

Are Turmo, ILS, UiO

# Internasjonale elevundersøkelser – trender og fortolkninger

## Innledning

Ved publiseringen av resultatene fra PISA- og TIMSS-undersøkelsene som ble gjennomført i 2003, fikk man i Norge for første gang omfattende og robuste data som kan si noe om utviklingen i norske skolelevers faglige kompetanse over tid i absolutt forstand. Riktignok hadde Norsk matematikkråd allerede i en årrekke publisert sin undersøkelse av grunnleggende matematikkunnskap blant studenter som begynner på matematikkrevende studier. Undersøkelsen viste blant annet at andelen korrekte svar innenfor et tema som prosentregning hadde sunket fra over 80 prosent i 1984 til under 50 prosent i 2003 (Rasch-Halvorsen og Johnsbråten 2007). De nye trenddataene fra PISA og TIMSS som ble publisert i desember 2004, viste en sammenfallende utvikling.

I de internasjonale undersøkelsene kan man ved bruk av såkalte linkoppgaver plassere elever som har tatt faglige prøver på ulike tidspunkter på samme prestasjonsskala ved hjelp av spesielle målemetoder og statistiske teknikker. I TIMSS-undersøkelsen kunne man sammenlikne elevene som deltok i undersøkelsen i 2003, med de som deltok i 1995, mens man i PISA-undersøkelsen kunne gjøre tilsvarende sammenlikninger mellom 2003 og 2000. Denne typen sammenlikninger kan også gjøres i leseundersøkelsen PIRLS som ble gjennomført i både 2001 og 2006. Til

sammen gir disse undersøkelsene trenddata for de sentrale fagområdene lesing, matematikk og naturfag. Definisjonene av lesekompetanse i PISA og PIRLS er rimelig sammenfallende, mens kompetanse i matematikk og naturfag i PISA og TIMSS er definert på litt forskjellig måte. Dette kommer vi tilbake til. Samlet dekker de tre undersøkelsene 4., 8. og 10. klassetrinn. Det er for øvrig et mål at tilsvarende metoder for å måle trender i fagkompetanse på sikt også skal innføres når det gjelder de nasjonale prøvene (se St.meld. nr. 31, 2007-2008).

Mange kritikere av de internasjonale undersøkelsene trekker fram problemer knyttet til det å sammenlikne de faglige prestasjonene til elever fra svært ulike kulturer og skolesystemer. Det hevdes også at de internasjonale rangeringslistene som undersøkelsene produserer, lett kan misbrukes og misforstås, ikke minst av media (for eksempel Sjøberg 2007). Når vi fokuserer på trenddataene fra undersøkelsene, sammenlikner vi derimot elever i samme kultur og skolesystem på to ulike tidspunkter. I den grad man oppfatter det som måles som relevant i det aktuelle skolesystemet, kan denne typen sammenlikninger gi svært relevant og viktig informasjon. Både PIRLS og PISA, og også i stor grad TIMSS, måler kompetanser som er en viktig del av det norske skole gjennom læreplanverket er forpliktet til å søke å utvikle hos elevene.

Men undersøkelsene dekker selvsagt ikke hele bredden av de målsettinger som er satt for norsk skole.

Dette kapitlet vil først gi en kortfattet framstilling av det metodiske grunnlaget for å kunne måle utvikling i faglig kompetanse over tid i absolutt forstand. Deretter gis en oversikt over og diskusjon av de trendene i faglig kompetanse som er påvist i perioden 1995-2007. Endelig gis en syntese av de forskningsbaserte fortolkningene av trendene i fagkompetanse som er blitt presentert.

### Det metodiske grunnlaget

Grunnlaget for at man kan si noe om utvikling av kompetanse over tid i absolutt forstand, er den testteoretiske metoden som på engelsk kalles «test equating». Det dreier seg her om en metode for å sammenlikne resultater fra to forskjellige tester. Test equating forkortes i det følgende av og til som TE. TE-metoder brukes ved mange typer standardiserte tester innenfor utdanning og psykologi for å sikre at resultater fra ulike tester kan sammenliknes direkte. I de siste årene har forskere innenfor utdanning, psykologi og statistikk bidratt til en rask vekst i statistiske og psykometriske metoder som kan brukes for TE (Kolen og Brennan 2004). Med psykometri menes vitenskapen om hvordan man kan måle psykologiske begreper på en mest mulig pålitelig og gyldig måte.

Å sammenholde informasjon fra tester som er tatt på ulike tidspunkter, innebærer åpenbare utfordringer. Hvis man for eksempel gir de samme oppgavene til elever som tar testen på ulike tidspunkter, er det en mulighet for at elevene kan informere hverandre. Tilsvarende risikerer man at en elev som tar testen på to tidspunkter, vil få samme oppgaver to ganger. I slike situasjoner kan en test i større grad bli et mål på hvor god eleven er til å løse akkurat de

oppgavene som ble gitt, heller enn det som oppgavene til sammen er ment å måle.

Det norske kvalitetsvurderingssystemet for grunnopplæringen baserer seg blant annet på ulike typer faglige prøver og tester: internasjonale tester, nasjonale prøver, diagnostiske prøver og eksamen (Udir 2008a). Blant disse er det først og fremst eksamen som kan ha betydelige direkte konsekvenser for enkeltelever, spesielt i videregående opplæring. I engelskspråklig litteratur omtales gjerne denne typen tester som «high stakes tests». Ikke alle elever består eksamen, og eksamenskarakterene har betydning ved opptak til flere typer studier. Det er derfor opplagt avgjørende at elevenes prestasjoner bedømmes så rettferdig som mulig. Som kjent eksisterer det et etablert system som forsøker å sikre dette i vårt land. Vi vet imidlertid at ikke alle avgangselever søker seg direkte videre til høyere utdanning. Det er derimot relativt vanlig å vente ett eller flere år etter videregående skole før man søker opptak. Ved hvert års opptak til høyere utdanning vil vi derfor ha søkere som har tatt eksamen på ulike tidspunkter. Vi skal se nærmere på et eksempel i det følgende.

Nina tok eksamen i matematikk på VK2 i 2005 og fikk karakteren 4, mens Asle tok eksamen i 2006 og fikk samme karakter. Nina og Asle fikk selvsagt forskjellige eksamenssett. Er vi da sikre på at Nina og Asle hadde tilsvarende kompetanse i matematikk (slik den er definert i læreplanens kompetansemål) ved utgangen av videregående opplæring? Vi kan selvsagt relatere de to elevenes prestasjoner på eksamen til fordelingen for alle elevene som tok eksamen de respektive to årene. Dette kalles gjerne grupperelatering. Men hvilket grunnlag har vi for å påstå at de to årskullene elever er gjennomsnittlig like dyktige i matematikk? Nina venter ett år med å søke seg videre til høyere utdanning, mens

Asle søker direkte. De konkurrerer da med samme karakter i matematikk, men er karakterene basert på et sammenliknbart og dermed rettferdig grunnlag? Mange land legger stor vekt på at eksamen faktisk skal være direkte sammenliknbar fra år til år, for eksempel Nederland, men i Norge har vi ingen tradisjon for dette. Det må imidlertid nevnes at eksamenskarakterene som kjent kun utgjør en liten del av grunnlaget for opptak til høyere utdanning. Sammenliknbarheten til standpunkt-karakterene kan generelt sies å være en betydelig større utfordring (se for eksempel Bonesrønning 1999).

Utfordringene er kanskje enda større i høyere utdanning. Vi kan tenke oss to studenter som tar mastergrad ved ulike institusjoner, og som senere vil søke seg videre til ph.d. ved samme lærested, der det stilles krav om karakterene A eller B fra mastergrad. Tabell 1 viser andeler studenter som fikk A eller B på mastergrad i 2005 ved tre ulike institusjoner. Tabellen synliggjør at Universitetet i Bergen gir høyest andel A eller B, mens NTNU gir klart lavest andel. Rapporten som resultatene er hentet fra (UHR 2006), påpeker imidlertid at det må tas forbehold om at det er eksamener på forskjellige nivåer som er rapportert; resultatene inneholder en blanding av 2-årige og 5-årige programmer. Like fullt indikerer resultatene ulike karaktertradisjoner ved institusjonene, noe man også har funnet i andre tilsvarende undersøkelser. Det er opplagt betenkelig hvis karakterene A og B representerer betydelig ulik kompetanse avhengig av hvilken institusjon utdanningen er tatt ved. I prinsippet har vi her samme type problem som i eksempelet fra videregående opplæring, men utfordringene på mastergradsnivå er betydelig mer utfordrende å løse, av ulike årsaker. Det kan også nevnes at Møen og Tjelta (2005) har påvist betydelige forskjeller i karakterbruk mellom høyskoler som tilbyr

økonomisk-administrativ utdanning på bachelornivå.

Eksempelene som er vist over, viser betydningen av teknikker for å kunne sammenlikne resultater fra tester tatt på ulike tidspunkter. I det følgende skal vi se nærmere på prinsippene for «test equating». Framstillingen er tilpasset fra beskrivelser gitt av Kjærnsli, Lie og Turmo (2005) og Turmo (2003).

Tabell 1. **Andel studenter som fikk karakterene A eller B på mastergrad i 2005 ved tre institusjoner (UHR 2006)**

Institusjon	Andel studenter som fikk karakteren A eller B på mastergrad i 2005
Samfunnsvitenskaplig fakultet, Universitetet i Oslo	41,6
Samfunnsvitenskaplig fakultet, Universitetet i Bergen	49,5
Samfunnsvitenskaplig fakultet, NTNU	33,8

Kilde: UHR 2006.

Som vi allerede har vært inne på, kan resultater i form av for eksempel antall riktige svar på to forskjellige prøver i utgangspunktet ikke fortelle oss noe om en eventuell fram- eller tilbakegang i et fagområde. Dette ville krevd at innholdet på prøvene i samme grad er representativt for fagområdet som helhet. Dessuten kan ikke antall riktige svar uten videre brukes som objektivt mål for prestasjonen, i og med at oppgavene kan ha forskjellig vanskelighetsgrad. Hvis en gruppe elever skårer høyere på en prøve enn på den forrige, er det i prinsippet umulig å si om det er fordi elevene er blitt dyktigere, eller om oppgavene er blitt lettere. Det viser seg dessuten i praksis svært vanskelig å forutsi hvor vanskelig en faglig oppgave er for en gruppe elever uten at man har prøvd dette ut empirisk (se Olsen, Turmo og Lie 2001). Små variasjoner i ordvalg og formulering-

ger kan påvirke vanskelighetsgraden til enkeltoppgaver betydelig.

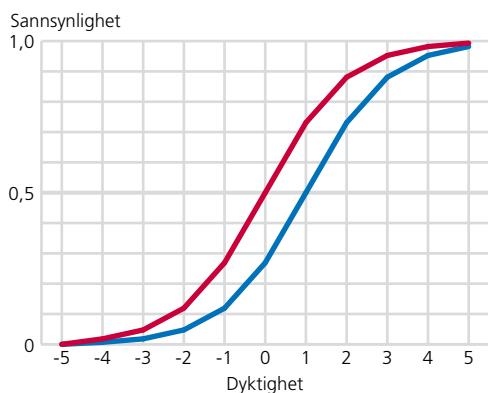
I det som gjerne kalles moderne test-teori (Item Response Theory, IRT), måles elevs dyktighet og oppgavers vanskelighetsgrad langs samme måleskala (se for eksempel Crocker og Algina 1986). Den såkalte Rasch-modellen er en variant av Item Response Theory som blant annet anvendes i PISA-undersøkelsen. I Rasch-modell bestemmer differansen mellom dyktigheten til eleven ( $d$ ) og vanskelighetsgraden ( $v$ ) til oppgaven sannsynligheten for at eleven vil løse oppgaven riktig. I Rasch-modellen er sannsynligheten for et riktig svar gitt som en matematisk funksjon av denne differansen. Fordelingen av sannsynligheter antar man følger kurver som vist på figur 1. Disse kurvene kalles karakteristiske kurver (på engelsk: Item Characteristic Curves, gjerne forkortet ICC).

I figur 1 er karakteristiske kurver for to forskjellige oppgaver tegnet inn. Figuren viser at når dyktigheten stiger, stiger også sannsynligheten for å svare riktig på oppgaven, men sammenhengen er ikke lineær. Det er også verdt å merke seg at sannsynligheten aldri blir 1, og den blir heller

aldri 0. Uansett hvor dyktig en elev er, er det alltid en liten mulighet for at eleven vil svare feil på en konkret enkeltoppgave, av ulike årsaker. På samme måte kan også en elev med lav dyktighet tenkes å svare riktig på enkeltoppgaver. I figur 1 har den «rød» oppgaven større sannsynligheter for riktig svar for alle dyktighetsnivåer enn det den «blå» har, og er følgelig en lettere oppgave. Ved at et sett av oppgaver gis til en gruppe elever, kan man bestemme hver elevs dyktighet og hver oppgaves vanskelighetsgrad. Dette medfører at hver oppgave som er brukt i en undersøkelse, ut fra resultatene får en kjent vanskelighetsgrad. Det er her bare snakk om relative vanskelighetsgrader, med andre ord hvor vanskelig oppgavene er sammenliknet med hverandre.

Poenget med «test equating» er at resultatene på to forskjellige prøver kan gis langs samme skala for dyktighet. For å få til dette er det imidlertid nødvendig at vanskelighetsgradene kan sammenliknes ved at noen av oppgavene forekommer i begge prøvene. Vanligvis gjøres dette ved at noen få oppgaver fra den første prøven holdes hemmelig for senere å brukes også i prøve nummer to. Ved hjelp av disse linkoppgavene kan vanskelighetsgradene til alle de nye oppgavene, og dermed hele den nye prøven, normeres slik at måleskalaen blir den samme som i den første prøven. Det er denne teknikken som brukes i de internasjonale undersøkelsene PISA, PIRLS og TIMSS. En annen teknikk som av og til brukes, er å la en gruppe elever ta to forskjellige tester. På denne måten kan testene kobles sammen via dyktighetene til de felles elevene. Denne metoden brukes blant annet i forbindelse med konstruksjon av eksamen i Nederland. En annen mulighet er at man fra to helt forskjellige prøver kan sette sammen en tilleggsprøve med oppgaver fra hver av dem. Ut fra de målte vanskelighetsgradene til oppgavene

Figur 1. Karakteristiske kurver for to oppgaver



Kilde: Lie mfl. 2001.

i tilleggsprøven kan så de to opprinnelige prøvene indirekte sammenliknes.

I prinsippet er «test equating» enkelt nok, men i praksis er det mange krav som må oppfylles, og problemer som må løses. Et spesielt problem oppstår når oppgavene forekommer på ulike steder i oppgaveheftene. En oppgave på slutten av et oppgavehefte får gjerne lavere løsningsfrekvens og fortøner seg dermed som vanskeligere. Åpenbare årsaker til dette kan være at elevene blir mer slitne utover i testen, eller de mister motivasjonen. Denne typen konteksteffekter er utfordrende fordi Rasch-modellen som ligger til grunn, i utgangspunktet forutsetter at oppgavene fungerer likt uavhengig av hvor i oppgaveheftene de forekommer. I de internasjonale undersøkelsene som PISA legger man derfor mye arbeid i å justere for denne typen effekter. Like fullt er dette et tema som diskuteres i stor detalj også i de faglige evalueringene av undersøkelsen som etter hvert foreligger (for eksempel i regi av Education Testing Service, ETS, i 2008).

Ufordringer knyttet til testmotivasjon ble nevnt over. I diskusjoner av resultater fra internasjonale komparative studier av læringsutbytte blir variasjoner i testmotivasjon av og til nevnt som mulig forklaringsfaktor. Enkelte hevder for eksempel at norske elever ikke er motiverte til å gjøre sitt beste på denne typen tester. I PISA-undersøkelsen er testmotivasjonen forsøkt målt ved at et såkalt «innsatstermometer» er lagt inn til slutt i oppgaveheftene. Her blir elevene blant annet bedt om å angi relativ innsats på en skala fra 1 til 10, hvor 10 skal betegne den innsatsen de ville lagt ned hvis prøven hadde hatt betydning for karakteren på skolen. Gjennomsnittsverdien for Norge er her av de laveste blant de deltakende landene, men variasjonen mellom landene er generelt liten. Det viser seg dessuten av innsatsfaktoren ikke kan

forklare noe av variasjonen i prestasjoner mellom land (OECD 2007, Kjærnsli mfl. 2007). Selv om det etter hvert finnes en del studier av testmotivasjon i forbindelse med internasjonale kunnskapstester, er dette opplagt et område hvor det er behov for mer forskning og kunnskap.

### **Trender for norske skoleelevers faglige kompetanse i perioden 1995-2007**

I det følgende skal vi studere i detalj de trenddataene som er framkommet gjennom de internasjonale studiene PISA, PIRLS og TIMSS i perioden 1995-2007. Når vi studerer trender, kan vi betrakte gruppen av elever under ett, eller vi kan se på trender for undergrupper av elever. For eksempel kan man analysere hvordan minoritetselevers prestasjoner utvikler seg over tid, og så sammenlikne med tilsvarende utvikling for majoriteten. I slike tilfeller er det selvsagt viktig at ikke gruppenes sammensetning også varierer betydelig over tid. I denne artikkelen velger vi å betrakte trender for alle elevene samlet, samt for jenter og gutter separat. Vi starter med en kort presentasjon av de tre studiene PIRLS, PISA og TIMSS.

I 1991 var The International Association for the Evaluation of Educational Achievement (IEA) ansvarlig for den første internasjonale leseundersøkelsen som Norge deltok i, en undersøkelse som omfattet 9-åringer og 14-åringer i 32 land. På slutten av 90-tallet tok IEA initiativet til en ny leseundersøkelse, PIRLS (Progress in International Reading Literacy Study). Undersøkelsen kartlegger leseforståelse blant 10-åringer og ble første gang gjennomført i 2001. PIRLS ble gjentatt i 2006, og studien vil etter planen gjennomføres hvert 5 år. I PIRLS trekkes hele klasser ut til å delta ved den enkelte skole.

PISA (Programme for International Student Assessment) er en internasjonal komparativ studie i regi av OECD. Undersøkelsen ble første gang gjennomført i 2000, og den innhenter nye data hvert tredje år. Populasjonen er 15-åringer, og studien fokuserer på kompetanser som ansees å være viktige for å kunne fungere som en aktiv og reflektert borger i morgendagens samfunn. Tre fagområder er i utgangspunktet valgt ut for kartlegging: lesing, matematikk og naturfag. I den siste undersøkelsen i 2006 var naturfag viet mest prøvetid (tilsvarende for lesing i 2000 og matematikk i 2003). Konkrete eksempler på faglige oppgaver kan finnes på [www.pisa.no](http://www.pisa.no). I PISA 2006 deltok 57 land, med en hovedvekt av OECD-land. Utvalget i PISA trekkes blant alle 15-åringer ved den enkelte skole, uavhengig av hvilken klasse elevene går i.

TIMSS (Trends in International Mathematics and Science Study) gjennomføres i regi av IEA, og studien dekker altså både matematikk og naturfag. Undersøkelsen har røtter tilbake til rundt 1970 da den første studien av realfagene i regi av IEA ble gjennomført. De faglige prøvene i TIMSS er mer læreplan-nære enn hva tilfellet er i PISA, og man undersøker to populasjoner, henholdsvis 9-åringer og 13-åringer. Konkrete eksempler på oppgaver kan finnes på [www.timss.no](http://www.timss.no). TIMSS «Trends» ble første gang gjennomført i 2003, og vil bli gjentatt hvert 4 år. Den første TIMSS-studien ble imidlertid gjennomført allerede i 1995, og Norge deltok også da. Den gangen sto forkortelsen for Third International Mathematics and Science Study. I TIMSS trekkes hele klasser ut til å delta ved den enkelte skole.

I de trendresultatene vi skal studere i det følgende, har man typisk fastsatt en prestasjonsskala på et gitt tidspunkt, og plasserer så elevene som tar tester på senere

tidspunkter på samme prestasjonsskala ved hjelp av de teknikkene som ble beskrevet innledningsvis i artikkelen. Tilkomst av nye land i studiene, for eksempel i TIMSS hvor gjennomsnittsverdier i utgangspunktet beregnes basert på alle deltagende land, påvirker derfor ikke trendvurderingene for det enkelte land.

### Trender i PIRLS: Lesekompetanse på 4. klasses-trinn

PIRLS-undersøkelsen kartlegger lesekompetanse blant elever på 4. klasses-trinn. Leseforståelse i PIRLS-studien vil si å ha evne til å forstå og bruke de skriftlige språkuttrykk som kreves i samfunnet og/eller verdsettes av den enkelte. Unge lesere bør ha evne til å skape mening fra et spekter av ulike typer tekster. De bør kunne lese for å lære, for å delta i samspill med andre, samt kunne lese for egen fornøyles skyld (Campbell mfl. 2001, van Daal mfl. 2007).

Tabell 2 viser endringene i lesekompetanse i Norge slik de framstår gjennom de to PIRLS-undersøkelsene. Den internasjonale prestasjonsskalaen er her i utgangspunktet standardisert med 500 som gjennomsnitt og 100 som standardavvik. Det bør for øvrig nevnes at det internasjonale gjennomsnittet for lesekompetanse ikke har endret seg signifikant fra 2001 til 2006. For Norges vedkommende er det en nedgang på 1 poeng på leseskalaen, men denne endringen er heller ikke statistisk signifikant. Hovedbudskapet er altså at norske fjerdeklassingers lesekompetanse er om lag den samme i 2006 som den var i 2001.

Tabell 2. Trender i lesekompetanse, PIRLS 2006 og PIRLS 2001 (4. klasse)

	2006	2001	Endring 2001-2006
<b>Alle elever</b>	498	499	-1
Jenter	508	511	-3
Gutter	489	489	0

Kilde: Mullis mfl. 2007, van Daal mfl. 2007.

Tabell 2 viser også tilsvarende resultater for jenter og gutter separat. Tabellen viser en liten tendens til at jentenes lesekompetanse er svakere i 2006 sammenliknet med i 2001, men heller ikke denne forskjellen er statistisk signifikant. Skårene for guttene i de to PIRLS-undersøkelsen er derimot helt tilsvarende.

Tabell 3 viser trender for ulike aspekter ved lesekompetanse: to lesehensikter og to leseprosesser. Tabellen viser en svak tilbakegang for «litterære hensikter» fra 2001 til 2006. Endringen er imidlertid ikke statistisk signifikant; den måtte vært i overkant av 7 poeng for å være signifikant (95 prosent-nivå). For «faktahensikter» kan vi registrere en svak og ikke signifikant framgang fra 2001 til 2006. En svak og ikke signifikant tilbakegang kan registreres for «hente ut informasjon og gjøre enkle inferenser», mens skåreverdiene er helt tilsvarende i 2006 og 2001 for «fortolke, integrere og vurdere».

Tabell 4 viser en mer detaljert sammenlikning av resultatene i de to PIRLS-undersøkelsene. Tabellen viser prosentandeler elever på ulike ferdighetsnivåer for lesekompetanse. De ulike ferdighetsnivåene er definert ut fra om elevene har nådd ulike nivåer på den internasjonale

prestasjonsskalaen for lesekompetanse: 625 for avansert nivå, 550 for høyt nivå, 475 for middels nivå og 400 for lavt nivå. De ulike ferdighetsnivåene er videre gitt beskrivelser av hva elever som befinner seg på nivået, typisk mestrer. Elever som når det avanserte nivået, kan blant annet tolke personers karaktertrekk, intensjoner og følelser i litterære tekster. De kan også tolke språklige bilder, samt begynne å undersøke og vurdere handlingsstrukturer. I møtet med faktatekster kan elevene blant annet fortolke sammensatt informasjon fra ulike deler av teksten. De kan også forstå betydningen av tekstens organisering (Mullis mfl. 2007).

Tabell 4 viser at andelene norske elever som når middels internasjonalt nivå, er rimelig tilsvarende i undersøkelsene fra 2001 og 2006. Det er imidlertid noe lavere andel norske elever som når avansert og høyt nivå i 2006 sammenliknet med i 2001. Disse forskjellene er dessuten statistisk signifikante. Imidlertid er det en noe større andel elever som når lavt nivå i 2006 sammenliknet med i 2001, og også denne endringen er statistisk signifikant.

Oppsummert kan vi si at PIRLS-undersøkelsene i 2001 og 2006 viser ubetydelige

Tabell 3. **Trender for typer lesekompetanse i Norge, PIRLS 2006 og PIRLS 2001 (4. klasse)**

	2006	2001	Endring 2001-2006
Hensikt 1: Litterære hensikter	501	506	-5
Hensikt 2: Faktahensikter	494	492	2
Prosess 1: Hente ut informasjon og gjøre enkle inferenser	502	505	-3
Prosess 2: Fortolke, integrere og vurdere	495	495	0

Kilde: Mullis mfl. 2007, van Daal mfl. 2007.

Tabell 4. **Trender i lesekompetanse på ulike nivåer, PIRLS 2006 og PIRLS 2001 (4. klasse)**

	Avansert nivå		Høyt nivå eller bedre		Middels nivå eller bedre		Lagt nivå eller bedre	
	2006	2001	2006	2001	2006	2001	2006	2001
Norge	2	4	22	28	67	65	92	88
Int. gj.	9	8	40	38	74	72	90	89

Kilde: Mullis mfl. 2007, van Daal mfl. 2007.

endringer i lesekompetanse blant fjerdeklassinger i Norge. Vi kan likevel registrere noen svake tendenser i negativ retning, og det er åpenbart interessant hvordan disse tendensene utvikler seg fram mot PIRLS-undersøkelsen som skal gjennomføres i 2011.

### Trender i PISA: Kompetanse i lesing, matematikk og naturfag på 10. klassetrinn

PISA-undersøkelsen måler, som tidligere nevnt, 15-åringers kompetanse i lesing, matematikk og naturfag. Nesten alle disse elevene går i Norge på 10. klassetrinn. Vi skal i det følgende studere trender for hvert av de tre fagområdene, og starter med lesing.

#### Lesekompetanse på 10. klassetrinn

Leseforståelse i PISA-studien vil si å kunne forstå, bruke og reflektere over skrevet tekst for å nå egne mål, utvikle egen kunnskap og potensial, samt delta i samfunnet (OECD 2006). Tabell 5 viser trender i lesekompetanse på 10. klassetrinn i perioden 2000-2006 slik de framstår gjennom de tre PISA-undersøkelsene i perioden. Prestasjonsskalaene i PISA er i utgangspunktet standardisert med 500 som gjennomsnitt for OECD-landene og 100 som standardavvik. Det må nevnes at det har vært en svekkelse av gjennomsnittlig lesekompetanse i OECD-landene i perioden. Vi finner samme tendens i Norge, men svekkelsen er betydelig sterkere. Fra å ligge noe over OECD-gjennomsnittet for leseforståelse i 2000 ligger norske 15-åringer i 2006 under gjennomsnittet. Det er også verd å merke seg at vi finner en klart større nedgang i perioden 2003-2006 sammenliknet med perioden 2000-2003.

Tabell 5 viser også utviklingen i lesekompetanse for henholdsvis jenter og gutter. Tabellen viser at tilbakegangen i perioden 2000-2006 har vært litt større for guttene

enn for jentene. For begge kjønn er nedgangen tiltakende i perioden, men dette er særlig tydelig for jentene. Det er interessant å observere at vi altså ser en tydelig svekkelse av lesekompetansen på 10. trinn i perioden 2000-2006, mens det derimot ikke er signifikant forskjell i lesekompetanse på 4. trinn i 2006 sammenliknet med i 2001 (se tabell 2). Det er åpenbart interessant om vi også vil se tilsvarende stabilisering av nivået i framtidige PISA-undersøkelser.

Tabell 5. Trender i lesekompetanse, PISA 2006, PISA 2003 og PISA 2000 (10. klasse)

	2006	2003	2000	Endring 2000- 2006
<b>Alle elever</b>	484	500	507	-23
Jenter	508	525	529	-21
Gutter	462	476	486	-24

Kilde: OECD 2007, [www.pisa.oecd.org](http://www.pisa.oecd.org), Kjærnsli mfl. 2007.

#### Matematikkkompetanse på 10. klassetrinn

Matematikkkompetanse i PISA-studien handler om en persons evne til å identifisere og forstå den rollen som matematikken spiller i verden, evne til å foreta velfunderte vurderinger, samt å kunne anvende og engasjere seg i matematikk på måter som fyller personens behov som et konstruktivt, bevisst og reflektert medlem av samfunnet (OECD 2006).

Tabell 6 viser endringer i matematikkkompetanse i perioden 2000-2006. I matematikk ble rammeverket justert mellom undersøkelsene i 2000 og 2003. Sammenlikninger av resultater fra disse to undersøkelsene må derfor gjøres med noe varsomhet. Resultatene i tabell 6 viser en nedgang i perioden som helhet for de norske elevene, men nedgangen er imidlertid ikke så stor som den vi så for lesing i samme periode. Tabellen viser også at nedgangen har vært noe større for guttene enn for jentene i perioden.



Tabell 6. **Trender i matematikkompetanse, PISA 2006, PISA 2003 og PISA 2000 (10. klasse)**

	2006	2003	2000	Endring 2000- 2006
<b>Alle elever</b>	490	495	500	-10
Jenter	487	492	494	-7
Gutter	493	498	506	-13

Kilde: OECD 2007, [www.pisa.oecd.org](http://www.pisa.oecd.org), Kjærnsli mfl. 2007.

### **Naturfagkompetanse på 10. klassetrinn**

Naturfagkompetanse i PISA-studien handler om evne til å bruke naturfaglig kunnskap til å identifisere spørsmål, få ny kunnskap, forklare naturfaglige fenomener og til å trekke evidensbaserte konklusjoner om naturfagrelaterte temaer. Det handler også om å forstå sentrale trekk ved naturvitenskap som kunnskapsform og virksomhet. Naturfagkompetanse innbefatter videre å ha bevissthet om hvordan naturvitenskap og teknologi former våre materielle, intellektuelle og kulturelle miljøer. Dessuten handler det om vilje til å engasjere seg i naturfagrelaterte temaer og sentrale ideer som et reflektert samfunnsmedlem (OECD 2006). Som det går fram, søker man også å måle affektive aspekter ved naturfaget, som vilje til engasjement og så videre. Disse aspektene ble introdusert som del av rammeverket for PISA-undersøkelsen i 2006, da naturfag var hovedområde. I det følgende vil vi fokusere på trender knyttet til de kognitive aspektene ved definisjonen presentert ovenfor.

Tabell 7 viser utviklingen i naturfagkompetanse på 10. klassetrinn i perioden 2000-2006. I naturfag ble, som nevnt, rammeverket justert mellom undersøkelsene i 2003 og 2006. Sammenlikninger av resultatene fra disse to undersøkelsene må derfor gjøres med noe varsomhet. Tabell 7 viser en nedgang i perioden som helhet, og den var markant fra 2000 til 2003. Nedgangen i prestasjoner i perioden er ganske lik for jenter og gutter.

Tabell 7. **Trender i naturfagkompetanse, PISA 2006, PISA 2003 og PISA 2000 (10. klasse)**

	2006	2003	2000	Endring 2000- 2006
<b>Alle elever</b>	487	484	502	-15
Jenter	489	483	505	-16
Gutter	484	485	499	-15

Kilde: OECD 2007, [www.pisa.oecd.org](http://www.pisa.oecd.org), Kjærnsli mfl. 2007.

Oppsummert kan vi si at PISA-undersøkelsene fra 2000, 2003 og 2006 viser tilbakegang i prestasjoner innenfor alle de tre fagområdene som er målt.

### **Trender i TIMSS: Kompetanse i matematikk og naturfag på 4. og 8. klassetrinn**

Som tidligere nevnt, måler man i TIMSS-undersøkelsene elevenes kompetanse i matematikk og naturfag. Vi starter med å se på trender for matematikk.

#### **Matematikkompetanse på 4. klassetrinn**

I TIMSS tar man utgangspunkt i læreplanene i matematikk i de landene som deltar. Basert på disse lager man et rammeverk for matematikkompetanse som søker å fange opp det som typisk undervises i faget på de to klassetrinnene som undersøkelsen tar for seg. I matematikk på 4. trinn opererer man med disse innholdsområdene i rammeverket for undersøkelsen: tall, geometriske former og mål og framstilling av data (Mullis mfl. 2005). I tillegg inneholder rammeverket en kognitiv dimensjon innenfor de faglige områdene med disse kategoriene: kunne, anvende og resonnere.

Tabell 8 viser utviklingen i matematikkompetanse på 4. trinn i perioden 1995-2007 slik den framstår gjennom de tre TIMSS-undersøkelsene som er gjennomført i perioden (Mullis mfl. 2008). Det internasjonale gjennomsnittet i 1995 er satt til 500 med standardavvik 100. Tabellen viser at

for perioden som helhet er det en svak, men ikke signifikant, nedgang i prestasjoner (3 poeng). Forskjellen måtte ha vært om lag 8 poeng for å være statistisk signifikant (95 prosent-nivå). Tabellen viser også at det var en relativt kraftig nedgang i matematikkprestasjoner fra 1995 til 2003, men at nivået i 2007 nesten er tilbake på 1995-nivå. Sagt på en annen måte har de norske 4. klassingenes prestasjoner gått fra å ligge om lag et halvt standardavvik under det internasjonale gjennomsnittet i 2003 til å ligge om lag et kvart standardavvik under gjennomsnittet i 2007. I TIMSS rapporterer man også elevenes prestasjoner i forhold til ulike ferdighetsnivåer definert ut fra hvor høyt elevene skårer på den internasjonale prestasjonsskalaen. Det er laget beskrivelser av hva elevene typisk mestrer i matematikk på de ulike ferdighetsnivåene. Sammenlikninger av andelene elever på de ulike ferdighetsnivåene i 2007 og 1995 viser imidlertid små og ubetydelige forskjeller på 4. trinn i Norge.

Tabell 8 viser også trender for henholdsvis jenter og gutter i matematikk på 4. trinn. Tabellen viser at guttene skårer noe høyere enn jentene i alle tre TIMSS-undersøkelser. For begge kjønn ligger skåreverdiene i 2007 litt lavere enn i 1995, men endringene for perioden som helhet er ikke signifikante.

### Matematikkkompetanse på 8. trinn

Matematikk på 8. trinn i TIMSS har disse innholdsområdene: tall, algebra, geometri samt data og sannsynlighet (Mullis mfl.

Tabell 8. Trender i matematikkkompetanse i TIMSS 2007, TIMSS 2003 og TIMSS 1995 (4. klasse)

	2007	2003	1995	Endring 1995- 2007
<b>Alle elever</b>	473	451	476	-3
Jenter	470	449	474	-4
Gutter	477	454	478	-1

Kilde: Mullis mfl. 2008, Grønmo mfl. 2009.

2005). Tabell 9 viser utviklingen i matematikkkompetanse på 8. klassetrinn i perioden 1995-2007. Tabellen viser en betydelig nedgang i prestasjoner for perioden sett under ett, selv om det er en svak tendens til bedring fra 2003 til 2007. I motsetning til på 4. trinn, ligger elevene på 8. trinn i 2007 fremdeles langt under nivået i 1995.

Tabell 9 viser også separate trendresultater for jenter og gutter på 8. klassetrinn. Tabellen viser at guttene som gruppe har gått noe mer tilbake i faglig nivå enn jentene i perioden som helhet. Det er også verdt å merke seg at det er en svak tendens til at jentene skårer høyere enn guttene i 2007, altså motsatt tendens av det vi så på 4. trinn.

Vi har tidligere sett at det er en tilbakegang i matematikknivå på 10. trinn i PISA fra 2003-2006 (se tabell 6), mens tabell 9 derimot viser en svak framgang fra 2003-2007 for 8. trinn i TIMSS. Det kan konstateres at elevene på 8. trinn har blitt undervist ett år etter Kunnskapsløftet, mens elevene på 10. trinn har fått all sin undervisning etter L97. Vi kommer nærmere tilbake til dette avslutningsvis i artikkelen.

Som på 4. trinn opererer man med ulike ferdighetsnivåer med tilhørende beskrivelser av hva elevene typisk mestrer i matematikk på de ulike nivåene. Elever på avansert nivå i matematikk på 8. trinn kan organisere og trekke konklusjoner basert på informasjon, gjøre generalise-

Tabell 9. Trender i matematikkkompetanse i TIMSS 2007, TIMSS 2003 og TIMSS 1995 (8. klasse)

	2007	2003	1995	Endring 1995- 2007
<b>Alle elever</b>	469	461	498	-29
Jenter	471	463	497	-26
Gutter	467	460	499	-32

Kilde: Mullis mfl. 2008, Grønmo mfl. 2009.

ringer, samt løse oppgaver de ikke har en forhåndsdefinert rutine for å løse. De kan også løse et bredt spekter av oppgaver som omhandler forhold, proporsjonalitet og prosent. De kan videre anvende kunnskaper om begreper og sammenhenger innenfor algebra og tall. Elevene kan dessuten uttrykke generaliseringer algebraisk og modellere situasjoner. De kan også anvende kunnskaper innenfor geometri på komplekse problemstillinger. Endelig kan de hente ut og bruke data fra ulike kilder til å løse oppgaver som består av flere trinn (Mullis mfl. 2008). I TIMSS 1995 når 4 prosent av norske 8. klassinger dette nivået. I 2003 har andelen sunket til 0 prosent, og dette er samme andel som i TIMSS 2007.

Elever som når høyt nivå i TIMSS, kan anvende sin forståelse og kunnskap i ulike relativt kompliserte situasjoner. De kan forholde seg til og regne med brøker, desimaltall og prosent, samt operere med negative tall og løse tekstoppgaver som involverer proporsjonalitet. Elevene kan også arbeide med algebraiske uttrykk og likninger av første grad, og de kan bruke kunnskaper om geometriske forhold til å løse oppgaver som inkluderer areal, volum og vinkler. Endelig kan disse elevene tolke data i ulike typer grafer og tabeller og løse enkle problemer som omhandler sannsynligheter. I TIMSS 1995 når omlag en firedel (26 prosent) av de norske 8. klassingene dette ferdighetsnivået. I undersøkelsen i 2003 hadde andelen sunket til 10 prosent, mens den er 11 prosent i TIMSS 2007 (Mullis mfl. 2008).

#### **Naturfagkompetanse på 4. klassetrinn**

Naturfag på 4. trinn i TIMSS har disse innholdsområdene: biologi, fysikk og geofag (Mullis mfl. 2005). Tabell 10 viser gjennomsnittresultater i naturfag for 4. klassetrinn i de tre TIMSS-undersøkelsene. Tabellen viser en markant nedgang i

prestasjoner for perioden som helhet. Fra å ligge noe over gjennomsnittet i 1995 skårer elevene i 2007 om lag et kvart standardavvik under gjennomsnittet. Like fullt er det en signifikant framgang i prestasjonene fra 2003 til 2007. Tabell 10 viser også separate trendresultater for jenter og gutter. Tabellen viser at guttene har en klart større nedgang i prestasjoner i perioden som helhet enn det jentene har.

Tabell 10. **Trender i naturfagkompetanse i TIMSS 2007, TIMSS 2003 og TIMSS 1995 (4. klasse)**

	2007	2003	1995	Endring 1995- 2007
<b>Alle elever</b>	477	466	504	-27
Jenter	475	467	496	-21
Gutter	478	466	510	-32

Kilde: Martin mfl. 2008, Grønmo mfl. 2009.

#### **Naturfagkompetanse på 8. klassetrinn**

Naturfag på 8. trinn i TIMSS har følgende innholdsområder: biologi, kjemi, fysikk og geofag (Martin mfl. 2005). Tabell 11 viser trender i naturfagkompetanse på 8. klassetrinn. Igjen ser vi en stor og betydelig nedgang i prestasjoner for perioden som helhet, men her ser vi også en signifikant nedgang fra 2003 til 2007. Tabell 11 viser også separate trendresultater for jenter og gutter på 8. klassetrinn. Resultatene viser en betydelig større nedgang i prestasjoner for guttene enn for jentene i perioden som helhet. Det er også interessant å merke seg at resultatene for jentene er tilsvarende i 2007 og 2003, mens det derimot er en markant nedgang for guttene i denne perioden.

Tabell 11. **Trender i naturfagkompetanse i TIMSS 2007, TIMSS 2003 og TIMSS 1995 (8. klasse)**

	2007	2003	1995	Endring 1995- 2007
<b>Alle elever</b>	487	494	514	-28
Jenter	487	489	505	-18
Gutter	486	498	523	-37

Kilde: Martin mfl. 2008, Grønmo mfl. 2009.

Elever som når avansert nivå i naturfag på 8. klassetrinn, kan vise forståelse av noen komplekse og abstrakte begreper i biologi, kjemi, fysikk og geofag. De har videre en forståelse av kompleksiteten til levende organismer og hvordan organismene samvirker med miljøet de lever i. Elevene viser forståelse av egenskaper ved magneter, lyd og lys, og de kan demonstrere forståelse av stoffers struktur, samt fysiske og kjemiske endringer. Elevene kan også anvende kunnskaper om solsystemet og jordas egenskaper og prosesser, og de kan anvende forståelse av store miljøutfordringer. Elever som når dette nivået, forstår noen grunnleggende trekk ved vitenskapelige metoder og kan anvende grunnleggende fysiske prinsipper til å løse noen kvantitative problemstillinger. De kan også kommunisere naturfaglige kunnskaper og forklaringer skriftlig. Resultatene viser at kun 2 prosent av norske 8. klassinger når avansert internasjonalt nivå i naturfag i 2007. I 1995 var andelen 6 prosent (Martin mfl. 2008).

Elever som når høyt nivå i naturfag på 8. trinn, kan demonstrere begrepsmessig forståelse av noen kretsløp, systemer og prinsipper i naturen. Elevene har også noe forståelse av biologiske begreper, prosesser i celler, humanbiologi og helse, samt samspillet mellom planter og dyr i økosystemer. Videre kan elevene anvende kunnskaper i situasjoner som involverer lyd og lys, demonstrere elementære kunnskaper om varme og krefter, samt vise noe forståelse av stoffers struktur, kjemiske og fysiske egenskaper og endringer. Elevene kan også demonstrere noe forståelse av solsystemet, jordas prosesser og ressurser, samt noe grunnleggende forståelse når det gjelder miljøspørsmål. Dessuten kan elevene demonstrere noe kompetanse når det gjelder vitenskapelige metoder. Elevene kan også kombinere informasjon for å trekke konklusjoner, tolke tabeller og

grafer, samt lage korte forklaringer som involverer naturfaglig kunnskap. I TIMSS 1995 når omlag en av tre norske elever dette nivået på 8. trinn (32 prosent). I 2007 gjelder det imidlertid kun en av fem (20 prosent) (Martin mfl. 2008).

Oppsummert kan man si at TIMSS-undersøkelsen viser nedgang i prestasjoner i matematikk og naturfag i perioden 1995-2007. Det er imidlertid noen tegn til bedring fra 2003 til 2007 (Grønmo mfl. 2009).

### **Fortolkninger av trendene i fagkompetanse**

Resultatene vi har studert så langt, viser interessante trender i fagkompetanse innenfor områdene lesing, matematikk og naturfag det siste drøye tiåret. I det følgende skal vi ta for oss noen av de fortolkningene og forklaringene som er blitt fremmet i kjølvannet av de negative trendene i de internasjonale undersøkelsene.

### **Generelle og fagspesifikke forklaringer**

De kvantitative undersøkelsene gir i seg selv begrenset informasjon som grunnlag for å fortolke og forklare hva som ligger bak trendresultatene som er presentert i denne artikkelen. Som en støtte til fortolkningene må man derfor også se til andre undersøkelser. PISA+ (Pluss står her for *Prosjekt om Lærings- og Undervisningsstrategier i Skolen*) er for eksempel en kvalitativ studie som studerer trekk ved livet i klasserommet, blant annet gjennom videoobservasjoner. Målet er å identifisere mulige faktorer som kan forklare de problematiske funnene i PISA-undersøkelsen (Klette mfl. 2008).

Et grunnleggende spørsmål er om man skal søke etter fagspesifikke eller mer generelle forklaringer på de nedadgående trendene i perioden 1995-2007. Innenfor alle de målte fagområdene ser vi negative

trender, riktignok av svært ulik styrke, både i realfagene og i en grunnleggende kompetanse som lesing. Selv om man i TIMSS-undersøkelsen fra 2007 ser noen tendenser til forbedring av resultatene siden 2003, er likevel det overordnede bildet for perioden 1995-2007 preget av nedgang. Resultatene peker tydelig i retning av at vi også bør søke etter mer generelle trekk ved det norske skolesystemet som potensielt kan bidra til svekkede fagkompetanser. I tillegg vil opplagt mer fagspesifikke forklaringer spille en rolle.

### **To læreplanreformer i perioden**

I perioden 1995-2007 har vi hatt to læreplanreformer: innføringen av Læreplanen av 1997 (L97) og Kunnskapsløftet fra 2006 (LK06). Forskning på læreplanimplementering indikerer at vi i svært liten grad kan vente å se effektene av Kunnskapsløftet på resultatene innhentet i 2007 (Gundem 1990, Engelsen 2008a, 2008b), da implementering av læreplaner er en langsom prosess. Det er imidlertid grunn til å tro at de positive tendensene vi ser i TIMSS 2007, i noen grad kan tilskrives Kunnskapsløftet. For elever på 4. trinn vil ett års opplæring etter denne nye læreplanen utgjøre en betydelig andel av deres totale skolegang. Men for perioden 1995-2007 som helhet er det naturlig å primært se mot L97 for mulige forklaringer. Enkelte vil imidlertid hevde at implementeringen av en læreplan tar så lang tid at vi heller ikke fullt ut kan se effektene av intensjonene i denne læreplanen.

### **Endrete lærer- og elevroller det siste tiåret**

Mange av de foreslåtte forklaringene på den svekkete fagkompetansen henger tett sammen med endrete elev- og lærerroller i norsk skole det siste drøye tiåret (se for eksempel Kjærnsli mfl. 2007). Noen av disse endringene ble direkte foreskrevet i L97, som for eksempel at man skulle

anvende en viss andel av tiden til prosjektarbeid. Andre endringer i skolen har i større grad kommet i form av det man kan kalle «bottom up»-initiativer. Dette gjelder for eksempel bruken av arbeidsplaner som verktøy for å imøtekomme læreplanens krav om tilpasset opplæring (Klette 2007, Bergem 2009).

Forskning viser at arbeidsplaner kan føre til liten innsats hos mange elever. For enkelte fag, som matematikk, kan noen former for arbeidsplaner virke direkte mot sin hensikt hva læring angår. Observasjoner viser at elevene når de selv kan disponere sin tid i arbeidet med arbeidsplanene, ofte ikke ser noen grunn til å arbeide med matematikk. Videre har ofte ikke læreren som er til stede i arbeidsøkta, faglig bakgrunn i matematikk. Arbeidsplanene stiller også store krav til at eleven kan organisere seg selv, noe som kan føre til større forskjeller i læringsutbytte langs sosiale skillelinjer. Det hevdes at i mange klasserom blir faglig forståelse lett et anliggende mellom elev, læreboktekst og arbeidsplan, der læreren spiller en perifer og tilbaketrukket rolle (Klette 2007, Bergem 2009).

Generelt har det de siste tiårene skjedd store endringer i måten elevene og undervisningen er organisert på. Fra en situasjon hvor det typiske var at en lærer underviste i en klasse i ett klasserom, kan elevene nå være organisert i grupper og baser. Gruppene kan være fleksible, både innenfor og på tvers av klassetrinnet. Å mestre slike nye undervisningsformer krever innsikt både hos ledelse og lærere, og flere har påpekt at det her er grunn til bekymring for hvorvidt alle lærere mestrer nettopp dette (for eksempel Kjærnsli mfl. 2004, 2007).

Det kan nevnes at i St.meld. nr. 31 (2007-2008) varsles det at man vil sende på høring et forslag til justering av opplæringsloven der det framgår at elevene i

opplæringen skal deles i klasser eller basisgrupper som skal ivareta elevenes behov for sosial tilhørighet. For deler av opplæringen kan elevene deles i andre grupper etter behov. Klassen/basisgruppen skal ha en eller flere lærere (kontaktlærere) som har særlig ansvar for de praktiske, administrative og sosialpedagogiske gjøremålene som gjelder klassen/basisgruppen, blant annet kontakten med hjemmet. Det påpekes at selv om det også i dag er mange som velger tradisjonell klasseorganisering, bør loven inkludere begrepet «klasse» som en mulig måte å organisere opplæringen på, slik at det ikke oppstår noen misforståelser knyttet til dette.

### **Aktiviteter uten klar hensikt**

Både PISA+-undersøkelsen (Klette mfl. 2008) og den forskningsbaserte evalueringen av L97 (Haug 2004) finner at klasserommene preges av mye aktivitet som engasjerer elevene. Selv om det er mye aktivitet og oppgaveløsning i klasserommene, er ofte aktivitetenes hensikt og mål uklart festet hos elevene. Lærerne fokuserer gjerne på hva man skal gjøre der og da, men aktiviteten settes i mindre grad inn i en faglig sammenheng. Læringsmålene for aktivitetene er med andre ord ofte utydelige. Videre mangler ofte lærerens aktive og systematiske introduksjon og oppsummering av aktivitetene. Dette fører til at mye av læringen blir «privatisert» og overlatt til elevene selv.

PISA+-undersøkelsen avdekker at lærernes systematiske og gjennomtenkte bruk av ulike læringsaktiviteter framstår som lite optimal. Dessuten blir systematisk bruk av medelever som læringsressurs og medspiller for lærings situasjoner sjelden observert. Det er interessant at man finner dette på tross av at elever ofte er plassert sammen i grupper slik at samarbeid faktisk er mulig. Lærerne stimulerer imidlertid i liten grad til felles problemløsning.

Studien finner videre at det generelt er få faglige elevsamtaler i naturfag og matematikk (Klette mfl. 2008, Bergem 2009).

### **Tverrfaglighetens problem**

Norsk grunnskole har lenge hatt fokus på tverrfaglighet. Gjennom L97 kom tverrfaglig prosjektarbeid for fullt inn i norsk skole med krav om at 20 prosent av tida skulle brukes til dette. Tverrfaglig arbeid fører naturlig nok til at man flytter fokuset fra de enkelte fagene til temaet for selve prosjektet. Det er opplagt at skolefagene i større eller mindre grad har en egen iboende struktur. Realfagene er nok i større grad preget av dette enn mange andre fag, og gjennom tverrfaglig arbeid kan denne strukturen i stor grad bli utfordret. Enkelte fag kan videre nedtones eller forsvinne i en prosjektperiode. Det er derfor blitt hevdet at det er en fare for at flerfaglig, temabasert undervisning kan svekke de faglige kunnskapene, særlig i realfag (Kjærnsli mfl. 2007).

Det har også blitt stilt spørsmål ved effektiviteten av prosjektarbeid generelt. Valdermo (2007) hevder for eksempel at det i praksis viser seg at forsøk med flerfaglighet ikke er så vellykket, og at et fag som matematikk ofte kommer dårlig ut i slike arbeidsformer. Det stiller i hvert fall svært store krav til lærerne når det gjelder å sikre at alle elever får utbytte av arbeidet. Undersøkelser påviser også et smalt repertoar av læringsstrategier på tvers av fag og klasserom. Det er for eksempel liten bruk av og opplæring i utdypningsstrategier, det vil si å knytte nytt fagstoff aktivt til det man allerede kan. Frie arbeidsformer som gruppearbeid og prosjektarbeid stiller desto større krav til elevenes egne strategier og eget driv i læringsarbeidet (Elstad og Turmo 2006, 2009).

### **Realfagene oppfattes som lite relevante**

The Relevance of Science Education (ROSE) er en undersøkelse som samler inn data fra elever i ulike land ved hjelp av et spørreskjema ([www.ils.uio.no/english/rose/](http://www.ils.uio.no/english/rose/)). Målgruppen for ROSE er elever i 15-årsalderen, med andre ord samme elevgruppe som i PISA. Gjennom spørreskjemaet søker man å gi en beskrivelse av hva elever bringer med seg av relevante erfaringer, hva slags interesser de har, hvilke framtidsplaner de har, og hva slags forestillinger og holdninger de har til naturvitenskap, teknologi, miljøutfordringer, forskning og forskere. En viktig bakgrunn for undersøkelsen er at man i mange vestlige, høyt industrialiserte land registrerer en avtakende interesse for studier innenfor naturvitenskap og teknologi. ROSE finner at interessen for naturfag og teknologi blant de unge gjerne avtar med økende utviklingsgrad. Realfagene framstår ikke som relevante bidragsyttere i unge menneskers identitetsprosjekt i et modernisert samfunn som det norske, der enkeltindividet i stor grad er kulturelt fristilt og må skape sin egen identitet (Schreiner og Sjøberg 2006).

### **Svakt læringstrykk**

Begrepet «læringstrykk» kom for alvor inn i norsk offentlighet i forbindelse med arbeidet med den første kompetanseberetningen (UFD 2003a, 2003b). Begrepet ble brukt om forventninger og læringstrykk til arbeidstakere, virksomheter, sektorer og så videre. Læringstrykk overfor elever, har kommet mer i fokus etter hvert. I St.meld. nr. 30 Kultur for læring framheves det at lærerne må sørge for tilstrekkelig ytre trykk i opplæringen (Elstad og Turmo 2008). Undersøkelser påviser svakt læringstrykk i norsk grunnskole, med andre ord at elevene møter lave krav. Men man finner også stor variasjon i læringstrykk på tvers av klasserom og fag (Klette mfl.

2008). En undersøkelse i Vg1 fra 2008 konkluderer imidlertid med at det er lite grunnlag for å hevde at læringstrykket er lavt i naturfagklasserommene på dette nivået. Målingene var vel og merke basert på elevenes vurderinger. Det er imidlertid grunnlag for å hevde at elevene i stor grad ønsker læringstrykk i naturfagklasserommet. Dette kan forstås som et uttrykk for en viljesstrategi, en selvønsket binding fra elevens side. Elevene kan med andre ord ønske at læreren skal «pushe» dem til å jobbe enda hardere (Elstad og Turmo 2008).

### **Lite tid til faglig læring**

Norge har lavt timetall i naturfagene sammenliknet med andre OECD-land. Det er blitt hevdet at dette til en viss grad kan forklare de svake norske resultatene i akkurat dette faget. I matematikk og morsmål er derimot timetallet omtrent som det internasjonale gjennomsnittet (Sjøberg 2007).

Undersøkelser viser dessuten at store deler av undervisningstiden i norsk skole går bort til ikke-faglige gjøremål, som administrasjon, fortelle hva som skal skje, irettesettelse, gi beskjeder osv. Det antydes at dette dreier seg om så mye som 30 prosent av tiden (Haug 2006). I PISA 2000 og 2003 inngikk det spørsmål om bråk, uro og sløsing med tid i elevspørreskjemaet. Norske elever er blant de som rapporterer om størst problemer her i OECD-området (Lie mfl. 2001, Kjærnsli mfl. 2004). Den årlige elevundersøkelsen i regi av Utdanningsdirektoratet bekrefter i stor grad disse funnene (Skaar mfl. 2008).

Flere har også hevdet at mange undervisningstimer går bort til andre aktiviteter, som idrettsdager, prosjektdager og lignende. Ikke minst har dette vært i søkelyset i videregående skole. Det har blitt fokusert sterkere på at elevene skal få det antall timer de har krav på gjennom

det nasjonale tilsynet. Nasjonalt tilsyn avdekker for øvrig at elever i videregående skole får færre timer til opplæring enn de har krav på. Det er spesielt ved avvikling av eksamen, oppstart av skoleåret, idrettsdager og andre aktiviteter at elever mister timer (Udir 2008b).

Det kan i denne forbindelse også nevnes at Kunnskapsdepartementet i 2008 nedsatte et utvalg som skulle foreslå tiltak som vil gi bedre utnyttelse av tidsressursene i grunnskolen for at elevene skal få gode læringsvilkår og gode læringsresultater («Tidsbruk-utvalget»). Utvalget skal etter planen avgi sin rapport i desember 2009.

### **Lærernes kompetanse**

Undersøkelser har vist lav faglig kompetanse blant mange lærere som underviser i matematikk og naturfag i grunnskolen. Særlig gjelder dette på barnetrinnet. Mange lærere har selv valgt bort realfagene i videregående skole og har liten eller ingen fordypning i fagene fra sin lærerutdanning (Sjøberg 2007).

PISA+ viser at det er stor variasjon i lærernes kunnskaper og forutsetninger i systematisk bruk og oppfølging av fagstoff, læringsaktiviteter og tilgjengelige læringsverktøy i ulike fag. Bare et fåtall av de observerte lærerne har for eksempel skole-ring i ny innsikt når det gjelder lesestrategier og leseforståelse (Klette mfl. 2008).

### **Svak kultur for vurdering**

Flere forskere, blant annet de som evaluerte L97, hevder at det eksisterer en svak vurderingskultur i norsk skole. Særlig på barnetrinnet er de tilbakemeldingene som gis, lite presise, og de er i liten grad læringsstøttende. Tilbakemeldingene er gjerne ensidig positive, relativt uavhengig av selve kvaliteten på elevenes arbeid (Klette 2003, Haug 2004). TIMSS-undersøkelsen har påvist lite oppfølging

og tilbakemelding på elevenes leksearbeid. Undersøkelsen viser også at norske elever er blant dem med høyest selvoppfatning i matematikk og naturfag, selv om elevene skårer som gjennomsnittet eller lavere rent faglig (Mullis mfl. 2008, Martin mfl. 2008). PISA+ viser at i noen klasserom er lærernes veiledning i hovedsak knyttet til emosjonell støtte og motivering, mens i andre klasserom er lærernes veiledning konsentrert rundt faglige anvisninger og kommentarer. Det er lite vanlig å veilede elevene i aktivt å overvåke egen problemløsning (Klette mfl. 2008).

### **Mangel på tidlig innsats**

Flere forskere har påpekt at det eksisterer en «vente og se»-holdning i norsk skole. I stedet for å ta tak i eventuelle problemer den enkelte elev måtte ha tidlig i utdanningsløpet, utsetter man å gripe inn med håp om at det vil bedre seg over tid. Det er blitt hevdet at norsk skole på dette punktet skiller seg betydelig fra for eksempel den finske skolen, der det i større grad legges vekt på tidlig innsats. Det kan i denne forbindelse nevnes at det i St.meld. nr. 16 (2006-2007) ble lansert flere tiltak for å styrke tidlig innsats i skolesystemet. Her inngår blant annet å videreutvikle nasjonale prøver og kartleggingsprøver tilpasset Kunnskapsløftet og å ta initiativ til at det blir utarbeidet en veileder med konkrete eksempler på hvordan kommunene kan jobbe med tidlig innsats i barnehagen og skolen.

### **Skolens og kunnskapens status i samfunnet**

I tillegg til å søke etter forklaringer i selve skolesystemet har også mer generelle trekk ved samfunnsutviklingen blitt trukket inn som forklaringsfaktorer. I sin bok om årsaker til de gode finske resultatene i PISA-undersøkelsen trekker også Välijärvi mfl. (2002) inn mer generelle samfunnsmessige forhold. Kunnskap og skole har



høy status i det finske samfunnet, noe som for eksempel setter sitt preg på rekrutteringen til læreryrket. Hvor viktig det er å lykkes i utdanningssystemet, kan også variere betydelig fra land til land. Det er blitt hevdet at vi i Norge etter hvert har det «for» godt, har nok penger og ser ingen grunn til ekstra anstrengelser. Begrepet «oljedopet» er blant annet brukt i denne sammenhengen (Henmo 2004).

### Avslutning

I denne artikkelen har vi fokusert på trender i fagkompetanse blant norske elever det siste drøye tiåret slik de er påvist gjennom internasjonale undersøkelser. Noen mulige forklaringsfaktorer bak trendene er også diskutert. Undersøkelsene dekker sentrale kompetanseområder som lesing, matematikk og naturfag, men omfatter selvsagt ikke hele bredden av de mål som norsk skole gjennom læreplanverket er forpliktet til å arbeide mot. Selv om vi i de omtalte undersøkelsene ser negative trender, er det viktig at vi ikke på dette grunnlaget alene feller en altomfattende negativ «dom» over utviklingen i norsk skole. Dessverre skjer dette ofte i massemedia. Det kan opplagt hende at norsk skole det siste drøye tiåret er blitt bedre på andre viktige områder, selv om dette i mange tilfeller er vanskelig å dokumentere på en overbevisende måte. Men det er også viktig at en eventuell forbedring på andre områder ikke brukes som bortforklaring av de negative trendene innenfor sentrale fagkompetanser. Imidlertid utgjør resultatene vi har diskutert i denne artikkelen, kun en del av det grunnlaget som bør ligge til grunn for en balansert diskusjon av kvalitetsutviklingen i norsk skole de siste årene.

### Referanser

- Bergem, Ole Kristian (2009): Individuelle versus kollektive arbeidsformer. En drøfting av aktuelle utfordringer i matematikkundervisningen i grunnskolen, PhD-avhandling, Universitetet i Oslo.
- Bonesrønning, Hans (1999): Variations in Teachers' Grading Practices: causes and consequences, *Economics of Education Review*, 18, 89-105.
- Campbell, Jay R., Dana L. Kelly, Ina V. S. Mullis, Michael O. Martin, Marian Sainsbury (2001): Framework and Specifications for PIRLS Assessment 2001. PIRLS International Study Center, Boston College.
- Crocker, Linda og James Algina (1986): Introduction to Classical and Modern Test Theory, Forth Worth: Harcourt Brace Jovanovich College Publishers.
- Elstad, Eyvind og Are Turmo (2006): Læringsstrategier. Søkelys på lærernes praksis, Oslo: Universitetsforlaget.
- Elstad, Eyvind og Are Turmo (2008): Er læringstrykket lavt i norske klasserom? *Utdanning* nr. 21, 2008.
- Elstad, Eyvind og Are Turmo (2009): Students' self-regulation and teachers' regulating approaches in science: Results from a Norwegian Study, *Journal of Baltic Science Education*, 8(1), 35-43.
- Engelsen, Britt Ulstrup (2008a): Kunnskapsløftet: Sentrale styringssignaler og lokale strategidokumenter, Oslo: Pedagogisk forskningsinstitutt, Universitetet i Oslo.

- Engelsen, Britt Ulstrup (2008b): Kunnskapsløftet og skoleeierne, *Bedre skole* nr. 4.
- Grønmo, Liv Sissel mfl. (2009): Tegn til bedring: norske elevers prestasjoner i matematikk og naturfag i TIMSS 2007, Oslo: Unipub.
- Gundem, Bjørg Brandtzæg (1990): Læreplanpraksis og læreplanteori: en introduksjon til læreplanområdet, Oslo: Universitetsforlaget.
- Haug, Peder (2004): Resultat frå evalueringa av Reform 97, Oslo: Norges forskningsråd.
- Haug, Peder (red.) (2006): Begynneropplæring og tilpassa undervisning – kva skjer i klasserommet? Bergen: Caspar forlag.
- Henmo, Ola (2004): Clemet krever større lærelyst, *Aftenposten* 24.06.2004.
- Kjærnsli, Marit, Svein Lie, Rolf Vegar Olsen, Astrid Roe og Are Turmo (2004): Rett spor eller ville veier? Oslo: Universitetsforlaget.
- Kjærnsli, Marit, Svein Lie og Are Turmo (2005): Kan elevene mindre enn før? Naturfagkompetanse i Norden i perioden 1995-2003, *NorDiNa: Nordic Studies in Science Education*, 1(2) s. 51-60.
- Kjærnsli, Marit, Svein Lie, Rolf Vegar Olsen og Astrid Roe (2007): Tid for tunge løft. Oslo: Universitetsforlaget.
- Klette, Kirsti (red.) (2003): Klasserommets praksisformer etter Reform 97, Oslo: Pedagogisk forskningsinstitutt.
- Klette, Kirsti (2007): Bruk av arbeidsplaner i skolen – et hovedverktøy for å realisere tilpasset opplæring? *Norsk pedagogisk tidsskrift*, nr. 4.
- Klette, Kirsti, Svein Lie, Marianne Ødegaard, Øistein Anmarkrud, Nina Arnesen, Ole Kristian Bergem og Astrid Roe (2008): PISA+: Lærings- og undervisningsstrategier i skolen, Oslo: Norges forskningsråd.
- Kolen, Michael J. og Robert L. Brennan (2004): *Test Equating, Scaling, and Linking. Methods and Practices*, New York: Springer.
- Lie, Svein, Marit Kjærnsli, Astrid Roe og Are Turmo (2001): Godt rustet for fremtida? Norske 15-åringers kompetanse i lesing og realfag i et internasjonalt perspektiv, Oslo: Institutt for lærerutdanning og skoleutvikling, Universitetet i Oslo.
- Martin, Michael O., Ina V.S. Mullis og Pierre Foy, P. (2008): TIMSS 2007 International Science Report: Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades, Chestnut Hill, MA: TIMSS og PIRLS International Study Center, Boston College.
- Mullis, Ina V.S., Michael O. Martin og Pierre Foy (2008): TIMSS 2007 International Mathematics Report: Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades, Chestnut Hill, MA: TIMSS og PIRLS International Study Center, Boston College.
- Mullis, Ina V.S., Michael O. Martin, Ann M. Kennedy og Pierre Foy (2007): IEA's Progress in International Reading Literacy Study in Primary School in 40 Countries, Chestnut Hill, MA: TIMSS og PIRLS International Study Center, Boston College.

Mullis, Ina V.S., Michael O. Martin, Graham J. Ruddock, Christine Y. O` Sullivan, Alka Arora og Ebru Erberber (2005): TIMSS 2007 Assessment Frameworks, Chestnut Hill, MA: TIMSS og PIRLS International Study Center, Boston College.

Møen, Jarle og Martin Tjelta (2005): Bruker ulike høgskoler karakterskalaen ulikt? En analyse av sammenhengen mellom skolebakgrunn og faglig suksess ved NHH, Økonomisk Forum nr. 6.

OECD (2006): Assessing Scientific, Reading and Mathematical Literacy A Framework for PISA 2006, Paris: OECD.

OECD (2007): PISA 2006: Science Competencies for Tomorrow's World. Paris: OECD Publications.

Olsen, Rolf Vegar, Are Turmo og Svein Lie (2001): Learning about students' knowledge and thinking in science through large-scale quantitative studies, European Journal of Psychology of Education, 16, 401-418.

Rasch-Halvorsen, Anne og Håvard Johnsråten (2007): Norsk matematikkråds undersøkelse høsten 2007. En undersøkelse av grunnleggende matematisk kunnskap for studenter som begynner på matematikkrevende studier i Norge, Notodden: Norsk matematikkråd.

Schreiner, Camilla og Svein Sjøberg (2006): JEG VELGER MEG NATURFAG! (... HVEM GJØR EGENTLIG DET?) En studie av ungdoms prioriteringer ved valg av utdanning og yrke. Basert på ROSE (The Relevance of Science Education), en internasjonal studie av 15-åringers oppfatning av naturvitenskap og teknologi, Oslo: Norges forskningsråd.

Sjøberg, Svein (2007): Internasjonale undersøkelser: «Grunnlaget for norsk utdanningspolitikk?», i Halvard Hølleland (red.): På vei mot Kunnskapsløftet: begrunnelser, løsninger og utfordringer, Oslo: Cappelen Akademisk Forlag.

Skaar, Karl, Tor Inge Viblemo og Einar M. Skaalvik (2008): Se den enkelte. Analyse av Elevundersøkelsen 2008, Kristiansand: Oxford Research.

Turmo, Are (2003): Naturfagdidaktikk og internasjonale studier, Dr.scient.-avhandling, Universitetet i Oslo.

Udir (2008a): Nasjonalt kvalitetsvurderingssystem for grunnsopplæringen. [http://www.udir.no/templates/udir/TM\\_Artikkel.aspx?id=1117](http://www.udir.no/templates/udir/TM_Artikkel.aspx?id=1117)

Udir (2008b): Elevene får ikke de timene de har krav på. [http://www.utdanningsdirektoratet.no/templates/udir/TM\\_Artikkel.aspx?id=3659](http://www.utdanningsdirektoratet.no/templates/udir/TM_Artikkel.aspx?id=3659)

UFD (2003a): Har kompetanseberetningen et grunnlag? En grunnlagsrapport fra Kompetanseberetningen 2003, Utdannings- og forskningsdepartementet.

UFD (2003b): Hvordan skapes vektløse verdier. En snarvei til kompetanseberetningen 2003, Utdannings- og forskningsdepartementet.

UHR (2006): Karakterbruk i UH-sektoren: Rapport fra en arbeidsgruppe oppnevnt av Universitets- og høgskolerådet, Oslo: UHR.

Valdermo, Odd (2007): Flerfaglighet i skole og lærerutdanning – oppskrytt og misforstått læringspotensial, Bedre skole nr. 2.

Väljjarvi, Jouni, Pirjo Linnakylä, Pekka Kupari, Pasi Reinikainen and Inga Arffman (2002): *The Finnish Success in PISA – and some Reasons Behind It*, Jyväskylä: University of Jyväskylä.

Van Daal, Victor, Ragnar Gees Solheim, Nina Nøttaasen Gabrielsen og Anne Charlotte Begnum (2007): *Norske elevers lese-innsats og leseferdigheter. Resultater for fjerde og femte trinn i den internasjonale studien PIRLS 2006*, Stavanger: Lesesenteret, Universitetet i Stavanger.