4,7	18,0	8,4	5,88	1,43

Tabell 8. Deskriptiv bild av elevmaterialet med genomsnittliga resultat i naturvetenskap efter årskurs och elevtyp. OBS! Nyskalade plausibla värden (ej jämförbara med Cliffordson).

Årskurs	Elevtyp	Replikering				
		Antal elever	Summa houwgt	Andel %	Resultat PVma	S.D.
6	Accelerated	10	10	0,3	487,0	59,0
	Normal	2724	2783	96,7	485,4	74,6
	Delayed	82	85	3,0	413,0	83,9
7	Accelerated	30	21	0,7	563,1	58,4
	Normal	3901	2816	96,1	524,9	74,6
	Delayed	130	94	3,2	459,1	84,4
8	Accelerated	13	20	0,7	594,0	63,3
	Normal	1872	2898	96,4	555,0	80,3
	Delayed	58	89	3,0	473,8	89,1

Programvara: IDB analyser, vikt: houseweight (houwgt), Plausibla värden nyskalade för 1995.

Tabell 9. Skattning av ålders-, resp. årskurseffekt för åk 6–7 och 7–8 i naturvetenskap. Alla regressioner är baserade på "normalelever". OBS. Nyskalade plausibla värden. Ej jämförbara med Cliffordson.

	Årskurs	Födelse- månad	N*	Koefficient				Koefficient			
				Konstant	AGE (månad)	s.e.	t-värde	AGE (år)**	GRADE (år)	s.e.	t-värde
Rep	6–7	jan–dec	6625	465,2	1,08	0,34	3,2	13,0	26,7	6,48	4,1
Rep	7–8	jan–dec	5773	471,6	1,74	0,37	4,7	20,9	8,9	5,78	1,5

Tabell 10. Jämförelse av resultat för originalsкала respektive nyskala

Årskurser			Ålderseffekt (AGE år)	t-värde	Årskurseffekt (Grade år)	t-värde	
Åk 6-7	PVma	Original	13,1	3,4	28,4	4,9	
		Ny	13,7	4,3	24,0	4,4	
	PVsci	Original	11,5	3,0	34,4	5,8	
		Ny	13,0	3,2	26,7	4,1	
	Åk 7-8	PVma	Original	16,8	3,8	17,5	2,8
			Ny	18,0	4,7	8,4	1,4
PVsci		Original	21,1	3,8	13,6	2,0	
		Ny	20,9	3,2	8,9	1,5	

Kommentar: För åk 6-7 är det inga större skillnader mellan originalsкалаn och nysкалаn även om årskurseffekten är något svagare med nyskalad data. Vad gäller åk 7-8 är däremot skillnaden i årskurseffekt betydande, både för ma och sci. Årskurseffekten är endast cirka hälften så stor med nyskalad data och inte signifikant.

Skolverket

www.skolverket.se