Marit Kjærnsli (f. 1955) er cand.scient. med hovedfag i realfagdidaktikk fra Universitetet i Oslo. Hun er norsk prosjektleder for PISA 2003 og PISA 2006. Siden 1990 har hun vært ansatt ved Institutt for lærerutdanning og skoleutvikling (ILS), Universitetet i Oslo og har vært med i den norske prosjektgruppa i TIMSS og PISA. Hun har flere års erfaring som lærer i skolen.

Svein Lie (f. 1941) har doktorgrad (dr. philos) i teoretisk kjernefysikk fra Universitetet i Oslo. Han har mange års erfaring som lærer i videregående skole der han særlig har undervist i matematikk og fysikk. Siden 1991 har han vært ansatt ved Institutt for lærerutdanning og skoleutvikling (ILS), Universitetet i Oslo, for å arbeide med de store internasjonale undersøkelsene TIMSS og PISA. I store deler av denne tiden har han vært leder for disse prosjektene. Han var også medlem av Kvalitetsutvalget for grunnopplæringen.

Are Turmo (f. 1973) har doktorgrad i realfagdidaktikk (dr.scient.) fra Universitetet i Oslo. Han har vært forsker ved Institutt for lærerutdanning og skoleutvikling (ILS), Universitetet i Oslo siden 1998. Forskningsområdet har hele tiden vært internasjonale, komparative storskalundersøkelser av skolelevers kompetanse i matematikk og naturfag: TIMSS, TIMSS Trends og PISA.

MARIT KJÆRNSLI

Institutt for lærerutdanning og skoleutvikling,
Universitetet i Oslo
marit.kjærnsli@ils.uio.no

ARE TURMO

Institutt for lærerutdanning og skoleutvikling,
Universitetet i Oslo
are.turmo@ils.uio.no

SVEIN LIE

Institutt for lærerutdanning og skoleutvikling,
Universitetet i Oslo
svein.lie@ils.uio.no

Kan elevene mindre enn før? Naturfagkompetanse i Norden i perioden 1995–2003

Abstract

This article presents and discusses changes in science competency in the period from 1995 to 2003 in the Nordic countries based on analyses of data from PISA and TIMSS. One central feature of the PISA and TIMSS surveys is to study changes in students' competency over time in absolute terms. Some items are kept secure from cycle to cycle, and based on these items it is possible to create trend data, i.e. to place students who have taken tests at different times on the same cognitive scale. This is done by the use of modern psychometrical methods. The results show a decline in Norway and Sweden, while a positive development is found in Finland.

INNLEDNING

Har skolelevers kompetanse i naturfag endret seg over tid? Kunne elevene mer før, eller er nye generasjoner av skoleelever stadig mer kompetente i faget? Denne typen spørsmål engasjerer naturlig nok fagfolk innen utdanning, men også ofte "folk flest". Imidlertid er utsagn om dette oftest basert på anekdotisk evidens av begrenset vitenskapelig verdi. I desember 2004 ble resultater fra to internasjonale komparative studier publisert; TIMSS og PISA. Både TIMSS og PISA søker å måle utvikling i naturfagkompetanse over tid. Dette gjøres ved at en kjerne av oppgaver går igjen fra undersøkelse til undersøkelse. Med utgangspunkt i disse oppgavene kan man ved hjelp av moderne skaleringsmetoder plassere elever som har tatt tester på ulike tidspunkter på samme prestasjonsskala. Slik kan man studere utvikling i faglig nivå over tid i absolutt forstand. Med TIMSS- og PISA-undersøkelsene som ble gjennomført i 2003 har vi for første gang denne typen absolutte

trenddata, henholdsvis tilbake til 1995 og 2000, for de nordiske land basert på internasjonale komparative studier. Denne artikkelen presenterer og diskuterer endringer i naturfagkompetanse fra 1995 til 2003 i de nordiske landene, basert på analyser av data fra PISA og TIMSS. Forskjeller i utviklingen for jenter og gutter belyses spesielt. Det gis også en introduksjon til det metodiske grunnlaget som gjør slike sammenlikninger mulige.

OM PISA OG TIMSS, OG STUDIENES DEFINISJON AV "NATURFAG"

Programme for International Student Assessment (PISA) og Trends in International Mathematics and Science Study (TIMSS) er to internasjonale komparative studier som begge måler skoleelevers kompetanse blant annet i naturfag på ulike trinn. Populasjonen i PISA er 15-åringer, mens TIMSS inkluderer to populasjoner, 9-åringer og 13-åringer. I denne artikkelen vil vi imidlertid ikke diskutere resultatene for 9-åringene. PISA ble første gang gjennomført i 2000 i regi av OECD, med 32 deltakende land, og gjennomføres hvert tredje år. TIMSS-studien har røtter tilbake til 1960-tallet hvor den første studien av naturfag i skolen ble påbegynt, First International Science Study (FISS). Undersøkelsene gjennomføres i regi av The International Association for the Evaluation of Educational Achievement (IEA). I TIMSS-studien i 2003 deltok omlag 50 land, og undersøkelsen vil bli gjentatt hvert fjerde år.

Det er både likheter og forskjeller mellom PISA- og TIMSS-undersøkelsene. PISA har utbyttet av den totale grunnutdanningen som fokus, mens TIMSS er laget for å gjøre det mulig å rapportere prestasjoner på forskjellige aldersnivåer og på den måten sammenlikne framgang med stigende alder for forskjellige land. PISA tester det man regner med kreves i framtiden, mens TIMSS tar utgangspunkt i en felles kjerne av læreplanelementer i de landene som deltar. TIMSS har som siktepunkt å måle så "rettferdig" som mulig det som undervises i naturfag i alle deltakerlandene. Det man vurderer i PISA, er ikke knyttet til en slik kjerne av felles læreplanelementer, men i stedet til et konsensus syn når det gjelder hva utdanningssystemet bør vektlegge for å forberede samfunnsborgerne for voksenlivet og livslang læring. PISA-oppgavene tar utgangspunkt i autentisk materiale (tekster), og man anvender grupper av oppgaver knyttet til samme materiale. TIMSS-undersøkelsen består derimot i hovedsak av enkeltstående oppgaver. Både TIMSS og PISA består av en serie undersøkelser, henholdsvis hvert fjerde år og hvert tredje år, noe som gir muligheten til å få sammenliknbare data fra undersøkelse til undersøkelse. Deltakerlandene har derfor mulighet til å bruke begge disse studiene til å studere tendenser over tid.

Rammeverket for naturfag i TIMSS har to organiserende dimensjoner, en *innholdsdimensjon* og en *kognitiv dimensjon* (Mullis m.fl., 2003). Begge disse har flere elementer som definerer henholdsvis de spesifikke naturfaglige fagområdene i undersøkelsen og den ulike kognitive atferden som forventes av elevene i møte med oppgavene. I tillegg inneholder rammeverket en tredje dimensjon, nemlig naturvitenskapelige arbeidsmetoder. En del oppgaver i TIMSS søker spesielt å vurdere elevenes kompetanse knyttet til naturvitenskapelige arbeidsmetoder. Alle disse oppgavene er imidlertid også klassifisert etter både innholdsdimensjonen og den kognitive dimensjonen. Innholdsdimensjonen i naturfag i TIMSS består av fem fagområder for 8. klasse: biologi, kjemi, fysikk, geofag og miljølære. For 4. klasse er fysikk og kjemi slått sammen, mens miljølære ikke er et eget fagområde. Rammeverket i TIMSS er svært detaljert når det gjelder hva man ønsker å måle innenfor de ulike faglige emneområdene.

Tilsvarende detaljgrad finner man ikke i PISA-studien. Rammeverket i PISA inneholder også både en innholdsdimensjon og en kognitiv dimensjon, men legger større vekt på den kognitive dimensjonen. Man tenker seg at spesifikke kognitive ferdigheter kan utøves i ulike faglige sammenhenger og kontekster, men akkurat *hvilke* områder og kontekster, er ikke like viktig. Eksempler på slike kognitive ferdigheter kan være å identifisere hvilke spørsmål som kan besvares av naturvitenskap, og å kunne skille mellom objektive observasjoner og subjektive meninger. På denne måten har

TIMSS og PISA noe ulik tilnærming til naturfaget. I PISA 2003 er det for få oppgaver til at man kan lage flere måleskalaer basert på den kognitive dimensjonen, noe som imidlertid er planlagt i PISA 2006, hvor naturfag vil ha hovedfokus.

PISA-studiens definisjon av naturfagkompetanse tar utgangspunkt i begrepet *scientific literacy*. Med utgangspunkt i den generelle kompetansedefinisjonen i PISA og med innspill fra alle deltakerlandene har en ekspertgruppe i naturfag kommet fram til følgende definisjon av begrepet *scientific literacy*:

Scientific literacy is the capacity to use scientific knowledge, to identify questions and to draw evidence-based conclusions in order to understand and help make decisions about the natural world and the changes made to it through human activity. (OECD, 2003, s. 133)

I PISAs definisjon av *scientific literacy* brukes naturfaglig kunnskap på en måte som innebærer mye mer enn kunnskap om fakta, navn og begreper. Det inkluderer en forståelse av fundamentale naturfaglige begreper og hvilke begrensninger naturfaglig kunnskap har. Å gjenkjenne spørsmål som kan besvares ved naturvitenskapelige metoder innebærer kunnskap om naturvitenskapens egenart, så vel som kunnskap om naturfaglige sider knyttet til spesifikke temaer. Å trekke konklusjoner basert på data betyr å kjenne til og anvende metoder knyttet til utvalg og vurdering av informasjon og data. Det innebærer også å gjenkjenne tilfeller som mangler tilstrekkelig informasjon til at man kan trekke endelige konklusjoner.

Det er et viktig poeng at PISA og TIMSS måler i henhold til gitte definisjoner av faglig kompetanse, og at elevene selv sagt kan tilegne seg kunnskaper og ferdigheter som ikke inngår i disse målingene. Til sammen dekker imidlertid PISA og TIMSS et svært bredt spekter av faglig kompetanse, både innholdsmessig og når det gjelder de kognitive prosessene som er involvert.

KVALITETSSIKRING I PISA OG TIMSS

For å få gode og sammenliknbare data er det i internasjonale studier som PISA og TIMSS lagt stor vekt på å kvalitetssikre alle prosedyrer; som utvalg av skoler og elever, oversetting av oppgaver og spørsmål, kontakt med skoler, gjennomføring av undersøkelsen i skolene, vurdering av åpne oppgaver samt innskriving av data. Alt dette er godt dokumentert i internasjonale tekniske rapporter (se Adams & Wu, 2002; Martin, Mullis & Chrostowski, 2004). Ikke minst er det viktig å sikre ekvivalente versjoner av de faglige oppgavene på ulike språk.

I internasjonale undersøkelser har man erfart at språklige forskjeller og kulturelle tilpasninger har vært et problem når det gjelder å framskaffe pålitelig informasjon, se for eksempel Grisay (2003). Samme tekst kan oversettes på ulike måter og likevel være "riktig" oversatt, men ulike oversettelser kan gjøre vanskelighetsgraden litt forskjellig. For å minimalisere disse problemene og for å imøtekomme mye av den kritikken som har vært reist mot komparative studier når det gjelder oversetting, har det vært svært strenge krav til prosedyrene for oversetting. I PISA-undersøkelsen utvikles alle oppgavetekster og spørsmål både på engelsk og fransk. Alle land må ha to uavhengige oversettelser som sammenfattes og deretter kontrolleres av språkeksperter. Oversettelsen blir også spesielt sjekket i tilfeller der det viser seg at en oppgave relativt sett faller vanskeligere eller lettere ut i det aktuelle landet.

For å kunne studere utvikling i kompetanse over tid beholdes noen av oppgavene uendret fra undersøkelse til undersøkelse. Basert på disse oppgavene kan man ved hjelp av "test-equating" plassere elever som har tatt tester på ulike tidspunkter på samme prestasjonsskala. Dette kommer vi nærmere tilbake til.

NORDISK DELTAKELSE I TIMSS OG PISA

Alle de nordiske landene har deltatt i en eller flere av PISA- og TIMSS- studiene, men graden av nordisk deltakelse har variert. Tabell 1 gir en oversikt over hvilke nordiske land som har deltatt i de ulike studiene. Finland var det eneste nordiske landet som deltok i TIMSS i 1999, så denne undersøkelsen vil derfor ikke bli omtalt ytterligere i denne artikkelen.

Tabell 1 viser at Norge og Sverige er de eneste nordiske landene som har deltatt i samtlige fire studier, mens Danmark og Island har deltatt i tre og Finland i to. Med andre ord vil vi kunne si mest om norske og svenske elevers kompetanseutvikling i naturfag. Videre ser vi at alle de nordiske landene har deltatt i begge PISA-undersøkelsene.

Tabell 1: Nordisk deltakelse i TIMSS og PISA

TIMSS		PISA	
1995	2003	2000	2003
Norge	Norge	Norge	Norge
Sverige	Sverige	Sverige	Sverige
Danmark		Danmark	Danmark
Island		Island	Island
		Finland	Finland

METODISK GRUNNLAG

Hva er grunnlaget for at man kan si noe absolutt om utvikling i naturfagkompetanse over tid i PISA- og TIMSS- studiene? Svaret er å finne i den testteoretiske metoden som internasjonalt kalles "test equating", og det dreier seg om en metode for å sammenlikne resultater fra to forskjellige prøver. Her skal vi si litt om prinsippet for dette.

Det er lett å skjønne at resultater i form av for eksempel antall riktige svar på to forskjellige prøver i utgangspunktet ikke kan fortelle noe om en eventuell fram- eller tilbakegang i et fagområde. For det første vil det kreve at innholdet på prøvene i samme grad er representativt for fagområdet. Og for det andre, og viktigere i denne sammenheng, kan ikke antall riktige svar være noe 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 ikke mulig å si om det er fordi elevene er blitt "bedre" eller om oppgavene er blitt "lettere".

I moderne testteori (Item Response Theory, IRT, hvorav såkalt Rasch-modell er et eksempel) er dette nettopp utgangspunktet (se for eksempel Allerup 1987; Lie, Kjærnsli, Roe & Turmo, 2001). Elevers dyktighet og oppgavers vanskelighetsgrad måles da langs samme måleskala, og i Rasch-modell er det *differansen* mellom dyktigheten til eleven og vanskelighetsgraden til oppgaven som bestemmer sannsynligheten for at eleven skal løse oppgaven riktig. I Rasch-modellen er sannsynligheten for et riktig svar gitt matematisk som en logistisk stigende funksjon av den nevnte differansen. Ved at et sett av oppgaver gis til en gruppe elever, kan man ut fra en tilpasning av empiriske fordelinger av riktige svar til de logistiske kurvene bestemme hver elevs dyktighet og hver oppgaves vanskelighetsgrad samtidig. Dette innebærer at hver oppgave som er brukt i en undersøkelse, ut fra resultatene får en kjent vanskelighetsgrad. Viktig er det imidlertid at det her bare er snakk om *relative* vanskelighetsgrader, hvor vanskelig oppgavene er sammenliknet med hverandre. Det

ligger en tilfeldighet i normeringen, idet den gjennomsnittlige vanskelighetsgraden til et datasett kan velges fritt. For en gruppe som er mer kompetent, vil oppgavene *fortune* seg som "lettere", men det er det samme som å si at elevene er dyktigere og derfor skårer høyere.

Så lenge den *samme* prøven blir brukt ved to anledninger, kan resultatene selvsagt lett sammenliknes, men på grunn av læringseffekten fra den første prøven vil det være vanskelig å knytte framgangen entydig til økt kompetanse. Hvis man prøver å måle individuelle elevers framgang, vil elevene på den andre prøven kunne huske svarene eller framgangsmåten fra den første gjennomføringen. Hvis det derimot er nye elever, slik som i TIMSS og PISA, vil læringseffekten ligge i at elevene kunne kjenne til oppgavene på forhånd, og det er vanskelig å unngå så lenge man legger vekt på å offentliggjøre i hvert fall en god del av oppgavene for at andre kan få innsyn i prøven. Poenget med "test equating" er at resultatene på to *forskjellige* prøver kan gis langs *samme* skala for dyktighet, og da kan ikke bare prestasjonene, men også kompetansene sammenliknes. For å få til dette er det imidlertid nødvendig at vanskelighetsgradene kan sammenliknes ved at i hvert fall *noen* av oppgavene forekommer i begge prøvene. Vanligvis gjøres dette som i TIMSS og PISA ved at et noen få oppgaver fra den første prøven holdes hemmelig for senere å brukes også i prøve nummer to. Ved hjelp av disse "link-oppgavene" 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.

Man forutsetter da at Rasch-modellen passer med dataene, noe som innebærer at alle oppgavene representerer en måling av det samme "latente trekket", nemlig den aktuelle fagkompetansen. Dette må vise seg ved at link-oppgavene fungerer på samme måte i de to prøvene. Alle vanskelighetsgradene til disse oppgavene kan godt ha endret seg (for eksempel i form av prosent riktige svar), men de må i tilfelle endre seg omtrent like mye. Ved en justering av skalaen for vanskelighetsgrad, og dermed dyktighet, sørger man for at vanskelighetsgraden for *samme* oppgave er lik i begge tilfellene. Og dermed kan resultatene på hele den nye prøven måles langs samme skala som den første, noe som muliggjør en måling av framgang eller tilbakegang over tid.

I figur 1 er prinsippet for test equating illustrert. De to linjene representerer to ulike prøver med hver sin skala for vanskelighetsgrad, og dermed dyktighet. For hver prøve kan de aktuelle skaleverdier langs aksene forskyves vilkårlig. Figuren viser hvordan man kan forskyve for eksempel verdiene langs den "nederste" prøven et stykke mot venstre slik at link-oppgavene (sirklene på figuren) kommer mest mulig overens. Men figuren viser også et eksempel på en oppgave (fylt sirkel) som ikke fungerer godt nok for dette formålet og derfor må ses bort fra i denne prosessen. Tydeligvis er det noe med den spesielle oppgaven som har fungert forskjellig for elevene i de to prøvesituasjonene. Oppgaven fortøner seg relativt sett lettere på den siste prøven, siden oppgaven er forskjøvet noe til venstre i forhold til de andre link-oppgavene. For eksempel kan det være fordi den relaterer seg sterkt til et spesifikt emne som er mer sentralt for elevene i det siste tilfellet.

I prinsippet er dette enkelt nok, men i praksis er det mange krav som må oppfylles og problemer som må løses. Spesielt krevende er det at prosedyren må fungere bra i hvert eneste deltakerland for at trend-dataene skal være pålitelige mål for å sammenliknes internasjonalt. Dersom det er få link-oppgaver, blir usikkerheten ved tilpassing av skalaene stor. Et spesielt problem oppstår når oppgavene forekommer på ulike steder i oppgaveheftene, idet en oppgave på slutten av et hefte får lavere løsningsfrekvens og dermed fortøner seg som "vanskeligere". Slike forhold må det justeres for.

Bruk av link-oppgaver er den viktigste metoden for test-equating. Men det går også an å gjøre det på andre måter. For eksempel kan man fra to helt forskjellige prøver sette sammen en tilleggsprøve med oppgaver fra hver av dem. Ut fra de målte vanskelighetsgradene til oppgavene i tilleggsprøven kan så de to opprinnelige prøvene indirekte sammenliknes.

Figur 1: Illustrasjon av linkoppgaver (O) som brukes sammen med de andre oppgavene (kvadrater) for å kalibrere to forskjellige prøver til samme skala. De horisontale linjene representerer skala for vanskelighetsgrad for hver av de to prøvene. "Link-oppgavene" er knyttet sammen med linjer.

EMPIRISKE RESULTATER

Figur 2: Endringer av prestasjoner i naturfag for de nordiske landene i PISA og TIMSS

Fordi mange av oppgavene fra PISA 2000 også er brukt i PISA 2003, kan man ved hjelp av "test-equating" plassere elever som har tatt testene i henholdsvis 2000 og 2003 på samme prestasjonsskala. Tilsvarende gjelder for TIMSS-undersøkelsen. Skalaene i begge undersøkelsene er standardisert ved at gjennomsnitt for elevskåre er satt til 500 med standardavvik på 100. Gjennomsnittsskåre for de to studiene kan imidlertid ikke sammenliknes direkte. Dette skyldes at gjennomsnittet i PISA er beregnet ut fra de OECD-landene som deltok, mens gjennomsnittet i TIMSS er beregnet ut fra alle deltakerlandene.

Figur 2 viser resultatene for de nordiske landene. Som nevnt tidligere, deltok alle de nordiske landene både i PISA 2000 og PISA 2003, mens det bare var Norge og Sverige som både deltok i TIMSS 1995 og TIMSS 2003. I TIMSS er det bare resultatene for populasjonen av 13-åringer som er vist. I Sverige testet de imidlertid elever som var ett år eldre, altså 14-åringer. I framstillingen er det lagt vekt på å vise endringer innad for hvert land ved å illustrere differansen for gjennomsnittsskåre mellom den første og siste undersøkelsen. Den tidligste undersøkelsen er konsekvent markert med en fylt firkant enten det gjelder PISA eller TIMSS, mens resultater fra 2003 er markert med en åpen sirkel. Resultatene fra PISA og TIMSS er gitt i samme målestokk slik at endringene også kan sammenliknes visuelt.

Konfidensintervaller for gjennomsnittsverdiene er ikke vist på figurene, men vi oppgir i stedet følgende: I PISA er Norges sterke tilbakegang statistisk signifikant ($p < 0,01$), likeledes den finske

framgangen ($p < 0,05$). Endringen for de andre landene er ikke signifikante. Men selv om endringen for ett og ett av disse landene ikke er signifikant, er likevel den samlede tendensen for disse landene signifikant. I TIMSS er den norske og svenske tilbakegangen klart signifikant ($p < 0,01$).

Figur 2 viser at det bare er de finske elevene som viser framgang fra PISA 2000. Dette ser vi til tross for at de presterte blant de beste i PISA 2000. De islandske elevene presterer likt i begge undersøkelsene, mens det i de tre andre nordiske landene er tendens til tilbakegang, og aller størst er den i Norge. I TIMSS er det lengre tidsintervaller, og det er også her man finner de store endringene, noe som gjelder både de svenske og norske elevprestasjonene. Når man skal diskutere resultatene i TIMSS, er det for øvrig viktig å ta hensyn til at det er betydelige aldersforskjeller mellom elever i ulike land. Sverige har testet elever som er ett år eldre enn de norske, så i denne sammenhengen er det kun de to differansene innad i hvert land som er interessante å sammenlikne. Internasjonalt er Norge og Sverige de to landene som har hatt størst tilbakegang fra TIMSS 1995, og vi ser at de svenske elevene har hatt noe større tilbakegang enn de norske.

For Norge svarer tilbakegangen omtrent til at elevene nå ligger et halvt år "etter" elevenes dyktighet i 1995. Tilsvarende resultater fant vi for øvrig også for de norske elevene fra 4. klasse som deltok. Resultatene viste langt større tilbakegang enn i alle de andre deltakerlandene (Grønmo, Bergem, Kjærnsli, Lie & Turmo, 2004). En tilbakegang på nesten 40 poeng svarer omtrent til at elevene i 4. klasse i 2003 er bortimot "et helt år svakere" enn de jevnaldrende i 3. klasse i 1995. Sverige deltok ikke i denne populasjonen.

Figur 3: Endringer av jenters og gutters prestasjoner fra PISA 2000 til PISA 2003

Både i PISA 2000 og PISA 2003 er det totalt sett relativt små forskjeller mellom jentenes og guttenes prestasjoner i naturfag, og de er ubetydelig i forhold til dem man finner i lesing (se OECD, 2001, 2004). Figur 3 viser endringer i jentenes og guttenes prestasjoner i de to undersøkelsene, og i likhet med framstillingen i figur 2, er PISA 2000 markert med en fylt firkant, mens PISA 2003 er markert med en åpen sirkel.

Figur 3 viser tydelig at både jentenes og guttenes prestasjoner er forbedret i Finland, og det er omtrent like stor framgang for begge kjønn. Videre ser vi at de norske jentene har større tilbakegang enn de norske guttene. Den samme tendensen ser vi både i Sverige og Danmark, men den er sterkest i Norge.

Figur 4: Endringer av jenters og gutters prestasjoner fra TIMSS 1995 til TIMSS 2003

I figur 4 er tilsvarende resultater vist for TIMSS. Bildet viser tydelig en tilbakegang for både jentenes og guttenes prestasjoner både i Sverige og Norge. Men i motsetning til hva vi så for PISA-resultatene, ser vi her at det er større tilbakegang for guttenes prestasjoner. Igjen er det viktig å huske at disse endringene strekker seg over lengre tid enn i PISA.

KONKLUSJON

Vi kan sammenfatte resultatene med å konstatere at det er en tydelig generell tendens til at naturfagkunnskapene, slik disse måles i TIMSS og PISA, i Sverige og Norge er på retur, mens finske elever har ytterligere styrket sin stilling fra et allerede høyt nivå. For Island er det ubetydelige forskjeller å spore. Målingene i TIMSS framstår som særlig betenkelig, siden Norge og Sverige framstår som de to deltakerlandene (med omtrent samme populasjon de to årene) som har sterkest tilbakegang. I PISA har Norge den nest sterkeste tilbakegangen av alle landene. For Danmarks og Sveriges vedkommende er det også målt en tilbakegang i PISA, men den er ikke signifikant.

I Finland har begge kjønn gått tilnærmet like mye fram fra 2000 til 2003 i PISA, mens jentene har gått mest tilbake i Danmark, Norge og Sverige. Guttene har gått mest tilbake i Norge og Sverige i TIMSS. Forklaringer på disse forskjellene mellom de to studiene kan ligge både i ulike tidsspenn og i ulike definisjoner av naturfag, samt at studiene kartlegger ulike alderstrinn.

Et nærliggende og viktig spørsmål er hva som kan ligge bak de tydelige negative tendensene i de skandinaviske landene, særlig Norge og Sverige. Data fra undersøkelsene kan ikke gi noe entydig svar på dette, men det er grunn til å trekke fram noen mulige forklaringer. I de norske rapportene fra TIMSS og PISA stilte vi spørsmålet om man generelt har beveget seg fra undervisning til aktivitet i stedet for fra undervisning til læring i norsk grunnskole. Med den nye læreplanen for grunnskolen som kom i 1997, L97, har det skjedd en endring av undervisningsformer. I L97 understrekes det: "(...) at elevane skal være aktive, handlande og sjølvstendige. Dei skal få lære ved å gjere, utforske og prøve ut i aktivt arbeid fram mot ny kunnskap og erkjenning" (L97, s. 75). At elevene skal være aktive, er ofte tolket som å drive med ulike aktiviteter av typen gruppearbeid,

prosjektarbeid, lek og eksperimenter. Faren ved å fokusere så sterkt på spesielle arbeidsmetoder er at de faglige læringsmålene kan bli nedprioritert. Bruk av ulike læringsaktiviteter synes å ha preg av å være mål i seg selv uten at de relateres til klare læringsmål:

”Generelt sitter vi imidlertid med et inntrykk av at det er lite systematisk og oppsummert refleksjon rundt de ulike aktivitetenes læringspotensiale, hvilket igjen bidrar til at elevene vanskelig kan akkumulere kunnskap basert på systematiske erfaringer. Det faktum at det brukes lite tid til avrundning og oppsummering av de ulike aktivitetene bidrar videre til at de ulike aktivitetenes intensjoner blir uklare for elevene, og det etableres en svak relasjon mellom å gjøre noe og å lære noe.” (Klette, 2003, s. 73)

Sitatet fra Klette (2003) ovenfor tyder på at det ofte mangler en oppsummering fra lærerens side som kan strukturere lærestoffet for elevene etter en aktivitetsøkt. ”Learning by doing” er et slagord som L97 langt på vei prøver å realisere. Men problemet oppstår når elevene er overlatt til seg selv med å konstruere sin kunnskap ut fra et mylder av erfaringsbiter. Det kan være et problem at de ulike aktivitetene ofte blir gjennomført isolert fra den øvrige undervisningen og ikke blir satt inn i en helhetlig faglig sammenheng med klart definerte læringsmål. Uten forklarende oppsummering kan resultatet av mye ”doing” lett bli ”confusion” istedenfor ”learning”.

I den svenske TIMSS-rapporten (Skolverket, 2004a) får også tilbakegangen fra 1995 behørig omtale. Dels peker rapporten på at en grundig analyse av oppgavene og deres relevans i forhold til læreplanen vil være en vesentlig utfordring før man kan si mye om årsaker og sammenhenger. Men særlig interessant er at det refereres til en nasjonal vurdering av blant annet naturfag i svensk skole i 2003, og der ble det påvist en betydelig nedgang i kunnskaper siden 1992. Riktignok var nedgangen for omtrent like gamle elever ikke like stor som i TIMSS, og resultatene er ikke direkte sammenliknbare, men det er likevel en tydelig sammenfallende tendens. Den svenske rapporten går for øvrig ikke inn på noen diskusjon om mulige pedagogiske prosesser i klasserommene som kan ligge bak denne tendens. I den svenske PISA-rapporten (Skolverket, 2004b) sier de at selv om det totalt sett ikke er signifikant tilbakegang i Sverige, så viser resultatene urovekkende tendenser i naturfag. Dette er også i tråd med deres nasjonale vurdering fra 2003 som viser tilbakegang i fysikk og kjemi. En mulig forklaring som trekkes fram, er den sterke betoningen av fagene svensk, matematikk og engelsk som har kommet i kjølvannet av det nåværende vurderingssystemet som ble innført i 1998. Forfatterne påpeker det vesentlige i at alle elever får gode, eller i det minste tilstrekkelige, kunnskaper i kjerneemnene svensk, matematikk og engelsk. Likevel må man spørre seg om dette med nødvendighet må skje på andre fags bekostning.

Den danske tilbakegangen i PISA er som nevnt ikke stor, og i den danske PISA-rapporten (Andersen & Sørensen, 2004) kommenteres den bare kort med at den ikke er signifikant. Det er likevel grunn til å peke på at situasjonen i Danmark må ses i lys av at prestasjonene tidligere har ligget og fortsatt ligger lavere enn i de andre nordiske landene.

I en studie av prestasjoner oppgave for oppgave for alle landene i TIMSS 1995 (Grønmo, Kjærnsli & Lie, 2004) ble det tydelig demonstrert at profilene for prestasjoner var påfallende like mellom Norge og Sverige, og at også danske og (i litt mindre grad) islandske data knyttet seg nært til disse. Dette ble tolket som et tegn på at vi kan snakke om en nordisk profil for kompetanse i naturfag. En tilsvarende analyse av data fra PISA 2000 (Kjærnsli & Lie 2004) var basert på mange færre oppgaver, siden naturfag bare var et sekundært fag. Men likevel ble de sterke likhetene mellom de nordiske landene langt på vei bekreftet, med ett viktig unntak: De finske elevenes kompetanse framsto som svært forskjellig fra sine nordiske medelever. I den grad vi kan snakke om en nordisk, eventuelt en skandinavisk versjon av naturfagutdanning i grunnskolen, synes denne ut fra data fra TIMSS og PISA å være i en alvorlig situasjon. Sett med internasjonale øyne er tilbakegangen påtakelig og blir behørig kommentert i internasjonale rapporter. Og det er etter vår mening stor grunn til å ta slike utfordringer alvorlig. Det er lite data i PISA og TIMSS selv som gir en entydig

forklaring på tilbakegangen, men i siste kapittel i våre nasjonale rapporter har vi likevel trukket fram flere mulige forklaringer som vi ikke kan komme inn på her, som for eksempel endrede lærer- og elevroller, noe som igjen kan henge sammen med endringer i samfunnet generelt, liten vekt på læringsstrategier samt problematisk arbeidsmiljø i klasserommet (Grønmo m.fl., 2004, Kjærnsli, Lie, Olsen, Roe & Turmo, 2004).

Vi håper denne artikkelen kan gi grunnlag for en debatt videre om nordiske elevers kunnskaper i naturfag, og en slik debatt mener vi det er viktig og nødvendig å ta.

REFERANSER

- Adams, R. & Wu, M. (2002). *PISA Technical Report*. Paris: OECD Publications.
- Allerup, P. (1987). *Raschmodeller – prinsipper og anvendelse*. Danmarks Pædagogiske Institut
- Andersen, A.M. & Sørensen, H. (2004). Naturvidenskapelig kompetence. I J. Mejdning (red.), *PISA 2003- Danske unge i en international sammenligning*. København: Danmarks Pædagogiske Universitets Forlag.
- Grisay, A. (2003). Translation procedures in OECD/PISA 2000 international assessment. *Language Testing*, 20(3), 225-240.
- Grønmo, L.S., Bergem, O.K., Kjærnsli, M., Lie, S. & Turmo, A. (2004). *Hva i all verden har skjedd i realfagene? Norkse elevers prestasjoner i matematikk og naturfag i TIMSS 2003*. Acta Didactica 5/2004. Institutt for lærerutdanning og skoleutvikling, Universitetet i Oslo.
- Grønmo, L.S., Kjærnsli, M. & Lie, S. (2004). Looking for cultural and geographical factors in patterns of responses to TIMSS items. I Papanastasiou, C. (red.), *Proceedings of the IEA International Research Conference-2004 . Vol. 1*. Kypros: Cyprus University Press.
- Kjærnsli, M., Lie, S., Olsen, R. V., Roe, A. & Turmo, A. (2004). *Rett spor eller ville veier? Norske elevers prestasjoner i matematikk, naturfag og lesing i PISA 2003*. Oslo: Universitetsforlaget
- Kjærnsli, M. & Lie, S. (2004). PISA and Scientific Literacy: similarities and differences between the Nordic countries. *Scandinavian Journal of Educational Research*, 48(3), 271-286.
- Klette, K. (2003). Lærerens klasseromsarbeid; Interaksjon og arbeidsformer i norske klasserom etter Reform 97. I K. Klette (red.), *Klasserommets praksisformer etter Reform 97* (ss. 39-76). Oslo: Pedagogisk forskningsinstitutt, Universitetet i Oslo.
- Martin, M.O., Mullis, I.V.S. & Chrostowski, S.J. (red.) (2004). *TIMSS 2003 Technical Report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center.
- Mullis, I.V.S., Martin, M.O., Smith, T.A., Garden, R.A., Gregory, K.D., Gonzales, E.J., Chrostowski, S.J. & O'Connor, K.M. (2003). *TIMSS Assessment Frameworks and Specifications 2003*. Boston: International Study Center, Lynch School of Education, Boston College.
- OECD (2001). *Knowledge and Skills for Life- First Results from PISA 2000*. Paris: OECD Publications.
- OECD (2003). *The PISA 2003 Assessment Framework*. Paris: OECD Publications.
- OECD (2004). *Learning for Tomorrow's World- First Results from PISA 2003*. Paris: OECD Publications.
- Skolverket (2004a). *TIMSS 2003. Svenska elevers kunnskaper i matematikk och naturvetenskap i skolor 8 i ett nationellt och internationellt perspektiv*. Stockholm: Rapport 255, Skolverket.
- Skolverket (2004b). *PISA 2003. Svenska femtonåringars kunnskaper och attityder i ett internationellt perspektiv*. Stockholm: Rapport 254, Skolverket.