

4

Fødselsmåned og skoleprestasjoner

ROLF VEGAR OLSEN OG JULIUS KRISTJAN BJØRNSSON

SAMMENDRAG I Norge går alle barn født det samme året på samme klasstrinn, med veldig få unntak. Dette betyr at forskjell i alder kan være bortimot tolv måneder på det samme trinnet. I idrett og utdanning viser eksisterende litteratur at generelt presterer de eldste på trinnet vesentlig bedre enn de yngste. Hensikten med denne studien er å bruke norske data fra TIMSS og PISA gjennom 20 år for å se om denne effekten (relativ alderseffekt – RAE) har endret seg over tid, og hvordan den varierer etter klasstrinn og alder. Studien bruker en såkalt regresjons-diskontinuitetsanalyse for å se på effekten over to klasstrinn, og resultatene indikerer at effekten er omtrent 60 % av effekten av et helt skoleår. RAE er to ganger større blant de yngste elevene (4. og 5. trinn) enn de eldste (8. og 9. trinn), men analysene viser at forholdet mellom effekten av et år på skolen og RAE er den samme i begge aldersgruppene. Resultatene er diskutert, og diverse implikasjoner av dem er omtalt.

SUMMARY In Norway all children born in the same year, attend the same grade level, with very few exceptions. This means that up to a 12 month difference in age can exist in the same grade. The existing literature both on sports and educational achievements shows that the older pupils within a grade, perform considerably better than the younger. The purpose of this study is to utilise Norwegian data from TIMSS and PISA from the last 20 years in order to study whether this relative age effect (RAE) is robust over time, and how it varies across grade and age. The study applies a regression discontinuity analysis to look at the effect across two adjacent grades, 4th and 5th and 8th and 9th. The results indicate that the RAE is approximately 60 % of the effect of a whole year in school. The RAE is twice as large among the youngest students (4th and 5th) compared to the older (8th and 9th), but the analysis shows that proportionally the size of the effect of a whole grade and the RAE is the same in both cases. These results are further discussed and some implications are underlined.

INNLEDNING

I dette kapitlet bruker vi data fra 20 år med TIMSS- og PISA-undersøkelser for å undersøke om barnets alder, gitt at andre variabler er like, er relatert til prestasjonene på testene. I alle land er skolestart sterkt knyttet til fødselsdato, og som regel er det slik at det finnes en bestemt dato som fungerer som et mer eller mindre skarpt alderskriterium: Barn født på eller inntil ett år før denne datoen begynner på skolen det samme året, mens barn som er født etter denne datoen, begynner det neste skoleåret. Dette gir et i utgangspunktet tydelig, men likevel vilkårlig skille hvor: (i) barn som er født på hver sin side av skillelinjen, går i ulike klassetrinn selv om de er nesten like gamle, og (ii) hvor barn i samme klasse har en variasjon i alder på ett år. I Norge er klassetrinn og alder omtrent perfekt sammenfallende, og repetisjon av klasse og forsinket eller akselerert skolestart eksisterer nesten ikke. Gjennom datasettene for de internasjonale undersøkelsene kan vi derfor estimere hvor sterkt relatert barnets alder er til prestasjonene på de faglige prøvene. Denne effekten kalles ofte den relative alderseffekten (RAE) i samme klassetrinn. I tillegg har TIMSS enkelte år hatt utvalg fra to tilgrensende klassetrinn, noe som gir gode muligheter for også å kunne separere effektene av henholdsvis alder og antall år på skolen, noe som utnyttes i denne studien.

Det finnes allerede en rik forskning på RAE-fenomenet, også gjennom analyser av data fra de internasjonale undersøkelsene. Imidlertid gir vi et mer komplett bilde gjennom analyser av svært mange datasett på tvers av studier, alderskohorter, fagområder og undersøkelsessyklus. Til sammen bidrar vi dermed med en analyse av hvor robust RAE er i slike datasett. I utgangspunktet har undersøkelser som TIMSS og PISA et potensiale for å studere effekter av ulik politikk på et område. For det første gir den internasjonale komparative konteksten en variasjon i variabler som innenfor ett enkelt land eller utdanningssystem er en konstant. For det andre har man gjennom tidsseriene på 20 år også muligheter for å analysere situasjonen innen et land før og etter en større endring i regelverk eller politikk. I vårt eget land hadde vi eksempelvis innføring av obligatorisk skolestart for 6-åringene i 1997/98. Både i TIMSS og PISA finnes det alderskohorter både før og etter innføringen av denne reformen. Imidlertid er RAE et komplekst fenomen, og vi identifiserer derfor også noen klare begrensninger i studiedesignet ved de internasjonale undersøkelsene, og foreslår hvilke endringer i design som kan bidra til økte muligheter for å studere dette og en rekke andre svært utdanningspolitisk relevante spørsmål.

RELATIV ALDERSEFFEKT (RAE)

Det er godt kjent fra tidligere forskning at relativt eldre elever presterer bedre på en rekke utfallsmål, gitt at alt annet er likt. Dette fenomenet er spesielt godt studert for ulike fysiske utfallsmål og ikke minst for sportslige idrettsprestasjoner. (For en oversikt se Cobley, Baker, Wattie, & McKenna, 2009.) I en norsk kontekst (i Nord-Trøndelag) har Aune mfl. (2017) vist at elever født i første halvdel av alderskohorten har betydelig høyere gjennomsnittskarakterer i kroppsøving på ungdomstrinnet, og denne tendensen var tydeligere for jenter på alderstrinnet enn for gutter. I tillegg er det også godt dokumentert at det innenfor aldersbaserte skolekohorter er en overvekt av ADHD blant de yngste elevene (Chen et al., 2016; Morrow et al., 2012), spesielt i land hvor forekomsten av diagnosen er høy. Forsknin-gen på innflytelsen av RAE på generell mental helse viser imidlertid sprikende resultater (Lien, Tambs, Oppedal, Heyerdahl, & Bjertnes, 2005).

Det finnes også en rekke studier som viser en tydelig RAE for skoleprestasjoner, og flere av disse har benyttet data fra de internasjonale studiene. Et flertall av studiene har imidlertid hatt som hovedhensikt å estimere effekten av det å gå på skole (se f.eks. Luyten, 2006). For å kunne beregne effekten av ett års skolegang (heretter referert til som trinneffekten) må man også korrigere for elevenes alder ved testtidspunktet og/eller alder ved skolestart. Det er derfor en fordel å ha datasett som inkluderer flere klassetrinn og alderskohorter. I slike datasett har man flere metodiske muligheter knyttet til det å utnytte den såkalte diskontinuiteten som oppstår i et multikohortutvalg (Cliffordson, 2010; Luyten & Veldkamp, 2011). Datasettene fra TIMSS 1995 er derfor spesielt mye brukt fordi man den gangen nettopp inkluderte to tilgrensende klassetrinn for hver av de to utvalgene i TIMSS 1995 (2. og 3. klasse og 6. og 7. klasse) i en rekke land. Generelt viser analyser av dette datasettet, og andre multikohortdata, at forskjeller i skoleprestasjoner mellom elever på to (eller flere) klassetrinn er knyttet både til en effekt av tid på skolen (heretter referert til som trinneffekten) og elevenes alder (Luyten, Merrell, & Tymms, 2017).

I en analyse av TIMSS 1995 hvor Sverige deltok med elever fra tre klassetrinn (6., 7. og 8. trinn), viser Cliffordson (2010) at RAE er omtrent en tredel av trinneffekten fra 6. til 7. trinn. I overgangen 7. til 8. trinn er RAE omtrent uendret, men her er RAE og trinneffekten like store. I Norge var effekten av alder i matematikk på 2./3. trinn²⁵ i TIMSS 1995 omtrent en tredel av effekten av ett år på skolen (Luyten & Veldkamp, 2011). Datasettet for PIRLS 2006 inkluderte også to klas-

25. Norge deltok med 2. og 3. trinn i denne yngre populasjonen i TIMSS 1995, men senere med 4. trinn og i de siste rundene av TIMSS med 4. og 5. trinn

setrinn i det norske utvalget, og i tilsvarende diskontinuitetsanalyser av dette datasettet er RAE omtrent dobbelt så stor som trinneffekten (Martin, Mullis, & Foy, 2011). Gabrielsen og Lundetræ (2017) har for øvrig presentert RAE for alle PIRLS-studiene i Norge, men ikke gjennom en kombinert analyse av både 4. og 5. trinn. De viser at elever på 4. trinn født i januar/februar presterer betydelig bedre på PIRLS-leseprøven enn medelever født i november/desember (typisk 30 poeng eller litt mindre). De viser også at effekten er rimelig stabil i perioden 2001–2016, og at effekten er ganske lik for gutter og jenter. Oppsummert viser altså forskningen på RAE knyttet til faglige prestasjoner at RAE er et merkbart fenomen, men at effekten varierer på tvers av land, aldersgrupper, studier og fagområder.

Det er rimelig å anta at skolestart vil fortsette å være definert av et relativt skarpt alderskriterium. Det er derfor viktig å få kunnskap om hvordan RAE utvikler seg oppover i skoleløpet. Foreløpig finnes det kun noen få studier som dokumenterer utviklingen i RAE etter hvert som elevene blir eldre. Gjennom å bruke data fra tre ulike datasett fra de internasjonale studiene finner Ponzio & Scoppa (2014) at RAE avtar for italienske elever med økende alder/klassestrinn. Luyten mfl. (2017) har tilsvarende funn i et interessant datasett som inkluderer de første seks årene i britiske skoler, og de viser samtidig at også trinneffekten reduseres tilsvarende oppover i skoleløpet. Det er imidlertid verdt å legge vekt på at det fortsatt er en tydelig RAE også for elever på slutten av skoleløpet i de fleste land (Sprietsma, 2010). I Norge har Strøm (2004) bekreftet dette i en analyse av PISA 2000-data. Andre studier har også vist at RAE kan vedvare lenger opp i utdanningssystemet ved at relativt eldre elever har noe større sannsynlighet for å begynne på en høyere utdanning (Bedard & Dhuey, 2006; Kawaguchi, 2006). I en norsk kontekst har Solli (2017) brukt registerdata og finner markerte relative aldersforskjeller i skoleprestasjon. Disse effektene er upåvirket av kjønn og sosial bakgrunn. Videre finner hun også langtidseffekter i form av at de som er født relativt tidligst, i større grad går rett fra videregående skole inn i høyere utdanning, og at de også har relativt høyere inntekt ved trettiårsalder. Denne innteksteffekten er noe sterkere for gutter. Black, Devereux og Salvanes (2011) finner det samme i sine analyser av registerdata. I tillegg finner de at relativt eldre gutter har høyere skårer på IQ-tester som gjennomføres i forbindelse med sesjoner til Forsvaret, dvs. når de er omtrent 18 år gamle. Til sammen indikerer dette at relativt yngre elever har et noe svakere utgangspunkt for videre studievalg og karriere i de kritiske overgangene mellom ungdomstrinn og videregående skole og mellom videregående skole og høyere utdanning. I den samme studien viser imidlertid Black mfl. (2011) at det ikke synes å være langtidseffekter gjennom livsløpet (inntekt som avhengig variabel).

Dette er for øvrig i tråd med en tilsvarende studie fra en japansk kontekst (Kawaguchi, 2006).

I denne artikkelen ønsker vi å supplere denne forskningen ved å dokumentere den relative alderseffekten for alle undersøkelsesår i TIMSS og PISA i Norge fordi estimatene av effektstørrelsene er sprikende. Gjennom å undersøke et stort antall datasett fra ett land kan vi se på hvor robust denne effekten er på tvers av studier, fagområder, alderskohorter og over tid. Det siste er spesielt interessant i en norsk kontekst fordi alderskriteriet for skolestart ble endret fra sju til seks år i 1997, noe som gjør at det finnes datasett før og etter at endringen ble implementert. I tillegg ønsker vi å vise hvordan en regresjonsdiskontinuitetsanalyse gir muligheter for å skille mellom en ren alderseffekt og effekten av tidslengde (antall år) på skolen.

METODE OG DATA

I dette kapitlet analyseres norske data fra samtlige TIMSS- og PISA-undersøkelser som er gjennomført så langt. Til sammen dekker disse to studiene tre aldersgrupper/klassestrinn, 4., 8. og 10. trinn. Enkelte år har Norge deltatt også med hele eller reduserte utvalg for henholdsvis 5. og 9. trinn (TIMSS). Det er i utgangspunktet en vesentlig forskjell mellom TIMSS og PISA ved at TIMSS har en utvalgsramme basert på klassestrinn, mens PISA har utvalg som representerer en kohort med elever som er født i det samme kalenderåret. I analyser av norske (og nordiske) data er denne forskjellen av minimal betydning siden klassestrinn og årskohorter er nesten perfekt sammenfallende fordi skilledatoen er 1. januar. Data fra Norge er derfor spesielt egnet til å studere RAE fordi elevene har hatt like lang tid i skolen da de deltok i studiene. Eventuelle forskjeller i prestasjoner knyttet til fødselsmåned og/eller trinn vil derfor ikke være påvirket av at elever har hatt ulik lengde på skolegangen. Imidlertid vil en internasjonalt sammenliknende analyse av RAE i de internasjonale undersøkelsene være komplisert fordi alderssammensetninger på ulike klassestrinn varierer mye mellom land. For mer spesifikke detaljer om utvalgene i disse årene viser vi til de nasjonale og internasjonale rapportene for hver av studiene.

For alle datasettene finnes elevens fødselsmåned og år, og disse variablene har i datasettene blitt omformet til en variabel som representerer *alder ved testtidspunktet*.²⁶ I Norge er som sagt alderskriteriet for skolestart gjennomført veldig strengt, og det er kun få avvik fra regelen om at fødselsårskohorter går på samme klassestrinn. For analysene som presenteres her, er de 1–2 % av elevene som er

over- eller underårig, utelatt; noe som vil ha minimal eller ingen påvirkning på resultater og konklusjoner knyttet til RAE i dette kapitlet.

For hver av studiene har en enkel lineær regresjonsanalyse av elevens alder på skår fra de ulike faglige områdene blitt gjennomført. I tillegg til slike enkle lineære regresjoner presenterer vi resultatene av en analyse hvor data fra to naboliggende klassetrinn for TIMSS-studien i 2015 er inkludert. Denne analysen er en såkalt regresjonsdiskontinuitetsanalyse (RDA). I disse analysene estimeres den samlede effekten av forskjellen mellom to alderskohorter i form av to delkomponenter, én knyttet til elevens alder og en annen knyttet til antall år på skolen. Detaljene for denne analysen presenteres sammen med resultatene.

Alle analysene er foretatt ved hjelp av programvaren IDB-Analyzer (IEA, 2017). Denne programvaren tar hensyn til at elevene representerer et klyngeutvalg, og at ikke alle elever og elevklynger er trukket med samme sannsynlighet. I tillegg ivaretar dette programmet at elevenes skårer i de internasjonale undersøkelsene er uttrykt ved hjelp av såkalte plausible verdier (von Davier, 2014). Ved hjelp av dette beregner IDB Analyzer robuste standardfeil som tar hensyn til kompleksiteter knyttet til både utvalgs- og prøvedesign i de internasjonale studiene.

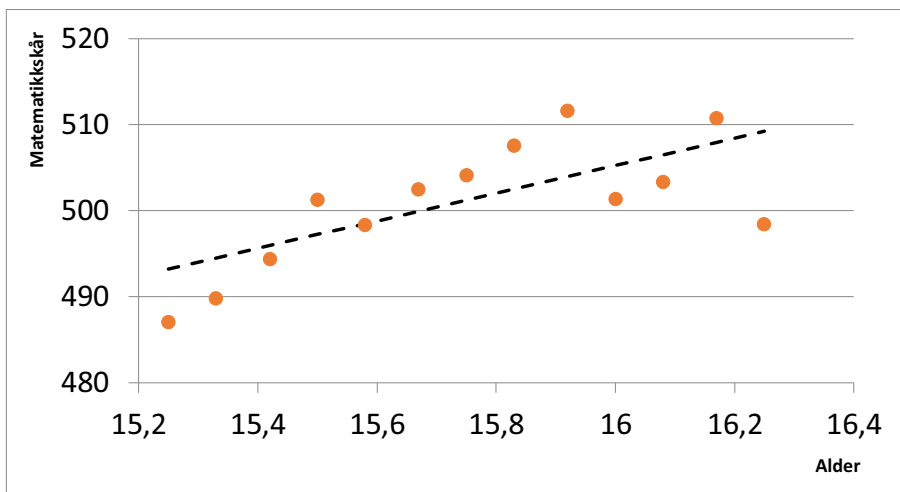
Regresjonsanalysene gir til slutt en verdi som representerer forskjellen i prestasjoner knyttet til en aldersforskjell på ett år. Hvor stor en slik effekt skal være før man kan kalle den substansielt meningsfull, er ikke helt lett å anslå. For det første viser standardfeilene fra de mange analysene at en RAE på 10 poeng eller mer gjennom året er statistisk signifikant. For det andre må man ha et grunnlag for å anslå hvor stor en meningsfull effekt er. Et utgangspunkt kan være å sammenlikne med typisk estimerte effekter av ett skoleår. På tvers av flere land i TIMSS 1995 er effekten av ett skoleår på 10–30 poeng for 4. trinn (Luyten & Veldkamp, 2011). Denne effekten knyttet til ett års skolegang er med stor sannsynlighet fallende oppover i skoleløpet (Luyten et al., 2017). Vi vil derfor mene at alle forskjeller som er statistisk signifikante, også er substansielt meningsfulle fordi de potensielt representerer effekten av et halvt til ett helt års skolegang.

26. Siden undersøkelsene er gjennomført over en testperiode på om lag en måned, medfører dette at elever født i samme måned har en liten variasjon i alder og lengde på skolegang ved testtidspunktet.

RESULTATER

OPPSUMMERING AV DE ENKLE LINEÆRE REGRESJONENE

Figur 4.1 illustrerer analysene som er gjort i et av datasettene. Hvert av punktene representerer gjennomsnittlig skår på en faglig prøve for elever med samme alder ved testtidspunktet. Den stiplede linjen visualiserer resultatet fra en ordinær lineær regresjon. Tendensen er tydelig: Jo eldre elevene er, jo bedre presterer de på testen. I dette tilfellet, med data fra matematikktesten i PISA 2015, representerer forskjellen mellom eldste og yngste elev omtrent 16 poeng på skalaen. Tilsvarende analyser er gjennomført for alle fagområder i alle TIMSS- og PISA-undersøkelsene, og i tillegg er analysene gjort uavhengig for jenter og gutter.



FIGUR 4.1 Sammenhengen mellom alder²⁷ ved testtidspunktet og skår i matematikk i PISA 2015 for norske elever

Tabell 4.1 viser regresjonskoeffisientene for alle de 38 uavhengige regresjonsanalysene. Oppsummert viser denne at estimatene for RAE gjennomgående er substansielt meningsfulle og statistisk signifikante. For PISA-undersøkelsene, altså 10. trinn, er bildet noe mindre tydelig, med gjennomgående lavere verdier. Av 18 analyser på PISA-data er 11 estimater statistisk signifikante. Tabell 4.1 viser også at estimatene innenfor samme trinn fluktuerer mye fra én syklus av studien til den neste. Det er derfor ingen merkbar trend i utviklingen i RAE over tid for noen av årstrinnene. Estimatene for 8. trinn er noe mer stabile over tid enn for de to andre

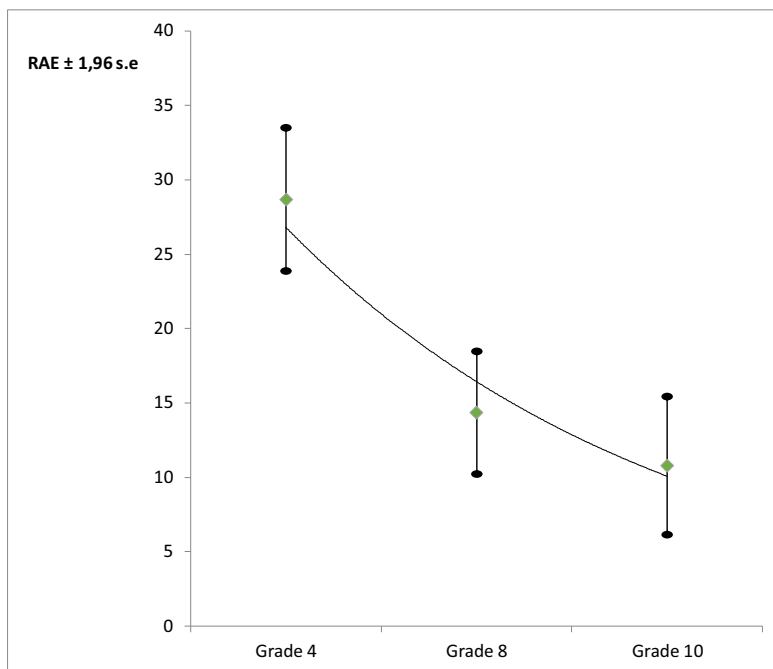
27. Alder i figur 4.1 er regnet med desimaler, dvs. en elev som er 15 år og seks måneder gammel, vises som 15,5.

trinnene. Det er heller ingen endring av RAE for kohorter før og etter innføringen av skolestart i Norge for 6-åringer. På tvers av fag innenfor samme studiesyklus er estimatene for RAE relativt like.

TABELL 4.1 Regresjonskoeffisienter fra alle analysene. Statistisk signifikante effekter er uthevet

TIMSS	1995	2003	2007	2011	2015	
4. trinn matematikk	20,6 (6,1)	27,6 (4,6)	43,8 (5,8)	19,4 (5,3)	31,9 (5,6)	
4. trinn naturfag	26,9 (8,0)	26,4 (6,6)	44,0 (6,4)	25,2 (5,0)	29,6 (5,5)	
8. trinn matematikk	14,6 (5,2)	12,4 (4,8)	9,0 (4,8)	14,6 (4,6)	13,7 (4,0)	
8. trinn naturfag	22,6 (5,3)	11,4 (5,2)	11,6 (4,3)	13,1 (5,5)	10,1 (6,5)	
PISA	2000	2003	2006	2009	2012	2015
10. trinn matematikk	21,0 (8,6)	2,0 (5,2)	13,4 (5,4)	8,3 (4,3)	6,2 (5,0)	16,9 (5,2)
10. trinn naturfag	29,8 (8,3)	1,0 (6,6)	12,8 (5,5)	19,8 (4,6)	5,4 (5,6)	20,8 (4,7)
10. trinn lesing	22,7 (6,2)	12,0 (6,0)	20,9 (6,1)	7,8 (4,4)	12,6 (6,0)	20,2 (6,6)

Gitt at det ikke synes å være noen tendens i hvordan RAE endrer seg fra ett undersøkelsesår til det neste, antar vi videre at disse fluktuasjonene reflekterer tilfeldige utvalgsfeil rundt et gjennomsnitt som er stabilt over tid. Med en slik antakelse kan vi uttrykke gjennomsnittlig RAE for hvert trinn over alle undersøkelsesår (med tilhørende standardfeil). Figur 4.2 viser gjennomsnittlig RAE for hvert trinn med tilhørende 95 % konfidensintervall. Figuren viser en tydelig tendens til at RAE faller fra omtrent 30 poeng for 4. trinn ned til 10–15 poeng for 8. og 10. trinn. Figuren oppsummerer kun estimatene for matematikk. RAE er som sagt veldig stabil over fagområder, og tilsvarende figur for naturfag ville vært omtrent identisk.



FIGUR 4.2 Oppsummering av RAE (matematikk) for de tre inkluderte årstrinnene. Intervallene angir et 95 % konfidensintervall. Se tekst for ytterligere forklaring av figuren.

Uavhengige regresjonsanalyser er også gjennomført for gutter og jenter for alle undersøkelsesår, trinn og fagområder. For 4. og 8. trinn i TIMSS-undersøkelsene er kjønnsforskjellene gjennomgående små (med noen få unntak) og i begge retninger. For 10. trinn i PISA-undersøkelsene er kjønnsforskjellene også moderate, men gjennomgående i retning av at RAE er høyere for gutter. Dette er oppsummert i tabell 4.2. Kjønnsforskjellene over fagområder er stabile, og for 10. trinn er gjennomsnittlig kjønnsforskjell i RAE på om lag 7–8 poeng.

TABELL 4.2 Oppsummering av kjønnsforskjeller i RAE. Positive verdier angir at RAE er høyere for gutter. Signifikante kjønnsforskjeller er uthevet, og standardfeil er oppgitt i parentes.

	Matematikk		Naturfag		Lesing	
4. trinn	1,5	(5,0)	-3,3	(7,6)		
8. trinn	-1,6	(1,9)	0,1	(1,9)		
10. trinn	7,5	(1,9)	8,0	(3,7)	7,0	(3,6)

Vi har i tillegg gjort regresjonsanalyser for alle datasett hvor elevenes sosiale bakgrunn også er inkludert. Resultater fra disse analysene rapporteres ikke her, men gjennomgående er estimatet av RAE uendret når SES blir inkludert. RAE er med andre ord upåvirket av elevenes hjemmebakgrunn.

REGRESJONSDISKONTINUITETSANALYSENE

For noen av undersøkelsesårene har TIMSS-studien blitt gjennomført med to årstrinn for hver av de to populasjonene. For noen av de norske datasettene har man derfor mulighet til å gjennomføre en regresjonsdiskontinuitetsanalyse (RD-analyse). Prinsippet for slike analyser er å bruke den muligheten som et plutselig sprang i en av variablene representerer: De yngste elevene på 5. trinn er kun én dag eldre enn de eldste elevene på 4. trinn. Det er med andre ord en fordeling i alder som er kontinuerlig og jevn over to kalenderår, men midt i denne fordelingen er det et punkt på tidsaksen hvor elevene som er bare litt eldre, har gått på skolen ett år lenger.

Denne situasjonen kan vi modellere ved følgende regresjonslikning i datasett som består av henholdsvis 4./5. trinn og 8./9. trinn:

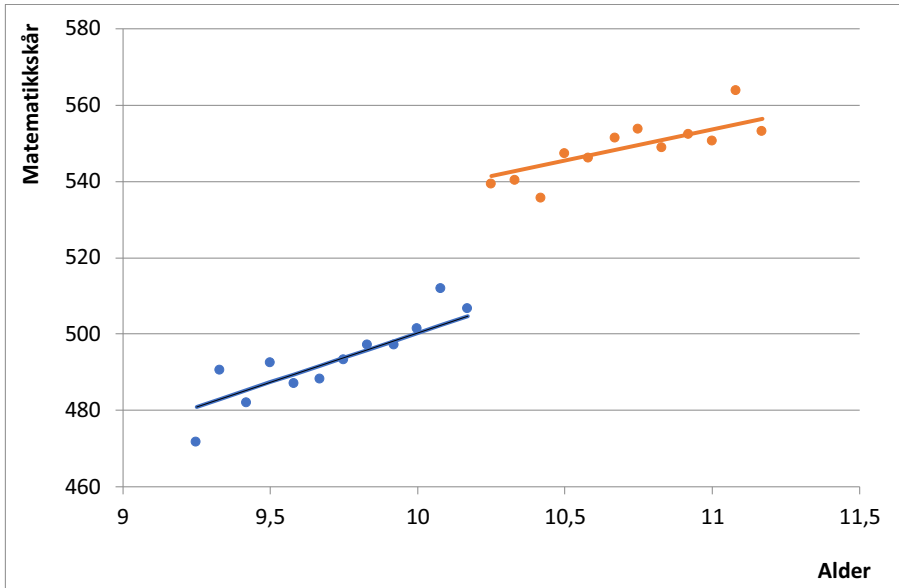
$$y = \beta_0 + \beta_1 \text{ Alder} + \beta_2 \text{ Trinn}(\text{dummy}) + \beta_3 \text{ Alder} \cdot \text{Trinn}$$

I tillegg til alder er her klassetrinn inkludert som en dummyvariabel med verdien 0 for elevene i det lavere alderstrinnet (henholdsvis 4. og 8.) og 1 for elever i det høyere klassetrinnet (5. og 9.). β_1 er et estimat for RAE, mens β_2 gir et estimat på trinneffekten. I tillegg er et interaksjonsledd mellom alder og trinn inkludert for å kunne teste om det er en signifikant endring i RAE fra det lavere alderstrinnet til det høyere. Resultatene for de to analysene er gjengitt i tabell 4.3, og de er også grafisk framstilt i figur 4.3 og 4.4.

TABELL 4.3 Resultater fra to regresjonsanalyser. Signifikante koeffisienter er uthevet, og standardfeil for estimatene er gitt i parentes.

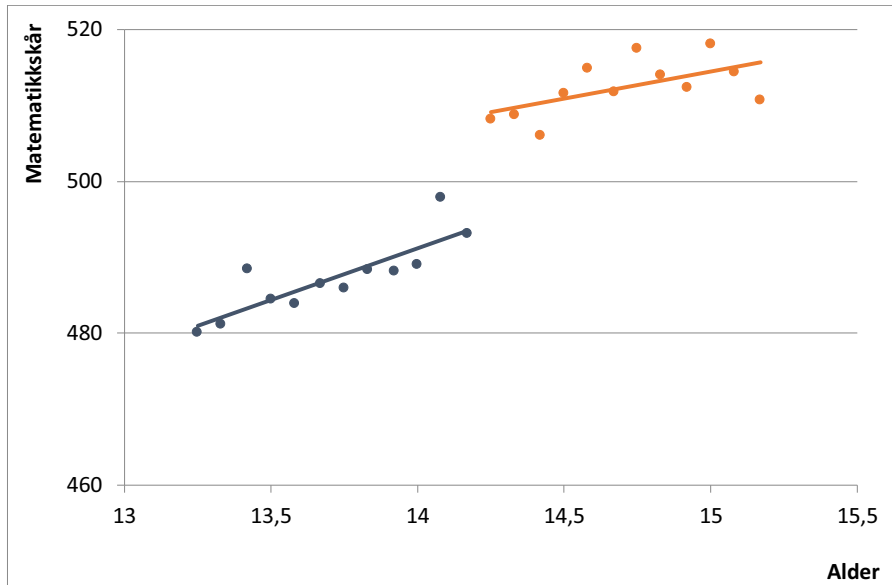
	4/5 trinn	8/9 trinn
β_0	476 (4,5)	481 (2,5)
β_1	25,9 (5,9)	13,6 (3,9)
β_2	45,9 (7,9)	21,0 (7,9)
β_3	-9,6 (6,9)	-6,5 (6,1)
R^2	0,14 (0,01)	0,04 (0,01)

I korte trekk viser analysen at effekten av ett år på skolen er om lag 46 poeng for overgangen 4. til 5. trinn, mens den er halvert for 8./9. trinn. Effekten av at elevene har blitt 12 måneder eldre, er henholdsvis 26 og 14 poeng. Koeffisienten for interaksjonsleddet (β_3) indikerer at RAE faller fra det lavere til det høyere av de to trinnene, et funn som er konsistent med den overordnede tendensen om at RAE faller gjennom skoleløpet (figur 4.2), men i RD-analysen er ikke dette funnet statistisk signifikant.



FIGUR 4.3 Diskontinuitetsanalyse for 4. og 5. trinn.

I begge de to RD-analysene er effekten av RAE over ett kalenderår om lag 60 % av effekten av ett års skolegang. Både RAE og trinneffekten er altså betydelig mindre i RD-analysen for den eldste aldersgruppen, men forholdet mellom RAE og trinneffekten er ganske stabil på 0,6 (60 %).



FIGUR 4.4 Diskontinuitetsanalyse for 8. og 9. trinn.

DISKUSJON

Våre analyser viser at estimatet for RAE varierer svært mye på tvers av undersøkelsesår for den samme alderspopulasjonen. RAE for 4. trinn i TIMSS matematikk varierer eksempelvis mellom 19 poeng i 2011 og 43 poeng i 2007. Det er vanskelig å finne en god, plausibel og substansiell forklaring for hvorfor RAE fluktuerer så mye fra én studie til den neste. Forklaringen kan ikke være at alderssammensetningen i de ulike årskullene varierer; det ble sjekket og viste seg ikke å være tilfellet. Standardfeilene for regresjonskoeffisientene er imidlertid så store at svingninger i denne størrelsesorden i stor grad kan være knyttet til utvalgstilfeldigheter alene. I lys av disse inkonsistente funnene bidrar vår analyse med mer stabile estimater for RAE i de internasjonale undersøkelsene gjennom å sammenholde analyser fra et større antall datasett. Innenfor samme undersøkelse og samme år er imidlertid estimatene for RAE svært stabile på tvers av ulike faglige domener.

Det er ingen kjønnsforskjeller i RAE for 4. og 8. trinn, men guttene har høyere RAE enn jentene på 10. trinn. Dette funnet er interessant i lys av noen av de mest vanlig foreslåtte forklaringsmekanismene for RAE. For det første kan større RAE for gutter gjenspeile at variasjonen for en rekke både biologiske og kognitive stør-

relser er større for gutter enn for jenter (Lehre, Lehre, Laake, & Danbolt, 2009). I tillegg fremsettes ofte hypotesen om at ulikheter i RAE mellom kjønn kan knyttes til ulik biologisk modningstakt. I Norge har dette fenomenet blitt aktualisert gjennom debatt i mediene i 2017, knyttet til at guttene presterer relativt svakere enn jentene i så å si alle fag på skolen. Utgangspunktet for debatten var en kronikk av lederen for Folkehelseinstituttet, Camilla Stoltenberg (2017), som ut fra et slikt argument løftet fram ideen om å vurdere mer fleksible ordninger for skolestart. Modningshypotesen tilsier dermed at flere av (de umodne) guttene med fordel kan starte senere på skolen. De resultatene vi presenterer fra de internasjonale undersøkelsene, kan ikke brukes for å tilføre denne debatten svært sikker kunnskap, men vi kan likevel slå fast at data fra de internasjonale undersøkelsene ikke gir støtte til en slik modningshypotese. Dersom en hypotese om at jenters og gutters ulike modningstakt skulle være en riktig og viktig faktor for å forstå kjønnsforskjeller i skolen, burde dette gjenspeiles i kjønnsforskjeller i både gjennomsnittlige prestasjoner og RAE for de yngste elevene. Kjønnsforskjeller for norske elever i de internasjonale undersøkelsene er generelt små, med unntak for lesing. For dette fagområdet er imidlertid kjønnsforskjellene betydelig større for de relativt eldre elevene i PISA (ca. 40 poeng) enn i PIRLS²⁸ (ca. 20 poeng). Videre ser vi også at kjønnsforskjellen i RAE kun er beskjedent statistisk signifikant for 15-åringer i PISA, mens det ikke er noen kjønnsforskjell i RAE for de yngre elevene.

De norske dataene er spesielt egnet til å studere RAE fordi man for flere år har inkludert to tilgrensende klassetrinn i TIMSS-undersøkelsene (1995, 2011 og 2015). I dette kapitlet presenterte vi derfor en såkalt regresjonsdiskontinuitetsanalyse for dataene fra 2015. I en slik analyse kan man separere alderseffekten (RAE) fra effekten av tid i skolen (trinneffekten). Et hovedfunn i analysen som er presentert i dette kapitlet, er at både RAE og trinneffekten er betydelige. Det er interessant å registrere at forholdet mellom RAE og trinneffekten er betydelig høyere i vår analyse enn i tilsvarende analyse av data fra TIMSS 1995. I analysen av data fra TIMSS 2015 var forholdet omtrent 0,6, mens analyser fra 1995-data gir at forholdet er omtrent 0,3 (Luyten & Veldkamp, 2011). Martin mfl. (2011) har gjort en tilsvarende analyse av RDA for data for 4. og 5. trinn i PIRLS 2006 med data fra Norge, og der er forholdet snudd helt på hodet med en RAE som er dobbelt så stor som trinneffekten. Det kan kanskje tillegges at et forhold hvor RAE er større enn trinneffekten for lesing, kan forstås ut fra at eldre elever vil ha lest mer også i hjemmet enn yngre elever, og at dette gir et viktig bidrag utover leseopplæring i skolen. Et slikt resonnement støttes imidlertid ikke av våre analyser av PISA-data,

28. Progress in International Reading Study, en internasjonal leseundersøkelse for 4./5. trinn.

som viser at alderseffekten er omtrent like stor for lesing som for fagområdene naturfag og matematikk.

Våre analyser bekrefter funn fra tidligere forskning om at alderseffekten avtar utover i skoleløpet, men det finnes fortsatt en ikke neglisjerbar effekt også i PISA-studiene hvor elever fra 10. trinn deltar. Riktignok gir 7 av de totalt 18 analysene av PISA-data på tvers av undersøkelsesår og fagområder en ikke-signifikant RAE. Det er opplagt noen tolkningsutfordringer knyttet til det å bruke disse dataene til å slutte at RAE avtar oppover i skoleløpet. For det første kan vi ikke her med full sikkerhet si at de faglige prøvene i matematikk eller naturfag i de ulike studiene måler det samme begrepet. Det er opplagt ulikheter i matematikkbegrepet som måles i PISA, sammenliknet med TIMSS, hvor man enkelt kan si at PISA i større grad måler grunnleggende regneferdigheter, mens TIMSS måler dyktighet i skolefaget matematikk. Prøvene for 4., 8. og 10. trinn er med andre ord ikke eksplisitt laget for å måle det samme begrepet. Det er likevel interessant at våre funn er konsistente med funn fra en engelsk studie hvor RAE ble undersøkt i et vertikalt lenket prøvesystem (Luyten et al., 2017). I likhet med Norge er også det engelske skolesystemet i stor grad basert på en relativt streng tolkning og bruk av aldersbestemte skolestartkriterier. Også i England er det relativt liten bruk av akselererte eller forsinkede skoleløp. Vi konkluderer derfor med at våre funn i stor grad gir støtte til den eksisterende forskningen, som viser at RAE avtar i løpet av den obligatoriske skolegangen. Den er likevel ikke helt borte på et tidspunkt hvor elevene er ferdig med sin obligatoriske skolegang. Vi kan derfor anta at relativt eldre elever har et noe bedre utgangspunkt for å søke opptak til videregående skoler.

Resultatene som er presentert her, er en del av et pågående forskningsprosjekt hvor det endelige formålet er å bruke den internasjonale variasjonen i relevante systemvariabler. Alder for skolestart varierer for eksempel typisk fra fem til sju år på tvers av land. Noen land har, som Norge, veldig rigide aldersbestemte kriterier for skolestart, mens andre land har ordninger hvor barn kan starte senere eller tidligere, og typisk vil dette gjøre seg gjeldende for barn som er født i måneden før/etter den definerte fødselsdatoen for skolestart. I noen få land har man ordninger med en mer gradvis innfasing i skolen ved at barnet kan møte opp på skolen den dagen det fyller et bestemt antall år (f.eks. i New Zealand og i Nederland). Videre finnes det variasjoner mellom land når det gjelder å la barn hoppe over eller sitte igjen et skoleår, en praksis som spesielt finner sted i løpet av de første skoleårene. Det er rimelig å anta at alle disse praksisene, på ett eller annet slags vis, er begrunnet ut fra en idé om hva som er best for barnets videre skolegang. En studie på tvers av land har derfor i utgangspunktet et potensiale for å belyse hva som er virkninger av ulik politikk på området. I tillegg har de internasjonale undersøkelsene

et annet gunstig designelement: Man har data fra de samme landene over tid. Det bør derfor være mulig å identifisere enkeltland som har hatt en (relativt plutselig) endring i politikk som det er grunn til å tro kan påvirke RAE. Norge er et slikt land hvor praksis har endret seg for flere av disse interessante systemvariablene. Denne historikken er grundigere beskrevet blant annet av Gabrielsen og Lundetræ (2017). Fra 1950-tallet og fram mot midten av 1970-tallet var det ikke uvanlig å bruke psykologisk testing for å bestemme barnets grad av skolemodenhet. I denne perioden var det derfor sannsynligvis slik at barn født i nærheten av alderskriteriet i større grad fordelte seg over to klassetrinn, og man fikk på denne måten et vagere aldersskille mellom klassetrinn. Endringen til dagens praksis med sammenholdte alderskohorter i klassetrinn skjedde imidlertid før den perioden vi dekker i våre analyser, og vi har derfor ikke mulighet til å studere effekten av denne endringen. Det er i stedet en annen endring som har skjedd i perioden, som i utgangspunktet kunne vært en interessant kilde til variasjon over tid. I 1997 ble skolestart for seksåringer innført. Dermed representerer noen av de tidlige datasettene kohorter med elever som begynte på skolen det året de fylte sju år, mens de resterende kohortene hadde skolestart som seksåringer. I analysene som vi har presentert her, finner vi imidlertid ikke at denne endringen i alder ved skolestart har påvirket RAE.

I vårt forskningsprosjekt ønsker vi videre helt spesifikt å undersøke hvilken betydning prestasjonsbasert seleksjon i tidlig alder har på RAE. Som vi har vist, er RAE svært stor for de yngste elevene. Det er derfor en fare for at man selekterer på aldersforskjellen i seg selv. På denne måten kan en slik relativt liten forskjell tidlig i skoleløpet bli forsterket og forstørret over tid. Dette vil i så fall være et fenomen parallelt til det som er dokumentert i konkurransepregede idretter med skarpe aldersklasser. Her er et gjennomgående funn at eliteutøvere i idretten på juniornivå og i voksen alder har en svært skjev aldersfordeling i retning av dem som er født tidligst i alderskohortene (Cobley et al., 2009).

Avslutningsvis vil vi imidlertid påpeke at heller ikke data fra de internasjonale undersøkelsene vil gi oss endelige svar på spørsmål om effekten av politikk knyttet til sen/tidlig/fleksibel skolestart, forsinket/forsert løp eller seleksjon av elever i ulike skoletilbud. Det er en rekke metodiske utfordringer knyttet til det å analysere slike effekter på tvers av land i tverrsnittundersøkelser av denne typen. En hovedutfordring er relatert til hvordan populasjonene i de internasjonale undersøkelsene er definert. I TIMSS dekker populasjonene nasjonalt representative utvalg på 4. og 8. trinn (5. og 9. i Norge fra og med 2015). I land hvor det har vært en tidlig seleksjonsmekanisme, vil man derfor ha en populasjon som ikke lenger inkluderer hele den aktuelle alderskohorten. I disse landene vil RAE typisk være forskjøvet mot lavere verdier fordi en del av de svakest presterende yngste elevene

i alderskohorten befinner seg på et lavere klassetrinn, mens flere av de skoleflinke eldre elevene følger et akselerert løp og er i et høyere trinn. I PISA er det et omvendt problem ved at utvalget dekker en komplett alderskohort basert på fødselsår. I mange land vil fødselsdato for skolestart være satt til august/september, og alderskohorten som er inkludert i PISA, vil derfor i utgangspunktet, selv i land uten seleksjonsmekanismer, være fordelt i to klassetrinn i TIMSS-studien. I tillegg vil seleksjonsmekanismene slå inn slik at populasjonen ikke lenger representerer ett eller to tilgrensende klassetrinn.

Selv om vi vil være i stand til å løse de analytiske utfordringene som slike utvalgstekniske utfordringer gir oss, er det likevel ikke sikkert man vil kunne få robust og sikker kunnskap. Tilsynelatende lik politikk på et område kan vise seg å være ganske ulike i praksis. Fleksibel skolestart kan for eksempel i noen land være et valg som foreldrene gjør på eget grunnlag, mens det i andre land vil være en pedagogisk-psykologisk tjeneste hvor resultater fra skolemodenhetstester og intervjuer med barn og foreldre blir brukt for å avgjøre tidlig eller utsatt skolestart. I tillegg vil det, fordi kunnskapen om hva som er god/dårlig politikk på området er usikker, sannsynligvis finnes ulike dominerende forestillinger om hva som er «godt for barnet». I noen land vil forsinket skolestart bli sett på som et «gode», mens tidlig skolestart i andre land vil oppfattes som en fordel. Slike forestillinger vil påvirke de valgene som foreldrene gjør. Til sammen kan dette gi skjevheter i hvem det er som ender opp med tidlig eller sen skolestart. I en preliminær analyse av PISA-data fra for eksempel Tyskland ser vi at elevgruppen med forsert løp i gjennomsnitt har foreldre med høyere utdanning. De motsatte karakteristikkenes finnes for elever med et forsinket løp. Til sammenlikning finnes det også mange elever i Irland med forserte og forsinkede løp, men her er sammenhengen med elevens bakgrunn merkbart svakere.

Ambisjonen vår er likevel å finne gode løsninger på disse analytiske og konseptuelle utfordringene i vårt videre arbeid. Det er opplagt et sterkere utvalgsdesign for de studiene hvor to naboliggende klassetrinn deltar, men dette har i liten grad blitt gjort i andre land enn Norge. En anbefaling er derfor å gjennomføre TIMSS i flere klassetrinn på tvers av land, slik det ble gjort i 1995. En annen mulighet som vi vil se på, er analyse av PISA-data for de landene som i 2009 og i 2012 hadde et utvalg som representerte både en alderskohort og et bestemt klassetrinn. I 2009 var det åtte land som hadde et slikt utvalg, og i 2012 var det fem land. Disse vil vi studere nærmere.

REFERANSER

- Aune, T. K., Pedersen, A. V., Ingvaldsen, R. P., & Dalen, T. (2017). Relative Age Effect and Gender Differences in Physical Education Attainment in Norwegian Schoolchildren. *Scandinavian Journal of Educational Research*, 61(3), 369–375. doi:10.1080/00313831.2016.1148073
- Bedard, K., & Dhuey, E. (2006). The persistence of early childhood maturity: international evidence of long-run age effects. *Quarterly Journal of Economics*, cxxi(4), 1437–1472.
- Black, S. E., Devereux, P. J., & Salvanes, K. G. (2011). Too Young to Leave the Nest? The Effects of School Starting Age. *The Review of Economics and Statistics*, 93(2), 455–467. doi:10.1162/REST_a_00081
- Chen, M.-H., Lan, W.-H., Bai, Y.-M., Huang, K.-L., Su, T.-P., Tsai, S.-J., ... Hsu, J.-W. (2016). Influence of Relative Age on Diagnosis and Treatment of Attention-Deficit Hyperactivity Disorder in Taiwanese Children. *The Journal of Pediatrics*, 172, 162–167.e161. doi:10.1016/j.jpeds.2016.02.012
- 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 1995 data. *Educational Research and Evaluation*, 16(1), 39–52. doi:10.1080/13803611003694391
- Cobley, S., Baker, J., Wattie, N., & McKenna, J. (2009). Annual Age-Grouping and Athlete Development. *Sports Medicine*, 39(3), 235–256. doi:10.2165/00007256-200939030-00005
- Gabrielsen, E., & Lundetræ, K. (2017), kap 11. Indikerer de norske PIRLS resultatene et behov for å justere retningslinjene for skolestartsalder? In E. Gabrielsen (Ed.), *Klar framgang!* (pp. 204–221). Oslo: Universitetsforlaget.
- Kawaguchi, D. (2006). The Effect of Age at School Entry on Education and Income. Lastet ned 3. januar 2018 fra <http://EconPapers.repec.org/RePEc:esj:esridp:162>
- Lehre, A.-C., Lehre, K. P., Laake, P., & Danbolt, N. C. (2009). Greater intrasex phenotype variability in males than in females is a fundamental aspect of the gender differences in humans. *Developmental Psychobiology*, 51(2), 198–206. doi:10.1002/dev.20358
- Lien, L., Tamsb, K., Oppedal, B., Heyerdahl, S., & Bjertnes, E. (2005). Is relatively young age within a school year a risk factor for mental health problems and poor school performance? A population-based cross-sectional study of Adolescents in Oslo, Norway. *BMC Public Health*, 5(102).
- Luyten, H. (2006). An empirical assessment of the absolute effect of schooling: regression? discontinuity applied to TIMSS?95. *Oxford Review of Education*, 32(3), 397–429. doi:10.1080/03054980600776589
- Luyten, H., Merrell, C., & Tymms, P. (2017). The contribution of schooling to learning gains of pupils in Years 1 to 6. *School Effectiveness and School Improvement*, 28(3), 374–405. doi:10.1080/09243453.2017.1297312
- Luyten, H., & Veldkamp, B. (2011). Assessing Effects of Schooling With Cross-Sectional Data: Between-Grades Differences Addressed as a Selection-Bias Problem. *Journal of Research on Educational Effectiveness*, 4(3), 264–288. doi:10.1080/19345747.2010.519825
- Martin, M. O., Mullis, I. V. S., & Foy, P. (2011). Age distribution and reading achievement configurations among fourth-grade students in PIRLS 2006. *IERI Monograph Series; Issues and Methodologies in Large-Scale Assessments*, 4.

- Morrow, R. L., Garland, E. J., Wright, J. M., Maclure, M., Taylor, S., & Dormuth, C. R. (2012). Influence of relative age on diagnosis and treatment of attention-deficit/hyperactivity disorder in children. *Canadian Medical Association Journal*, 184(7), 755–762. doi:<http://dx.doi.org/10.1503/cmaj.111619>
- Ponzo, M., & Scoppa, V. (2014). The long-lasting effects of school entry age: Evidence from Italian students. *Journal of Policy Modeling*, 36(3), 578–599. doi:<http://dx.doi.org/10.1016/j.jpolmod.2014.04.001>
- Solli, I. F. (2017). Left behind by birth month. *Education Economics*, 25(4), 323–346. doi:[10.1080/09645292.2017.1287881](https://doi.org/10.1080/09645292.2017.1287881)
- Sprietsma, M. (2010). Effect of relative age in the first grade of primary school on long-term scholastic results: international comparative evidence using PISA 2003. *Education Economics*, 18(1), 1–32. doi:[10.1080/09645290802201961](https://doi.org/10.1080/09645290802201961)
- Stoltenberg, C. (2017). Vi har skapt et nytt kjønns gap før vi har rukket å kvitte oss med det gamle. *Morgenbladet*, 10. februar.
- Strøm, B. (2004). Student achievement and birthday effects. *Norwegian University for Science and Technology working paper*.
- von Davier, M. (2014). Imputing Proficiency data under Planned Missingness in Population Models. In: Rutkowski, L., Davier, M. v. & Rutkowski, D. (2014). *Handbook of International Large-scale Assessment: background, technical issues, and methods of data analysis*. In Statistics in the social and behavioral sciences series. Boca Raton: Chapman & Hall/CRC Press.