

6

Hva forklarer endringer i elevenes matematikkprestasjoner over tid?

ROLF VEGAR OLSEN OG SIGRID BLÖMEKE

SAMMENDRAG TIMSS-undersøkelsene har vist at norske elever på 4. og 8. trinn har hatt en framgang i matematikk- og naturfagprestasjoner siden 2003. Dette kom etter en tilsvarende stor tilbakegang fra 1995 til 2003. I dette kapitlet presenteres en analyse hvor endringen i gjennomsnittlig matematikkskår for 8. trinn fra 2003 til 2015 blir knyttet til endringer for 22 bakgrunnsvariabler. Generelt kan endringer i elevprestasjoner knyttes til flere fundamentalt ulike kilder: Sammensetningen av elevkohortene kan endre seg, undervisningens innhold og kvalitet kan endre seg, og sammenhengene mellom bakgrunnsfaktorer og prestasjoner i matematikk kan endre seg over tid. Gjennom å anvende Oaxaca-Blinder-dekomponering (OBD) kan slike mulige bidrag til framgangen separeres og tallfestes. Analysen gir ikke et entydig bidrag til å forstå den store framgangen. Et funn som kan framheves, er at faktorer knyttet til elevenes selvfølelse og trivsel framstår som viktige.

SUMMARY The scores in mathematics and science have increased for Norwegian grade 4 and 8 students from the 2003 to the 2015 TIMSS study. This increase followed an equally large decline from 1995 to 2003. This chapter presents and discusses the results from an analysis where the change in mathematics performance from 2003 to 2015 is coupled with changes for 22 background variables. In general, a change can be associated with a few fundamentally different sources: Characteristics of the students' cohort are not necessarily constant over time, the content and quality of instruction may change, and the relationship between background characteristics and students' performance may change over time. By applying Oaxaca-Blinder decomposition (OBD) the contributions from these various sources are separated and estimated. However, the analysis does not provide a very clear picture. One finding can be emphasised: Factors related to students' self-concept and well-being appear as important.

INNLEDNING

I perioden som har gått siden den første internasjonale storskalaundersøkelsen ble gjennomført i 1995, kan norsk skole sies å være preget av både stabilitet og endring. I hele perioden har norsk skole holdt fast på enhetsskoleprinsippet, utdanning er i all hovedsak offentlig finansiert, bosted avgjør skoletilhørighet, det er sentralgitte læreplaner, det gis ingen karakterer på barnetrinnet, individuelt tilpasset opplæring er lovpålagt, og grunnskolen har i all hovedsak lærere med allmennlærerbakgrunn, for å nevne noen egenskaper som det er stor politisk enighet om. (Se for eksempel St.meld. nr. 16, 2006–2007.) Gjennom norsk deltakelse i de internasjonale undersøkelsene er dette karakteristikk ved vårt skolesystem som i sammenlikning med andre land framstår som langt fra selvsagte.

Samtidig har perioden vært preget av to læreplanreformer (L97 og Kunnskapsløftet) og mange andre tilknyttede endringer i skolesystem og politikk. Det er imidlertid kjent at det tar lang tid før effekter av reformer i utdanningssektoren kan observeres (Fullan, 2007; Tyack & Cuban, 1995). I tillegg er slike reformer ofte bredt anlagt, og de forandres underveis i implementeringsprosessen, noe som også gjør det vanskelig å studere hvilken effekt de har (Sarason, 1990). Analyser av de norske reformene har også kommet til tilsvarende konklusjoner (Aasen et al., 2012; Haug, 2003), og reformene bør derfor heller oppfattes som bevegelige mål (Olsen, Hopfenbeck, & Lillejord, 2013). Endringer i perioden som helt eller delvis er tatt opp i reformene, er for eksempel tidligere skolestart, innføring av nye læreplandokumenter med sterkere vektlegging av kompetanser og grunnleggende ferdigheter i fag, innføring av et nasjonalt kvalitetsvurderingssystem (NKVS) med blant annet nasjonale prøver i sentrale fagområder og satsing på vurdering for læring. Samtidig er det andre gradvise endringer som i mindre grad kan knyttes til konkret politikkutforming. Eksempler på dette kan være endringer i lærernes kompetanser, hjemmets holdninger og støtte til barnets skolegang og prioriteringer for fordeling av ressurser m.m. Til sammen tilsier dette at det ikke er helt enkelt å identifisere kilder eller mulige årsaker til endringer i de norske resultatene i de internasjonale undersøkelsene i perioden.

Vi presenterer i dette kapitlet en analyse som nettopp har som formål å identifisere mulige kilder til endringer i gjennomsnittsprestasjonene for Norge i perioden 2003 til 2015. Nærmere bestemt har vi valgt å se på endringene i matematikkprestasjoner i TIMSS-undersøkelsen for 8. trinn fra 2003 til 2015. Dette er ikke et helt tilfeldig valg. For det første valgte vi å bruke data fra TIMSS fordi resultatene er preget av relativt store endringer med en tydelig trend over tid. I PISA-undersøkelsen fluktuierer resultatene rundt det internasjonale gjennomsnittet uten en tydelig retning for endringen (Kjærnsli & Jensen, 2016). Det er derfor vanskelig

å bruke data fra PISA for å studere mulige kilder til endringer, utover den opplagte konklusjonen at prestasjonsnivået synes å være tilnærmet konstant over tid. For det andre gir TIMSS oss en mulighet til å inkludere lærer- og klasseromvariabler. Dersom endringer på systemnivå over tid skal få effekter, er det grunn til å vektlegge analyser av nettopp klasseromnivået fordi variablene på dette nivået ligger nærmere elevenes prestasjonsutvikling i en hypotetisk kausal sammenheng enn variabler på skolenivå. PISA har interessante variabler på skolenivå, men dersom disse til sist skal ha en effekt på elevprestasjoner, må de medieres gjennom lærere, elever og foreldre.

For det tredje har vi valgt kun å inkludere matematikk i perioden 2003–2015 fordi dette gir en periode med en tilnærmet konstant positiv endring over tid. I analysene skulle vi nok helst ha inkludert også perioden 1995–2003 hvor det var en tilsvarende stor nedgang i de norske elevenes prestasjoner, men vi har valgt bort denne perioden fordi spørreskjemaene til lærere og elever er lite stabile over tid. Ved å inkludere datasett fra flere år i analysene ville det knapt vært sammenliknbare data over hele perioden. For det fjerde har vi valgt å se på 8. trinn til tross for at endringene i perioden er større for 4. trinn. Dette skyldes i all hovedsak at informasjonen fra elevspørreskjemaene er rikere og sannsynligvis også mer valide for elevene på dette trinnet.

I de nasjonale og internasjonale rapportene er det gjort få analyser av hvordan framgangen i matematikkprestasjoner i perioden eventuelt er relatert til bakgrunnsinformasjon om elevene, lærerne og skolene. Det eneste eksemplet vi kjenner til, er Nilsen, Grønmo og Hole (2013), som viste at en samlevariabel for skoleledernes vurdering av lærernes, skolenes og foreldrenes vektlegging av faglige prestasjoner (eller deres akademiske ambisjoner,³¹ som vi i det videre vil betegne dette som) økte fra 2007 til 2011, og videre viste de at denne økningen var sterkt relatert også til bedringen i skår på matematikkprøven for norske elever. Videre viser Nilsen og Gustafsson (2016) i en analyse av endringer på landsnivå at lærernes utdanningsnivå, lærernes deltakelse i etterutdanning og foreldrenes akademiske ambisjoner er relatert til økning i matematikkskår på 8. trinn fra 2007 til 2011.

Resultatene viser for øvrig at norske elever er blitt dyktigere på oppgaver knyttet til områdene statistikk, tall og geometri (Bergem, 2016). I tillegg er det tydelig at det norske gjennomsnittet hovedsakelig har økt fordi færre elever har svært lave skårer. Andelen elever som skårer i det laveste kompetansenivået (under 400 poeng), har falt fra 19 % til 10 % fra 2003 til 2015. (Se for øvrig også prosentil-

31. På engelsk er begrepet kalt «School emphasis on academic success» (SEAS), mens Nilsen mfl. her oversatte dette til «læringstrykk».

fordelingen i figur 6.1.) Det er med andre ord et behov for analyser som i sterkere grad bruker dataene fra de internasjonale undersøkelsene for å belyse mulige årsaker til endringer over tid.

UTVALG AV VARIABLER I ANALYSEN

I utgangspunktet tilsier den metoden vi bruker (OBD) at det er mest interessant å inkludere forklaringsvariabler som selv endrer seg i perioden, og/eller hvor sammenhengen med prestasjoner i matematikk endrer seg i perioden. Vi ønsket også at analysen skal være basert på variabler som reflekterer pedagogisk interessante og policyrelevante variabler som i en hypotetisk årsakskjede ligger nær elevens verden. I tillegg er det viktig å inkludere elev- og lærerkaraktistikker som også kan ha endret seg i perioden.

For å kunne bruke OBD er det også nødvendig å ta hensyn til mer tekniske sider ved svarfordelingene til variablene:

- ▶ Variablene må være med i datasettene både fra 2003 og 2015 med identiske spørsmål og svarskalaer. Dette kravet utelukker de aller fleste variablene fra de to datasettene.
- ▶ Variabler hvor store andeler av elevene ikke har svart, må utelates fordi de bidrar til at utvalget i analysen reduseres. Foreldrenes utdanningsbakgrunn er et eksempel på en variabel som vi ikke har kunnet inkludere fordi om lag halvparten av elevene enten ikke svarer eller oppgir «vet ikke». Vi bruker i stedet antall bøker hjemme som er sterkt relatert til utdanningsbakgrunn. Det finnes også lave andeler *ikke-svar* for de variablene som til slutt har blitt inkludert i regresjonsmodellen, og for disse har vi valgt å erstatte (eller imputere) manglende verdier ved å bruke *Expectation Maximization*-metoden som er implementert i SPSS, for å unngå at vi må eliminere alle elever som har en manglende verdi. EM-metoden bruker verdiene fra alle andre variabler for en elev for å finne ut hva den mest sannsynlige eller forventede verdien hadde vært, gitt alt vi ellers vet om eleven. Dette er en metode som fungerer godt når andel manglende svar er lav, slik det er i vårt tilfelle (typisk 1–3 %).
- ▶ Et annet forhold som kan bidra til å destabilisere regresjonsanalysen, er variabler med merkelige fordelinger. Slike variabler er vanskelig å tolke i en lineær modell fordi svarene fra en gruppe elever med ekstremverdier kan påvirke løsningen uforholdsmessig mye. Et eksempel på en variabel som er utelatt av denne grunnen, er elevenes selvrapporterte tid brukt på hjemmelekser. Denne variabelen har en bimodal fordeling med elever som klumper seg med hen-

- holdsvis svært lave verdier og svært høye verdier, men med få elever i midten. En liten, men betydelig, andel av verdiene er så høye at de sannsynligvis er et resultat av misforståelser hos elevene, f.eks. ved at de har rapportert total tid brukt på alle lekser i stedet for tiden brukt på matematikklekser (som spørsmålet var rettet mot).
- ▀ Et siste teknisk hensyn vi har måttet ta, er å unngå å ha med forklaringsvariabler med svært høye innbyrdes korrelasjoner. Dette gir såkalte multikollinearitetsproblemer (Cohen, Cohen, West, & Aiken, 2003). Variablene som beskriver innvandringsstatus og språk i hjemmet, er et eksempel på et slikt variabelpar som er høyt korrelert med hverandre. Vi har derfor valgt å utelate variabelen for språk i hjemmet.

Oppsummert inkluderer den endelige regresjonsmodellen et sett med variabler som beskriver elevene (kjønn, alder, innvandringsbakgrunn og antall bøker i hjemmet) og deres lærere (kjønn, alder, erfaring, utdanningsbakgrunn og deltakelse i etterutdanning). I tillegg er noen variabler knyttet til elevenes affektive vurderinger av faget og skolen (liker seg på skolen, selvbilde i og verdsetting av matematikkfaget), skolemiljø (klassestørrelse, lærere føler seg trygge på skolen, læreres, foreldres og elevers akademiske ambisjoner) og undervisningen (bruk av åpne problemer, bruk av kalkulator og lekser) inkludert i modellen. En mer detaljert beskrivelse av variablene som inngår i modellen, er gjengitt i vedlegg 1.

OAXACA-BLINDER-DEKOMPONERING

Utgangspunktet for metoden som vi bruker i denne analysen, er en ordinær regresjonsanalyse.³² Den avhengige variabelen er elevenes skår i matematikk, og de uavhengige variablene er et sett med variabler fra spørreskjemaene til lærere og elever. To separate regresjonsanalyser ble først gjennomført i datasettene som representerer de to gruppene som skal bli sammenliknet, her utvalgene av elever i TIMSS-studien for 8. trinn matematikk i henholdsvis 2003 og 2015. Regresjonsanalysene er gjort med programvaren IDB Analyzer. Denne programvaren er spesifikt laget for å håndtere analyser fra de internasjonale undersøkelsene gjennom å inkludere de fem plausible verdiene for matematikkskår og vektene som representerer det komplekse utvalgsdesignet. Siden vi har inkludert både elev- og lærerdata i analysen, bruker vi de såkalte Math Teacher Weights (Foy, 2017, s. 32

32. Vi har også gjort en flernivåanalyse i programvaren Mplus. En sammenlikning mellom denne analysen og den analysen vi presenterer her, viser at både regresjonskoeffisientene og standardfeilene er relativt like.

og 65). Dette sikrer at data fra lærerne blir inkludert i analysen ut fra hvor stor andel av elevene de representerer.

Resultatene fra dette første steget i analysen blir i et neste steg dekomponert i tre ledd som beskrevet nedenfor. Det er dette som er kjent som Oaxaca-Blinder-dekomponering (Blinder, 1973; Elder, Goddeeris, & Haider, 2010; Oaxaca, 1973). OBD er en analyse som spesifikt forsøker å svare på spørsmålet: Hva ville skjedd dersom verdiene for henholdsvis gjennomsnittet (\bar{X}) og regresjonskoeffisientene (β) hadde vært de samme i 2015 som i 2003? Eller med andre ord: I hvilken grad kan endringene i matematikkskår forstås ved at det er den uavhengige variabelen som i perioden har endret seg, både ved at gjennomsnittet har endret seg, og/eller ved at styrken i relasjonen med den avhengige variabelen har endret seg? Det er med andre ord en metode som stiller et *kontrafaktisk* spørsmål. I likningen nedenfor vises prinsippene for dekomponeringen i det enkle tilfellet hvor det kun er én enkelt uavhengig variabel X :

$$\bar{Y}_{15} - \bar{Y}_{03} = \underbrace{(\bar{X}_{15} - \bar{X}_{03}) \cdot \beta_{03}}_{\text{Gjennomsnittsleddet}} + \underbrace{(\beta_{15} - \beta_{03}) \cdot \bar{X}_{03}}_{\text{Koeffisientleddet}} + \underbrace{(\bar{X}_{15} - \bar{X}_{03}) \cdot (\beta_{15} - \beta_{03})}_{\text{Interaksjonsleddet}}$$

I denne likningen modelleres forskjellen i *gjennomsnittene* for den avhengige variabelen (\bar{Y}) for to grupper (her matematikkskår i TIMSS 8. trinn i hhv. 2003 og 2015) som en sum av tre ulike komponenter. Disse komponentene inkluderer gjennomsnittsverdiene i de to gruppene for den uavhengige variabelen (\bar{X} ; f.eks. elevenes bakgrunn eller lærerkvalitet) og koeffisientene fra regresjonsanalysene (β) som uttrykker sammenhengen mellom den uavhengige variabelen og matematikkskåren.

Det første leddet på høyresiden i uttrykket angir den delen av differansen i matematikkskåren som kan knyttes til endringer i gjennomsnittsverdien av den uavhengige variabelen, gitt at styrken i sammenhengen er den samme begge årene. Vi vil her benevne dette som *gjennomsnittsleddet*³³. Denne komponenten kalles ofte også for det forklarte leddet. I vår analyse kan dette eksempelvis være endringer som reflekterer at sammensetningen av elevene har endret seg i perioden, for eksempel at man i 2015 har flere elever med innvandringsbakgrunn enn i 2003.

Den andre komponenten uttrykker at endringen i matematikkskår også kan knyttes til at styrken i sammenhengen mellom den avhengige og den uavhengige variabelen har endret seg over tid – av en eller annen (gjerne uobservert) årsak. Eller sagt litt annerledes: Hvor stor del av endringen i matematikkskår kan knyttes til at

33. I engelskspråklig litteratur omtales denne komponenten ofte som «endowments» – et ord som ikke helt enkelt lar seg oversette til norsk.

regresjonskoeffisienten har blitt større eller mindre? Et illustrerende og hypotetisk eksempel kunne være at det å ha en datamaskin tilgjengelig hjemme er sterkere relatert til prestasjoner i 2015 enn i 2003 fordi teknologien i sterkere grad blir brukt som et viktig pedagogisk verktøy i undervisningen. Et annet eksempel er samfunnsendringer på et mer overordnet nivå som medfører at elevers sosioøkonomiske bakgrunn har en svekket eller styrket sammenheng med prestasjonene i matematikk. Det siste er i så fall et eksempel på at disse leddene indirekte inkluderer også endringer i ikke-observerte variabler, og i så måte representerer dette leddet i likningen ikke-forklarte endringer. Dette leddet kalles ofte *koeffisientleddet*, med andre ord endringene som er knyttet til at regresjonskoeffisientene endrer seg.

En hovedutfordring med å tolke koeffisientleddene i OBD er knyttet til valg av skala for de uavhengige variablene (Jann, 2008). Et illustrerende eksempel på dette er variabelen for elevenes alder. Man kan velge å sette nullpunktet ved fødselsåret, men det kan være mer relevant å oppgi alderen som relativt til den yngste eleven. I det første tilfellet vil man i formelen få høye verdier for \bar{X} (i vår studie ca. 13,5), mens man med en relativ alder får en mye lavere verdi (her ca. 0,5 fordi elevene på ett klassetrinn stort sett er født i samme kalenderår). Dette er et valg som ikke påvirker styrken i sammenhengene. Det påvirker heller ikke forskjellene i gjennomsnittlig alder for de to gruppene, men koeffisientleddet påvirkes av dette valget fordi man i dette leddet multipliserer med \bar{X} . I vår analyse har vi derfor valgt nullpunkter for skalaene som er så meningsfulle som mulig i den gitte konteksten. Gitt for eksempel at nesten alle norske elever på 8. trinn også representerer samme fødselsår, har vi satt aldersvariabelen til 0 for de yngste elevene født i dette «normalåret». Det er noen få enkeltelever som er yngre enn dette, og dermed får en negativ alder, og noen ytterst få som er eldre, og dermed får en verdi for aldersvariabelen som er høyere enn 1. Ut fra samme logikk har vi valgt å sette nullpunktet for klassestørrelse til 0 for en klasse med 15 elever siden nesten ingen klasserom har færre elever enn dette. Av samme grunn har vi omformet alle kategoriske variabler med mer enn to verdier til dikotome variabler. Det er derfor viktig å minne om at størrelsene for ustandardiserte regresjonskoeffisienter også i vår analyse er vanskelig å tolke på tvers av variabler.

I tillegg er det en tredje komponent som angir at det også kan være et samspill mellom endringene i de uavhengige variablene og i regresjonskoeffisientene. Dette leddet kalles derfor *interaksjonsleddet*. Dette leddet er imidlertid også vanskelig å tolke direkte fordi det i likhet med koeffisientleddet er knyttet til at modellen ikke er fullstendig. Koeffisient- og interaksjonsleddene utgjør derfor det som til sammen betegnes som «det uforklarte leddet».

Likningen gitt ovenfor gjelder som sagt for det enkle tilfellet hvor vi kun har én uavhengig variabel (X), men likningen kan generaliseres til å inkludere flere uavhengige variable, i vårt tilfelle 22 stykker. Man legger bare til tilsvarende ledd for hver ny variabel som er inkludert i regresjonsmodellen. Gjennom dette kan man til slutt summere opp alle komponentene som er knyttet til gjennomsnitt-, koeffisient- og interaksjonsleddene.³⁴ Alle estimater kan leses som ustandardiserte parametere.

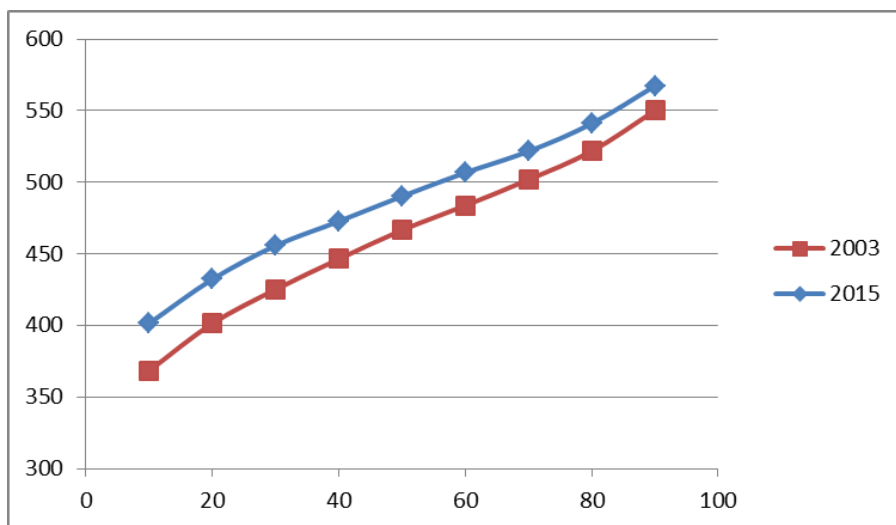
Vedlegg 1 presenterer de detaljerte resultatene for alle de 22 enkeltvariablene som er inkludert i analysen. I teksten og i tabell 6.1 har vi imidlertid valgt å lage delsummer for grupper av variabler med samme type informasjon, dvs. at vi har slått sammen alle leddene for alle variablene knyttet til henholdsvis elevenes bakgrunn (kjønn, alder, innvandringsbakgrunn og antall bøker i hjemmet), affektive faktorer (liker seg på skolen, selvbilde i og verdsetting av matematikkfaget), lærernes bakgrunn (kjønn, alder, erfaring, utdanningsbakgrunn og deltakelse i to typer etterutdanning) og undervisningspraksis (bruk av åpne problemer, to typer bruk av kalkulator og lekser) og skolemiljø (klassestørrelse, lærere føler seg trygge på skolen, læreres, foreldres og elevers akademiske ambisjoner). Selv om OBD er en metode som primært har vært brukt av økonomer for å studere fordeling av goder i samfunnet mellom to grupper, finnes det etter hvert flere studier hvor OBD og tilsvarende dekomponeringsmetoder er brukt på data fra de store internasjonale undersøkelsene, og da spesielt på data fra PISA. I disse artiklene studeres eksempelvis kilder til ulike prestasjoner for elever med bakgrunn fra by eller landsbygd (Liao, Chang, Wang, & Sun, 2016; Lounkaew, 2013; Ramos, Duque, & Nieto, 2012), ulikheter mellom elever med minoritets- eller majoritetsbakgrunn (Ammermueller, 2007b; Lock & Gibson, 2008; Song, Perry, & McConney, 2014), forskjeller i prestasjoner mellom to land (Ammermueller, 2007a), kilder til sosial ulikhet mellom land eller hvorfor sosial ulikhet endrer seg over tid (Martins & Veiga, 2010; Nieto & Ramos, 2015; Oppedisano & Turati, 2015), innvirkning på skår fordi utvalgene over tid ikke er sammenliknbare (Freitas, Nunes, Balcão Reis, Seabra, & Ferro, 2016). Og av spesiell relevans for vårt kapittel: Studier som benytter metoden for å studere kilder til endringer i prestasjoner over tid (Barrera-Osorio, Garcia-Moreno, Patrinos, & Porta, 2011; Cattaneo & Wolter,

34. I tillegg finnes en rekke andre beslektede dekomponeringsmetoder hvor man ikke kun ser på hvordan gjennomsnittet endrer seg, men også inkluderer hvordan spredningen innen to grupper endrer seg. (For en oversikt, se Fortin, Lemieux, & Firpo, 2011.) Modellene blir enda kompleksere og tolkningen av de enkelte leddene enda vanskeligere på denne måten, og vi har derfor her kun valgt en enkel lineær modellering, men vi viser i figur 6.1 at endringen mellom 2003 og 2015 har vært litt ujevn langs prestasjonsskalaen.

2012; Ehmke, Klieme, & Stanat, 2013; Valenzuela, Gómez Vera, & Sotomayor, 2015).

RESULTATER

Figur 6.1 viser hvordan matematikkprestasjonene til norske elever har endret seg. De to linjene viser gjennomsnittlig matematikkskår for elevene i 10.–90. prosentilene for 8. trinns elever i henholdsvis 2003 og 2015. Figuren viser tydelig at forbedringen har skjedd langs hele fordelingen, men med en relativt større endring for elevene lengst nede i fordelingen. I vår analyse modellerer vi endringen for gjennomsnittene i de to årene og tar derfor ikke hensyn til at endringen kan være ulikt knyttet til ulike deler av denne fordelingen. Dermed kan vi forklare størsteparten av endringen, den som går langs hele prestasjonsskalaen, men ikke forskjellen mellom endringene på øvre og nedre delen av fordelingen.



FIGUR 6.1 Matematikkskår for elever i 10.–90. prosentilene for 8. trinn i 2003 og 2015.

Gjennomsnittsverdiene for alle de uavhengige variablene som er inkludert i analysen, er gjengitt i vedlegg 1 (se kolonner \bar{X}_{03} og \bar{X}_{15}), og for de fleste variablene er endringene små, men likevel statistisk signifikante. Det er verdt å merke seg at andel elever med en innvandringsbakgrunn er nesten doblet i perioden (6 % vs. 11 %), og at antall bøker i hjemmet rapporteres å være lavere. Dette er endringer i en retning som vil gi en forventning om nedgang i matematikkprestasjoner fra

2003 til 2015. Elevene oppgir ellers å ha noe høyere selvtillit og verdsetting av matematikkfaget i 2015, og de liker seg bedre på skolen. En langt større andel av lærerne er kvinner i 2015 (59 %) sammenlignet med 2003 (36 %), og lærerne i dag har mindre erfaring (18 vs. 12 år) og noe lengre utdanning enn i 2003. En betydelig større andel av lærerne har nå minst 60 studiepoeng i matematikk (38 % vs. 69 %) eller matematikdidaktikk (3 % vs. 23 %), og lærerne rapporterer om høyere akademiske ambisjoner for seg selv og foreldrene enn før. Dette er i seg selv interessante endringer, men om dette er endringer som bidrar til å forstå framgangen i elevenes matematikkprestasjoner, må undersøkes. Lærerne rapporterer også om mindre bruk av kalkulator, at de bruker åpne problemer og gir lekser litt oftere, og at de føler seg tryggere på skolen enn før.

Resultatene fra de to regresjonsanalysene for 2003 og 2015 er også inkludert i tabellen i vedlegg 1. Modellen forklarer henholdsvis 41 % og 45 % av variansen i matematikkskår for de to årene 2003 og 2015. Dette er betydelige andeler forklart varians, men gitt det store antallet variabler som er inkludert, er dette helt som forventet. De predikerte matematikkskårene fra regresjonsmodellene er i tråd med de observerte (henholdsvis 462 og 487 i 2003 og 2015). Den relativt høye andelen forklart varians og nærheten mellom predikerte og observerte matematikkskårer indikerer at modellen fungerer godt. Generelt er det korrelasjoner mellom de uavhengige variablene, og dette bidrar til at estimatene for regresjonskoeffisientene har relativt store standardfeil.³⁵

Ikke uventet er regresjonskoeffisientene for flere av variablene for elevenes bakgrunn statistisk signifikante i begge analysene (2003 og 2015; se kolonner β_{03} og β_{15} i vedlegg 1). Det gjelder særlig innvandringsbakgrunn (negativ sammenheng) og antall bøker hjemme (positiv sammenheng). Vi ser også at elevenes selvbilde står fram med en betydelig positiv effekt, mens elevenes kjønn har liten betydning i modellene. Det er kun noen få av lærervariablene som gir signifikante koeffisienter både i 2003 og 2015. Lærernes vurdering av foreldrenes ambisjoner, samt et enkelt spørsmål om lærerne føler seg trygge på skolen, framstår med de største og positive bidragene av lærervariablene.

Når det gjelder endringer i regresjonskoeffisienter over tid, er det kun noen få som er statistisk signifikante. (Se vedlegg 1, kolonne *Koeffisientleddet*.) Samtidig som at elevene rapporterer om høyere verdier for noen av de affektive faktorene i 2015, svekkes effektene av disse variablene sammenlignet med modellen fra 2003. Dette gjelder spesielt selvbildet i matematikk og det at eleven liker seg på skolen. Blant lærervariablene ser vi at betydningen av etterutdanning endrer seg i

35. I teksten bruker vi *effekt*, *bidrag* og *betydningen av* som synonymt med *regresjonskoeffisient*.

ulik retning for de to ulike komponentene som inngår i modellen (generell og spesifikk matematikkrelatert etterutdanning). Vi ser også at effekten av antall elever i klassen endrer seg fra å være ubetydelig i 2003 til en moderat positiv effekt i 2015. Når vi i det følgende presenterer hovedresultatene fra OBD-analysen, er det likevel grunn til å være forsiktig i tolkningen av koeffisientleddene. Samvariasjonene mellom den uavhengige og den avhengige variabelen kan endre seg i en slik modell fordi de inngår i en kompleks sammenheng med andre inkluderte eller ikke-inkluderte variabler.³⁶

TABELL 6.1 Resultater fra Oaxaca-Blinder-dekomponeringen for grupper av variabler. Bokstavene [A] til [E] referer til vedlegg 1 for å vise hvilke variabler som inngår i de ulike kategoriene

Oaxaca-Blinder-dekomponeringen			
	Gjennomsnittsled	Koeffisientledd	Interaksjonsledd
	$(\bar{X}_{15} - \bar{X}_{03}) \cdot \beta_{03}$	$(\beta_{15} - \beta_{03}) \cdot \bar{X}_{03}$	$(\bar{X}_{15} - \bar{X}_{03}) \cdot (\beta_{15} - \beta_{03})$
<i>Forskjell i konstant 2015–2003</i>		29,40	
[A] Elevbakgrunn	-5,05	2,63	0,18
[B] Elevenes affektive faktorer	11,04	-15,73	-2,33
[C] Lærerbakgrunn	-3,17	-20,62	-0,43
[D] Skolemiljø	1,08	28,48	1,75
[E] Lærernes undervisningspraksis	-1,03	1,06	-1,58
Sum av endringer per ledd	2,87	-4,18	-2,42
Totalsum av endringer inkludert forskjell i konstant			25,66

Tabell 6.1 oppsummerer resultatene for dekomponeringen av endringen i gjennomsnittlig matematikkprestasjon for norske elever på 8. trinn fra 2003 til 2015. Her er komponenter knyttet til de uavhengige variablene summert i noen overordnede kategorier. En fullstendig analyse og en oversikt over hva som inngår i hver av kategoriene, finnes i vedlegg 1.

36. Av samme grunn inkluderer vi i det videre ikke eksplisitte tolkninger av interaksjonsleddene. De er imidlertid inkludert for å få renere estimater for de to andre leddene. Størrelsen av interaksjonsleddene er imidlertid oftest ubetydelige sammenlignet med de to andre.

Det første man bør notere seg, er den store forskjellen i konstantleddene (interceptene) mellom 2003 og 2015. Forskjellen på 29,40 poeng er faktisk større enn den totale endringen i matematikkskår fra 2003 til 2015 (som er 25,66 poeng). Vi kommer tilbake til dette i diskusjonen. Når det gjelder gjennomsnittsledet ser det ved første øyekast ut som om kun 2,87 av 25,66 poeng, altså bare 11 % av endringen i elevenes gjennomsnittsskår i matematikk, kan knyttes til endringer i gjennomsnittsverdiene for de uavhengige variablene. Men gjennomsnittsledene for de enkelte faktorene varierer mye og går både i negative og positive retninger (fra $-5,05$ til $11,04$). Det betyr at endringer som bidrar positivt til elevenes matematikkskår, må først kompensere for endringer som bidrar negativt før de faktisk kan bidra til å forklare framgangen. Det er derfor lett å undervurdere de positive bidragene, men totalt sett betyr dette at modellen er ufullstendig, og at det finnes andre viktige variabler som ikke er inkludert.

Det gjennomsnittsledet som står fram med størst positivt bidrag, er knyttet til *Affektive faktorer*. Isolert sett er dette en komponent som tilsvarer 43 % ($11,04$ av $25,66$ poeng) av den totale økningen i matematikkprestasjonene. Fra tabellen i vedlegg 1 kommer det fram at dette leddet nesten utelukkende er knyttet til en beskjeden økning i elevenes selvbilde i faget. Men fordi effekten for denne variabelen er så stor, vil selv en slik liten endring i gjennomsnittet for variabelen gi et stort bidrag. De to andre affektive variablene (*Liker å være på skolen* og *Verdsetting av matematikk*) bidrar med signifikante, men bare små positive eller negative effekter. Gjennomsnittsledet med størst negativ endring er elevenes bakgrunn ($-5,05$ poeng). Som forventet gir nedgangen i *Bøker hjemme* og økningen i andelen elever med *Innvandringsbakgrunn* negative statistisk signifikante, men små bidrag.

Vi ser at koeffisientleddene har gjennomgående enda større utslag enn gjennomsnittsledene, både totalt og innenfor flere av samlekategoriene som vi har valgt å organisere resultatene etter. Koeffisientleddene går i begge retninger fra $-20,62$ til $28,48$. Samlekategorien *Skolemiljø* gir det største positive bidraget. Ser vi nærmere på resultatene for enkeltvariablene som inngår i denne kategorien, er det imidlertid få av koeffisientleddene som hver for seg er statistisk signifikante, men alle bidrar i positiv retning, noe som gjør at denne delsummen blir relativt stor og positiv. Det eneste enkeltleddet med et signifikant positivt bidrag er endringen i regresjonskoeffisienten for antall elever i klassen.

I tillegg legger vi merke til at det er store negative koeffisientledd for elevenes *Affektive faktorer* ($-15,73$) og *Lærerbakgrunn* ($-20,62$), noe som altså betyr at sammenhengene med prestasjoner for noen av variablene som inngår i disse kategoriene, er svekket fra 2003 til 2015. To av de tre enkeltvariablene innen de affektive faktorene har signifikante og negative koeffisientledd (elevenes *Selvbilde i*

matematikk og Liker å være på skolen). Til tross for at *Lærebakgrunn* med sine mange enkeltvariabler samlet framstår med et negativt koeffisientledd, er dette primært drevet av et stort negativt, men ikke-signifikant bidrag knyttet til lærernes utdanningsnivå. Dette illustrerer godt at standardfeilene for en del av regresjonskoeffisientene i modellen er relativt store, og at koeffisientleddene derfor generelt må oppfattes som lite robuste.

DISKUSJON

I dette kapitlet har vi forsøkt å kaste lys over den relativt store og positive endringen i norske elevers matematikkprestasjoner i 8. trinn slik dette måles i TIMSS-undersøkelsen. Gjennomsnittlig matematikkskår var 25,5 poeng høyere i 2015 enn i 2003, en økning som tilsvarer forskjellen mellom norske 8. og 9. trinns elever i 2015 (Bergem, 2016). Endringen er derfor betydelig, og de norske elevenes framgang for dette trinnet er blant de aller største internasjonalt (Mullis, Martin, Foy & Hooper, 2016).

Analysene i dette kapitlet tok utgangspunkt i to uavhengige regresjonsanalyser hvor den samme modellen med 22 enkeltvariabler ble anvendt på data fra hvert av de to årene. Analysene viste at gjennomsnittene for langt de fleste variablene som ble inkludert i modellen, hadde en signifikant endring fra 2003 til 2015, og i stor grad med en retning som skulle tilsi at kvaliteten på opplæringen er bedre i 2015 enn den var i 2003: Elevene verdsetter matematikk høyere og har et bedre selv-bilde knyttet til matematikkfaget, elevene liker seg bedre på skolen, og lærerne føler i enda sterkere grad at de er trygge på skolen; lærere har høyere formell kompetanse med mer spesialisering i matematikk og matematikdidaktikk, og lærerne rapporterer at de selv, elevene og foreldrene har høyere akademiske ambisjoner. Samtidig er det flere elever med en innvandringsbakgrunn, og ifølge spørsmålet om antall bøker i hjemmet er det flere elever med en relativt lavere sosioøkonomisk hjemmebakgrunn. Hver for seg og samlet skulle disse siste endringene tilsi at matematikkprestasjonene i Norge burde synke i perioden.³⁷ Økningen i norske elevers matematikkprestasjoner har derfor skjedd til tross for disse gjennomsnittlige endringene i hjemmebakgrunn. Dette kan i seg selv tolkes som at kvaliteten på opplæringen i norsk skole har økt enda mer mellom 2003 og 2015 enn det som direkte kommer fram av endringen i matematikkskår. (Se for øvrig kapittel 7 om likeverd for en mer spesifikk analyse av forskjellene som kan knyttes til elevenes

37. Så kan det innvendes at spørsmålet om bøker hjemme sannsynligvis ikke fanger inn sosioøkonomisk status på samme måte i 2015 hvor papir i stor grad er erstattet av elektroniske medier. Imidlertid skal man legge merke til at variabelen er like sterkt relatert til prestasjoner i 2015 som i 2003, gitt at alt annet i modellen er konstant.

hjemmebakgrunn.) I tillegg ser vi at modellen også inkluderer et stort antall variabler med relativt store, og spesielt for elevvariablene, også statistisk signifikante positive og negative regresjonskoeffisienter. Modellene «forklarer» hver for seg mer enn 40 % av variansen i elevskår og er derfor i utgangspunktet en substansielt relevant modell.

Oaxaca-Blinder-dekomponering bruker resultatene fra disse to uavhengige regresjonsanalysene. Dette gir en løsning hvor den gjennomsnittlige forskjellen i matematikkprestasjoner blir delt opp i ulike komponenter: én komponent knyttet til endringer i gjennomsnittsverdiene for de uavhengige variablene som er forventet å forklare den norske oppturen, én komponent knyttet til endringer i styrken av effektene/regresjonskoeffisientene, og en siste komponent knyttet til interaksjonen mellom disse to endringene.

Vår analyse kan tolkes som å støtte tidligere funn om at den norske framgangen i deler av perioden kan knyttes til en sterkere vektlegging av akademiske ambisjoner (Gustafsson & Nilsen, 2016; Nilsen et al., 2013). Det er rimelig å hevde at dette gjenspeiler at faglige utbytter av skolegangen er løftet fram i de senere årenes politikktutforming i norsk skole, og vår analyse gir ytterligere støtte til å hevde at denne vektleggingen er en viktig årsak til den faglige framgangen. Denne utviklingen har skjedd samtidig med en positiv utvikling for affektive faktorer i den samme perioden, og da særlig for elevenes selvbylde i faget. Denne kombinasjonen, at en sterkere vektlegging av faglige prestasjoner fra foreldre, lærere og skoler kan skje samtidig med et løft i faglig selvbylde, er et viktig resultat. Faglig utvikling skjer i positivt samspill med enkeltelevers utvikling i motivasjon, selvbylde og mestringsopplevelse (Elliot, Dweck, & Yeager, 2017; Petersen & Hyde, 2017; Snow, 1989).

Til tross for disse enkeltstående resultatene i vår analyse er det riktig å konkludere med at analysen ikke ga et tydelig og enkelt tolkbart bilde. I modellen er det store utslag i begge retninger knyttet til endringer i regresjonskoeffisientene. Men som vi allerede har nevnt flere ganger, er koeffisientleddene vanskelig å tolke, både fordi de har relativt store standardfeil og ikke minst fordi relasjonen mellom to variabler i stor grad også påvirkes av andre variabler som ikke er inkludert. Et eksempel er det statistisk signifikante koeffisientleddet for klassestørrelse. I 2003 var sammenhengen mellom klassestørrelse og elevprestasjoner ubetydelig, men svakt negativt, mens sammenhengen var positiv og statistisk signifikant i 2015. Det er vanskelig å se at den eventuelt direkte *kausale* effekten av klassestørrelse skal ha endret seg mellom de to årene. En mer sannsynlig tolkning er at sammenhengen mellom klassestørrelse og prestasjoner er kompleks, og at denne sammenhengen er knyttet til en rekke andre karakteristikk som også er endret mellom

disse to årene – og som det ikke har vært mulig å inkludere i modellen. Indirekte fanger derfor disse koeffisientleddene også opp det som kan betegnes som uforklarte ledd i dekomponeringen.

Det er to hovedfunn som står fram i en overordnet tolkning av resultatene fra OBD-analysen: En første viktig observasjon er at de få merkbare gjennomsnittsleddene «viskes ut» av tilsvarende store (eller større) negative koeffisientledd. Et eksempel er gjennomsnitt- og koeffisientleddene for elevenes selvbylde. I gjennomsnitt rapporterer elevene om et bedre selvbylde, noe som kan knyttes til en økning i matematikkskår på 11 poeng. Samtidig svekkes sammenhengen fra 2003 til 2015, noe som isolert sett tilsier en reduksjon av matematikkskåren på 12 poeng. Videre kan vi observere at de to regresjonsmodellene predikerer en forskjell på 25,6 poeng mellom 2003 og 2015. Denne forskjellen kan faktisk forklares ved hjelp av forskjellen i konstantleddet i regresjonsmodellene alene. Den direkte tolkningen av dette er at selv om gjennomsnittsverdiene og regresjonskoeffisientene til variablene som inngår i modellen endrer seg over tid, så gjenstår hele forskjellen på 25,6 poeng når alle variablene i regresjonsmodellen har «sagt sitt». Dette er en ytterligere forsterkning av at koeffisientleddene også gjenspeiler endringer i variabler som ikke er inkludert i analysen.

I tillegg bør det pekes på at den ordinære regresjonsanalysen som vi brukte, antar en lineær relasjon mellom de uavhengige variablene og prestasjoner i matematikk. Nå kan det med gode grunner betviles om dette er en forutsetning som gjelder for alle de inkluderte variablene. Analyser basert på instrumenter som forsøker å fange inn frekvensen av ulike undervisningsaktiviteter (brukt som indikatorer for undervisningskvalitet), viser for eksempel ofte at det ikke nødvendigvis er best å ha mer av alt. Relasjonen mellom undervisningsaktiviteter og prestasjoner kan gjerne være ikke-lineær: Det er bra med noe, men effekten avtar eller kan til og med vende seg til det negative hvis det blir for mye. Et eksempel på en aktuell analyse av norske data er Teig, Scherer, og Nilsen (2018), som viser at sammenhengen mellom frekvensen av utforskende arbeidsmåter og prestasjoner i PISA er kurvelineær.

Det er derfor viktig å framheve at modellen som det har vært mulig å utvikle med de tilgjengelige dataene, er begrenset. Hovedutfordring når man skal se på endringer over tid, er at instrumentene er lite stabile fra en syklus til den neste, og da spesielt for de variablene som kan oppfattes å være mer direkte mål på opplæringens kvalitet. Et eksempel som godt illustrerer denne begrensningen i de tilgjengelige dataene, er de omfattende målene av både kvalitet og innhold i undervisningen som er inkludert i TIMSS-undersøkelsene. I hver av studiene har både lærere og elever blitt bedt om å angi i hvilken grad det har blitt undervist i spesi-

fikke temaer/emner i matematikk, og de har også blitt bedt om å beskrive hyppigheten av bestemte undervisningsmetoder. En stor andel av spørreskjemaene er viet disse to temaene, noe som reflekterer at undervisningskvalitet og læringsmuligheter (*opportunity to learn*) er to av de mest sentrale begrepene i det teoretiske rammeverket som ligger til grunn for TIMSS-studiene (Mullis & Martin, 2013). Det finnes imidlertid ingen spørsmål innen disse to hovedtemaene som var identiske i spørreskjemaene i 2003 og 2015. Det er alltid avveininger mellom på den ene siden å holde spørsmålene uendret og på den andre siden inkludere nye spørsmål for å fange inn fenomener med aktualitet og relevans (både for forskning og politikkutforming). Konsekvensen er at vi i vår modell i all hovedsak har inkludert variabler som reflekterer indirekte eller bakenforliggende forutsetninger for gode læringsresultater. De eneste variablene som i noen grad kan knyttes til undervisningen, er noen enkle og smale spørsmål om bruk av kalkulator og lekser. Ingen av disse variablene fanger inn sentrale forhold ved undervisningskvalitet, slik dette er definert i forskningslitteraturen (se for eksempel Nilsen & Gustafsson, 2016).

Modellen som er analysert, må derfor forstås som en åpen og empirisk analyse basert på pragmatiske valg – gitt de variablene som er tilgjengelige. Dette kan imidlertid ikke tolkes som å underminere det faktum at norske elever har hatt en bemerkelsesverdig framgang i matematikkprestasjoner i perioden, og at flere variabler som beskriver kvaliteten på opplæringen i norsk skole, har endret seg til det positive mellom 2003 og 2015. Det betyr bare at vi ikke kan finne støtte for en kausal hypotese om hva denne endringen kan skyldes. Gitt at informasjonen om utvikling over tid er et av de viktigste bidragene fra de internasjonale undersøkelsene, vil vi likevel anbefale at man i framtiden har en noe mer konservativ holdning spesielt når det gjelder de omfattende og svært sentrale komponentene i spørreskjemaene knyttet til undervisningens innhold og kvalitet.

VEDLEGG 1

Vedlegget gir mer detaljer om analysen som er gjort:

- Variablene og svarskalaene er oppgitt i kolonnen lengst til venstre. Bokstavene i klammer, viser til tabell 6.1 i teksten, hvor resultatene i analysen er summert i noen samlende kategorier. Kategoriene [A] «Elevbakgrunn» og [B] «Elevenes affektive faktorer» er basert på spørsmål fra elevspørreskjemaet, mens kategoriene [C] «Lærerbakgrunn», [D] «Skolemiljø» og [E] «Lærernes undervisningspraksis» er basert på spørsmål i lærerspørreskjemaet.
- De fire kolonnene som gir detaljer fra regresjonsanalysen, oppgir gjennomsnittsverdier og regresjonskoeffisienter for de to uavhengige analysene som er gjort på det norske datasettet fra henholdsvis 2003 og 2015. Alle de 22 variablene ble inkludert samtidig. Alle gjennomsnittsverdier (\bar{X}_{03} og \bar{X}_{15}) er statistisk signifikant forskjellig fra 0, og dette er derfor ikke markert i tabellen. Regresjonskoeffisienter (β_{03} og β_{15}) som er statistisk signifikant forskjellig fra 0, er markert med en asterisk (*). I tillegg gir de nederste radene i tabellen henholdsvis observert og predikert skår, samt modellenes totale forklarte varians (R^2).
- De tre kolonnene til høyre gir de detaljerte dekomponeringene for hver variabel. Dersom endringen i gjennomsnittet er statistisk signifikant, er dette markert i kolonnen for gjennomsnittsleddet med ^a. Dersom endringen i regresjonskoeffisienten er statistisk signifikant, er dette markert i kolonnen for koeffisientleddet med ^b. Dersom begge faktorene som inngår i de ulike leddene er statistisk signifikante, har vi markert dette ved å utheve verdien med fete typer. Siden hvert av disse leddene er et produkt av to faktorer, har vi gjort en enkel vurdering av statistisk signifikans av OBD-komponentene ved å kreve at begge faktorene må være statistisk signifikante. Denne signifikansvurderingen er derfor basert på en noe dristig antakelse om at de to faktorene som inngår i de ulike produktene, er uavhengige av hverandre. Det er derfor noe usikkerhet knyttet til disse signifikansvurderingene.

	Regresjonsanalysen				Oaxaca-Blinder-dekomponeringen		
	TIMSS 2003		TIMSS 2015		Gjennomsnittsleddet	Koeffisientleddet	Interaksjonsleddet
	\bar{X}_{03}	β_{03}	\bar{X}_{15}	β_{15}	$(\bar{X}_{15} - \bar{X}_{03}) \cdot \beta_{03}$	$(\beta_{15} - \beta_{03}) \cdot \bar{X}_{03}$	$(\bar{X}_{15} - \bar{X}_{03}) \cdot (\beta_{15} - \beta_{03})$
Variabler							
Konstant		295,69		325,08		29,40	
[A] Elevers kjønn (1=kvinne)	0,50	7,85*	0,50	4,52*	0,03	-1,65	-0,01
[A] Antall bøker hjemme (0=0–10 bøker, 4=Mer enn 200 bøker)	2,52	10,20*	2,13	10,73*	-3,99 ^a	1,35	-0,21
[A] Elevers alder (0=laveste alder i normalkohort)	0,47	-0,73	0,47	4,32	0,00	2,38	0,00
[A] Innvandringsbakgrunn (1= begge foreldre født i et annet land enn Norge)	0,06	-23,88*	0,11	-15,14*	-1,09 ^a	0,56	0,40
[B] Selvbilde matematikk (gj. snitt fire spørsmål, 0=svært uenig, 3=svært enig)	1,77	51,77*	1,98	44,79*	10,91 ^a	-12,34 ^b	-1,47
[B] Verdssetting av matematikk (gj. snitt fire spørsmål, 0=svært uenig, 3=svært enig)	1,98	-3,85	2,18	-1,48	-0,77 ^a	4,70	0,48
[B] Liker å være på skolen (0=svært uenig, 4=svært enig)	1,94	2,82*	2,26	-1,35	0,91 ^a	-8,08 ^b	-1,34
[C] Lærers kjønn (1=kvinne)	0,36	-0,34	0,59	-2,17	-0,08 ^a	-0,65	-0,43
[C] Antall år som lærer	18,12	0,25	12,09	0,04	-1,51 ^a	-3,86	1,28
[C] Lærers utdanningsnivå (0=ikke fullført vgs., 6= minst mastergrad)	3,78	1,96	4,19	-2,72	0,80 ^a	-17,68	-1,91
[C] Matematikk i lærerutdanningen (1=60 studiepoeng eller mer)	0,38	-2,99	0,69	-0,37	-0,92 ^a	1,00	0,80
[C] Matematikdidaktikk i lærerutdanningen (1=60 studiepoeng eller mer)	0,03	-8,60	0,23	-6,10	-1,74 ^a	0,07	0,51

	Regresjonsanalysen		Oaxaca-Blinder-dekomponeringen				
	TIMSS 2003	TIMSS 2015	Gjennomsnitts-leddet	Koeffisient-leddet	Interaksjons-leddet		
[C] Deltaelse i generell matematikkrelatert etterutdanning siste to år (gj.snitt tre spørsmål: 1=ja)	0,30	4,09	0,28	-9,00	-0,10	-3,98 ^b	0,31
[C] Deltaelse i spesifikke matematikkrelatert etterutdanning siste to år (gj.snitt tre spørsmål: 1=ja)	0,19	-8,67	0,15	14,52*	0,37	4,47 ^b	-0,99
[D] Lærer føler seg trygg på skolen (0=svært uenig, 4=svært enig)	2,75	7,81*	2,94	8,18*	1,45 ^a	1,04	0,07
[D] Lærerne har høye akademiske ambisjoner (gj.snitt tre spørsmål: 0=veldig lav, 4=veldig høy)	2,57	-4,37	3,02	0,05	-1,97 ^a	11,36	1,99
[D] Foreldre/elever har høye akademiske ambisjoner (gj.snitt tre spørsmål: 0=veldig lav, 4=veldig høy)	2,24	9,11*	2,40	12,52*	1,46a	7,64	0,55
[D] Antall elever i klassen (antall elever over 15)	10,42	-0,13	9,36	0,68*	0,14 ^a	8,44 ^b	-0,86
[E] Lærer gir elevene åpne problemer (0=aldri, 3=i alle eller nesten alle timene)	1,03	2,70	1,11	3,59	0,22 ^a	0,91	0,07
[E] Ubegrenset bruk av kalkulator (1=ja)	0,72	1,97	0,12	4,60	-1,18 ^a	1,89	-1,57
[E] Variert bruk av kalkulator (gj.snitt fire spørsmål: 0=aldri, 3=hver/nesten hver time)	2,18	0,44	1,72	0,50	-0,20 ^a	0,13	-0,03
[E] Lærer gir lekser (1=Ja)	0,93	4,44	0,96	2,44	0,12 ^a	-1,87	-0,05
Observert matematikkskår		461,5		487,0			
Predikert matematikkskår		461,5		487,1			
R²		0,41		0,45			

REFERANSER

- Ammermueller, Andreas. (2007a). PISA: What makes the difference? *Empirical Economics*, 33(2), 263–287. doi: [10.1007/s00181-006-0102-5](https://doi.org/10.1007/s00181-006-0102-5)
- Ammermueller, Andreas. (2007b). Poor Background or Low Returns? Why Immigrant Students in Germany Perform so Poorly in the Programme for International Student Assessment. *Education Economics*, 15(2), 215–230. doi: [10.1080/09645290701263161](https://doi.org/10.1080/09645290701263161)
- Barrera-Osorio, Felipe, Garcia-Moreno, Vicente, Patrinos, Harry Anthony, & Porta, Emilio. (2011). Using the Oaxaca-Blinder Decomposition Technique to Analyze Learning Outcomes Changes over Time. An Application to Indonesia's Results in PISA Mathematics. *Policy Research Working Paper*(5584). doi: <https://doi.org/10.1596/1813-9450-5584>
- Bergem, Ole Kristian. (2016). Hovedresultater i matematikk. I Ole Kristian Bergem, Hege Kaarstein & Trude Nilsen (Red.), *Vi kan lykkes i realfag. Resultater og analyser fra TIMSS 2015* (s. 22–43). Oslo: Universitetsforlaget.
- Blinder, Alan S. (1973). Wage Discrimination: Reduced Form and Structural Estimates. *The Journal of Human Resources*, 8(4), 436–455. doi: [10.2307/144855](https://doi.org/10.2307/144855)
- Cattaneo, Maria Alejandra, & Wolter, Stefan C. (2012). *Migration Policy Can Boost PISA Results: Findings from a Natural Experiment*. Aarau: Institute for the Study of Labor (IZA).
- Cohen, Jacob, Cohen, Patricia, West, Stephen G., & Aiken, Leona S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd utg.). Mahwah: Lawrence Erlbaum Associates, Publishers.
- Ehmke, Timo, Klieme, Eckhard, & Stanat, Petra. (2013). Veränderungen der Lesekompetenz von PISA 2000 nach PISA 2009. Die Rolle von Unterschieden in den Bildungswegen und in der Zusammensetzung der Schülerschaft. I N. Jude & Eckhard Klieme (Red.), *PISA 2009 – Impulse für die Schul- und Unterrichtsforschung* (Vol. 59, s. 132–150). Weinheim: Beltz.
- Elder, Todd E., Goddeeris, John H., & Haider, Steven J. (2010). Unexplained gaps and Oaxaca-Blinder decompositions. *Labour Economics*, 17(1), 284–290. doi: <http://dx.doi.org/10.1016/j.labeco.2009.11.002>
- Elliot, A. J., Dweck, C. S., & Yeager, D. S. (2017). *Handbook of Competence and Motivation: Theory and Application*. New York: Guilford Publications.
- Fortin, Nicole, Lemieux, Thomas, & Firpo, Sergio. (2011). Decomposition Methods in Economics. I Orley Ashenfelter & David Card (Red.), *Handbook of Labor Economics* (Vol. 4A, s. 1–102). Oxford: Elsevier.
- Foy, P. (2017). *TIMSS 2015 User Guide for the International Database*. Boston: TIMSS & PIRLS International Study Center.
- Freitas, Pedro, Nunes, Luís Catela, Balcão Reis, Ana, Seabra, Carmo, & Ferro, Adriana. (2016). Correcting for sample problems in PISA and the improvement in Portuguese students' performance. *Assessment in Education: Principles, Policy & Practice*, 23(4), 456–472. doi: [10.1080/0969594X.2015.1105784](https://doi.org/10.1080/0969594X.2015.1105784)
- Fullan, Michael. (2007). *The new meaning of educational change*. London New York: Routledge Teachers College Press.
- Gustafsson, Jan Eric, & Nilsen, Trude. (2016). The Impact of School Climate and Teacher Quality on Mathematics Achievement: A Difference-in-Differences Approach. I Trude Nilsen &

- Jan-Eric Gustafsson (Red.), *Teacher Quality, Instructional Quality and Student Outcomes: Relationships Across Countries, Cohorts and Time* (s. 81–95). Cham: Springer International Publishing.
- Haug, Peder. (2003). *Evaluering av Reform 97. Sluttrapport frå styret for Program for evaluering av Reform 97*. Oslo: Norges Forskningsråd.
- Jann, B. (2008). The Blinder-Oaxaca decomposition for linear regression models. *Stata Journal*, 8(4), 453–479.
- Kjærnsli, Marit, & Jensen, Fredrik. (2016). PISA 2015 – Gjennomføring og noen sentrale resultater. I Marit Kjærnsli & Fredrik Jensen (Red.), *Stø kurs. Norske elevers kompetanse i naturfag, matematikk og lesing i PISA 2015* (s. 11–31). Oslo: Universitetsforlaget.
- Liao, Pei-An, Chang, Hung-Hao, Wang, Jiun-Hao, & Sun, Lih-Chyun. (2016). What are the determinants of rural-urban digital inequality among schoolchildren in Taiwan? Insights from Blinder-Oaxaca decomposition. *Computers & Education*, 95, 123–133. doi: <http://dx.doi.org/10.1016/j.compedu.2016.01.002>
- Lock, Kelly J., & Gibson, John K. (2008). Explaining Maori underachievement in standardised reading tests: The role of social and individual characteristics. *Kōtuitui: New Zealand Journal of Social Sciences Online*, 3(1), 1–13. doi: [10.1080/1177083X.2008.9522428](https://doi.org/10.1080/1177083X.2008.9522428)
- Lounkaew, Kiatanantha. (2013). Explaining urban–rural differences in educational achievement in Thailand: Evidence from PISA literacy data. *Economics of Education Review*, 37, 213–225. doi: <http://dx.doi.org/10.1016/j.econedurev.2013.09.003>
- Martins, Lurdes, & Veiga, Paula. (2010). Do inequalities in parents' education play an important role in PISA students' mathematics achievement test score disparities? *Economics of Education Review*, 29(6), 1016–1033. doi: <http://dx.doi.org/10.1016/j.econedurev.2010.05.001>
- Mullis, Ina V.S., Martin, M.O., Foy, P., & Hooper, M. (2016). *TIMSS 2015 International Results in Mathematics*. Hentet fra <http://timssandpirls.bc.edu/timss2015/international-results/download-center/>
- Mullis, Ina V.S., & Martin, Michael O. (Red.). (2013). *TIMSS 2015 Assessment Framework*. Boston: TIMSS&PIRLS International Study Center.
- Nieto, Sandra, & Ramos, Raúl. (2015). Educational outcomes and socioeconomic status: A decomposition analysis for middle-income countries. *PROSPECTS*, 45(3), 325–343. doi: [10.1007/s11125-015-9357-y](https://doi.org/10.1007/s11125-015-9357-y)
- Nilsen, Trude, Grønmo, Liv Sissel, & Hole, Arne. (2013). Læringstrykk og prestasjoner i matematikk og naturfag. I Liv Sissel Grønmo & Torgeir Onstad (Red.), *Opptur og nedtur. Analyser av TIMSS-data for Norge og Sverige* (s. 19–51). Oslo: Akademika forlag.
- Nilsen, Trude, & Gustafsson, Jan Eric (Red.). (2016). *Teacher Quality, Instructional Quality and Student Outcomes. Relationships Across Countries, Cohorts and Time*. Cham: Springer International Publishing.
- Oaxaca, Ronald. (1973). Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review*, 14(3), 693–709.
- Olsen, Rolf Vegar, Hopfenbeck, Therese Nerheim, & Lillejord, Sølvi. (2013). Elevenes lærings-situasjon etter Kunnskapsløftet. *Norsk pedagogisk tidsskrift*, 97(6), 355–369.

- Oppedisano, Veruska, & Turati, Gilberto. (2015). What are the causes of educational inequality and of its evolution over time in Europe? Evidence from PISA. *Education Economics*, 23(1), 3–24. doi: <https://doi.org/10.1080/09645292.2012.736475>
- Petersen, J. L., & Hyde, J. S. (2017). Trajectories of self-perceived math ability, utility value and interest across middle school as predictors of high school math performance. *Educational Psychology*, 37(4), 438–456.
- Ramos, Raul, Duque, Juan C., & Nieto, Sandra. (2012). Decomposing the Rural-Urban Differential in Student Achievement in Colombia Using PISA Microdata. *IZA Discussion Paper* (No. 6515).
- Sarason, Seymour B. (1990). *The predictable failure of educational reform: Can we change course before it's too late?* San Francisco: Jossey-Bass.
- Snow, R. E. (1989). Cognitive-conative aptitude interactions in learning. I R. Kanfer, P.L. Ackerman & R. Cudeck (Red.), *Abilities, motivation, and methodology* (s. 435–474). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Song, Steve, Perry, Laura B., & McConney, Andrew. (2014). Explaining the achievement gap between Indigenous and non-Indigenous students: an analysis of PISA 2009 results for Australia and New Zealand. *Educational Research and Evaluation*, 20(3), 178–198. doi: [10.1080/13803611.2014.892432](https://doi.org/10.1080/13803611.2014.892432)
- St.meld. nr. 16. (2006–2007). *...og ingen stod igjen: Tidlig innsats for livslang læring*. Oslo: Kunnskapsdepartementet. Lastet ned fra <http://www.regjeringen.no/nb/dep/kd/dok/regpubl/stmeld/2006-2007/Stmeld-nr-16-2006-2007-.html?id=441395>
- Teig, Nani, Scherer, Ronny, & Nilsen, Trude. (2018). More isn't always better: The curvilinear relationship between inquiry-based teaching and student achievement in science. *Learning and Instruction*, 56, 20–29.
- Tyack, David B., & Cuban, Larry. (1995). *Tinkering toward utopia : a century of public school reform*. Cambridge, Mass.: Harvard University Press.
- Valenzuela, Juan Pablo, Gómez Vera, Gabriela, & Sotomayor, Carmen. (2015). The role of reading engagement in improving national achievement: An analysis of Chile's 2000–2009 PISA results. *International Journal of Educational Development*, 40, 28–39. doi: <http://dx.doi.org/10.1016/j.ijedudev.2014.11.011>
- Aasen, Petter, Møller, Jorunn, Rye, Ellen, Ottesen, Eli, Prøitz, Tine Sophie, & Hertzberg, Frøydís (2012). *Kunnskapsløftet som styringsreform – et løft eller et løfte? Forvaltningsnivåenes og institusjonenes rolle i implementeringen av reformen. (Rapport 20/2012)*. Oslo: NIFU.