

1

Tjue år med internasjonale skoleundersøkelser i Norge: Bakgrunn, læringspunkter og veien videre

ROLF VEGAR OLSEN OG JULIUS KRISTJAN BJÖRNSSON

SAMMENDRAG TIMSS 1995 kan regnes som den første internasjonale undersøkelsen av sitt slag, og i de mer enn 20 årene som har gått, har vi sett en framvekst av flere og mer varierte internasjonale undersøkelser. I dette kapitlet gir vi en bakgrunn for resten av boka ved å beskrive noen typiske kjennetegn ved de internasjonale undersøkelsene, og vi beskriver hvordan disse undersøkelsene spesielt legger til rette for to typer sammenlikninger – mellom land og over tid; to referanserammer som på hver sine måter gir nyttige linser for å tolke resultater. Vi presenterer også det vi mener er de mest sentrale funnene for norsk skole gjennom disse 20 årene, og på tvers av de ulike undersøkelsene. Sammenlikninger over tid viser at norsk skole har vært preget av en nedgang etterfulgt av en oppgang, og vi ser at læringsmiljø og skoleklima rapporteres å være betydelig bedre enn tidligere. Sammenlikninger med andre viser at norsk skole er preget av relativt små forskjeller: mellom skoler, mellom kjønn (med lesing som unntak) og mellom elever med ulik hjemmebakgrunn. Vi avrunder kapitlet med å foreslå hvordan undersøkelsene kan endre design for å styrke gyldigheten til sammenlikninger innenfor begge disse to referanserammene.

SUMMARY TIMSS 1995 may be regarded as the first international large-scale assessment (ILSA) of its kind. Since then we have witnessed a growth both in terms of the number and variation of ILSAs. This chapter seeks to lay a foundation for the rest of the book by describing some of the key features of ILSAs. Furthermore, we describe how ILSAs facilitate two fundamentally different types of comparisons: between countries and over time. These two frames of reference give different lenses for interpreting the results from ILSAs. Having done that, we present some of the main findings for the Nor-

wegian school system over this period. Comparisons over time show that Norwegian students' performance declined in the first half, and increased with (almost) the same amount over the last half of the period. In addition, over time Norwegian students report more positively on questions relating to the learning environment and school climate. Comparisons with other countries reveal that the Norwegian school system is characterized by relatively small differences: between schools, between genders (except for reading), and between students with different home backgrounds. The chapter ends with suggestions for how the studies over the next years may change or tweak parts of their design to strengthen the validity of claims based on the two frames of reference

OPPSTARTEN I 1995

I 2015 ble både TIMSS og PISA gjennomført i et stort antall land for sjette gang. Internasjonale storskalastudier av denne typen har røtter tilbake til 1960-tallet. Det som den gangen startet som et samarbeid mellom forskere fra flere land, førte til slutt til etableringen av organisasjonen The International Association for the Evaluation of Educational Achievement (IEA) (Husén & Tuijnman, 1994). Dette er i dag en organisasjon hvor både akademiske institusjoner og representanter for skolemyndigheter er medlemmer.¹ Til tross for at røttene til disse undersøkelsene går nesten 60 år tilbake i tid,² er det likevel naturlig å regne 1995 som startpunktet for denne typen undersøkelser slik vi kjenner dem i dag. Dette året ble TIMSS³ for første gang gjennomført i Norge (Lie, Kjærnsli, & Brekke, 1997) og 44 andre land.⁴ I tillegg til at dette var den første studien med et så stort antall deltakende land, kan noen viktige kjennetegn ved denne – og de påfølgende studiene – listes opp:

- ▶ Detaljerte rammeverk som beskriver hva som skal måles, og hvordan undersøkelsen skulle gjennomføres (Robitaille & Garden, 1996; Robitaille & Maxwell, 1996; Robitaille et al., 1993).

1. Danmark og Finland er eksempelvis representert gjennom de akademiske institusjonene Danmarks Pedagogiske Universitet og Universitetet i Jyväskylä, mens Norge og Sverige er representert ved at Utdanningsdirektoratet og Skolverket er medlemmer.
2. For en kort historisk oversikt, se også <http://www.iea.nl/brief-history-iea-more>
3. I 1995 sto akronymet for Third International Mathematics and Science Study. I senere undersøkelser er akronymet beholdt, men det står nå for Trends in International Mathematics and Science Study
4. Alle de internasjonale rapportene som er utgitt fra TIMSS, finnes på <https://timssand-pirls.bc.edu>, og tilsvarende for PISA finnes på <http://www.oecd.org/pisa>

- ▶ Utvalg basert på godt definerte populasjoner, grundig kvalitetssikring av instrumenter og data med anvendelse av avanserte og robuste psykometriske metoder (Martin & Kelly, 1997a, 1997b).
- ▶ Åpen tilgang til data, inkludert grundig teknisk dokumentasjon og hvordan data fra undersøkelsen bør analyseres (Gonzalez & Smith, 1997).
- ▶ Etableringen av en ny type samarbeid mellom politikktutformere og forskere som også førte til at disse og andre tilsvarende undersøkelser har fått stor oppmerksomhet i media og det offentlige ordskiftet om kvaliteten i utdanningen (Olsen, 2005).

TIMSS-studien i 1995 ble fulgt opp av tilsvarende studier hvert fjerde år. I tillegg har IEA også etablert en rekke andre studier med tilsvarende utforming og design knyttet til andre fagområder. (Se tabell 1.1 for en oversikt over studier.) En viktig grunn til at resultater fra IEA-undersøkelsene har blitt lagt merke til, er at Organisation for Economic Co-operation and Development (OECD) ganske tidlig ble en aktiv bruker av dataene. OECD er en organisasjon med sterk forankring til myndigheter i medlemslandene, og organisasjonen utgir et stort antall analyser for en rekke politikkområder, inkludert utdanning. Helt konkret fikk resultater fra TIMSS i 1995 en sentral funksjon i *Education at a Glance*, en rapportserie som denne organisasjonen gir ut med jevne mellomrom (se f.eks. OECD, 1997). Etter hvert bestemte OECD seg for å etablere også egne undersøkelser av samme type, og den første PISA-undersøkelsen ble gjennomført i 2000. Flere andre studier med relevans for utdanningsområdet er siden blitt etablert gjennom OECD. (Se tabell 1.1 for en oversikt.) I likhet med IEA-undersøkelsene gjennomføres også OECD-undersøkelsene med jevne mellomrom (PISA hvert tredje år). Også land utenfor OECD kan delta i undersøkelsene.

Undersøkelsene er altså sykliske, og PISA og TIMSS⁵ har begge vært gjennomført seks ganger siden oppstarten. Dette at undersøkelsene gjennomføres jevnlig, er nyttig og viktig fordi studiene benytter design⁶ og psykometriske metoder⁷ som sikrer at resultatene for elevenes faglige prestasjoner refererer til den samme faste underliggende skalaen over tid. Undersøkelsene inkluderer også (med noen variasjoner) spørreskjemaer til elever, foreldre, lærere og skoleledere. Dette gir et stort tilfang av variabler som beskriver kjennetegn ved utdanningssystemene, og som

5. Norge var imidlertid ikke med i TIMSS 1999 og har dermed vært med i fem av de seks undersøkelsene

6. Noen oppgaver gjentas over tid, og disse utgjør dermed det som kalles et anker.

7. Gjennom bruk av såkalt Item Response Theory, IRT, kan man beregne skårer til elever som refererer til den samme underliggende skalaen – selv om elevene svarer på litt ulike oppgaver.

gir uttrykk for konteksten som utdanningen skjer innenfor. De internasjonale studiene har dermed en utforming som gjør det mulig å beskrive og analysere utviklingstrekk over tid på systemnivå. De internasjonale og nasjonale rapportene vier derfor stadig mer oppmerksomhet til disse tidsseriene av data. Det er også etter hvert et økende antall av vitenskapelige artikler som benytter data fra flere (minst to) tidspunkter for å studere endringer (se f.eks. Strietholt & Rosén, 2016). Et poeng som flere av kapitlene i denne boka tar opp, er imidlertid at spørsmålene som stilles til elevene i spørreskjemaene, endres over tid, noe som begrenser mulighetene til å gjøre gode analyser av endringene som skjer. Dette siste momentet er en spenning som ofte er gjenstand for diskusjon i de faglige komiteene som utformer instrumentene i undersøkelsene: Balansen mellom forbedring/oppdatering av eksisterende instrumenter og ønsket om å inkludere nye variabler versus behovet for å holde variabler stabile over tid for å muliggjøre analyser av utvikling over tid.

Hovedformålet med dette innledningskapitlet til boka er å gi et bilde med lav oppløsning hvor konturene av norsk skole skisseres gjennom resultater fra de internasjonale undersøkelsene. De resterende kapitlene kan med samme metafor sies å representere mer høyoppløselige bilder av enkeltelementer. Vi begynner med en kort drøfting av hva slags type informasjon disse undersøkelsene gir. Videre gir kapitlet en oppsummering av resultater i elevenes faglige prestasjoner for perioden 1995–2015. Gjennom disse resultatpresentasjonene, og ved å vise til analyser i de andre kapitlene i boka samt til tidligere presenterte analyser, drister vi oss også til å løfte fram det vi mener er noen av de mest sentrale og betydningsfulle funnene fra disse undersøkelsene i en norsk kontekst i et tjuetårs perspektiv. Dette er funn eller karakteristikk som vi vil hevde (i) er relevante å inkludere i overordnede refleksjoner om tilstanden til norsk skole, (ii) er knyttet til fenomener som bør følges videre, og som bør suppleres med annen og ny kunnskap, og som derfor (iii) er knyttet til fenomener som er relevant for politikktutforming og politisk debatt om norsk skole. De internasjonale undersøkelsene gir opplagt ingen endelig dom over norsk skole, og undersøkelsene i seg selv definerer ikke målene og formålene for norsk skole, noe man kan få inntrykk av hvis man kun forholder seg til tabloide presentasjoner i nyhetsmedia eller de mest høylytte og retorisk spissede delene av skoledebatten. Men det bør også legges vekt på at undersøkelsene inkluderer gode indikatorer for sentrale og viktige sider ved utdannings-systemer i alle land. Basert på erfaringene vi har gjennom disse 20 årene, tillater vi oss også å identifisere noen utviklingstrekk ved denne typen studier i årene som kommer.

HVA SLAGS INFORMASJON GIR DISSE STUDIENE?

Det unike med de internasjonale undersøkelsene er at de for et stort antall land samler inn informasjon fra representative utvalg fra godt avgrensede populasjoner (eksempelvis «15-åringer», «4. trinnse elever» eller «lærere på ungdomstrinnet»). Denne informasjonen samles inn gjennom standardiserte, godt kvalitetssikrede og dokumenterte prosedyrer og verktøy. Det viktigste bidraget fra de internasjonale undersøkelsene er derfor gode deskriptive data som kan brukes for å utvikle indikatorer for en rekke forhold som det er viktig å ha oversikt over på nasjonalt nivå. I tillegg inkluderer undersøkelsene bakgrunnsdata som gjør det mulig også å studere sammenhenger på systemnivå som politikken har som ambisjon å forsterke eller svekke, eksempelvis ønsket om å redusere ulikheter knyttet til kjønn eller elevers hjemmebakgrunn.

Det er for øvrig viktig å legge til at selv om man har robuste indikatorer for viktige kvaliteter i et utdanningssystem, så er slike tallverdier ofte utfordrende å tolke: Hva er egentlig et akseptabelt faglig nivå, hvilke verdier på en skala for faglig motivasjon kan oppfattes å være så lave at vi bør bekymre oss, når kan en korrelasjon eller regresjonskoeffisient for en sammenheng sies å indikere at en sammenheng er svak eller sterk? Designet til de internasjonale undersøkelsene gir mulighet for å knytte substansiell mening til slike resultater gjennom to typer sammenlikninger:

1. Det at undersøkelsene gjennomføres i en rekke land, gir mulighet for å gi *relative tolkninger* av gjennomsnittsverdier, regresjonskoeffisienter og andre statistiske estimater. Dette gjør det mulig å uttrykke mening i tallverdiene gjennom utsagn som for eksempel at «kjønnsforskjellene i Norge er typisk mindre enn ...». I tillegg gir dette internasjonale sammenlikningsgrunnlaget også en mulighet for å drøfte forhold eller variabler som er tilnærmet konstante i ett enkelt land, eksempelvis lengden for læreres utdanning, alder ved skolestart eller andre mer eller mindre fastlagte og sentralt bestemte retningslinjer (Hanushek & Wössmann, 2011). For å vende tilbake til metaforen med fotografi: Sammenlikninger med andre land gir oss kontraster i bildet som gjør at det som er i forgrunnen, vises bedre.
2. Som tidligere påpekt gir undersøkelsene relativt lange tidsserier med data som kan lenkes sammen. Dette gjør det mulig å sammenlikne en enhet (her land eller utdanningssystemer) med seg selv over tid. Norge og de skandinaviske landene er kjennetegnet ved å ha svært gode registerdata, og for mange typiske ressurskvaliteter i norsk skole har vi allerede gode muligheter til å beskrive og eventuelt framskrive utviklingstrekk. Men inntil ganske nylig har det ikke vært

mulig å studere hvordan utviklingen for prosess- og resultatkvantiteter utvikler seg over tid gjennom nasjonale data. Det er kun data fra de internasjonale undersøkelsene som foreløpig gir muligheter til å studere denne typen karakteristikk ved norsk skole over de siste 20 årene. Sammenlikninger over tid gir oss dermed også en serie med bilder som kan settes sammen til en film med objekter som beveger seg.

Dette gir to ulike tolknings- eller referanserammer for å analysere data fra undersøkelsen. I tillegg til å gi utdanningsmyndigheter nyttig informasjon om deres eget utdanningssystem har de internasjonale undersøkelsene et viktig sekundært forskningsformål gjennom at data blir gjort tilgjengelige for forskere som ønsker å studere mer overordnede utdanningsvitenskapelige problemstillinger.⁸ Norske forskere har i stor grad benyttet data fra disse undersøkelsene for denne typen analyser, noe som blant annet er tydelig gjennom et stort antall publikasjoner i vitenskapelige antologier og tidsskrifter. Kun et lite utvalg av disse artiklene er referert i dette kapitlet, men de andre bidragene i boka gir ytterligere eksempler på slike referanser. Det er også avlagt et stort antall doktorgrader som er helt eller delvis basert på analyser av data fra de internasjonale undersøkelsene (Angell, 1996; Gabrielsen, 2002; Hopfenbeck, 2009; Kaarstein, 2015; Kind, 1996; Lundetræ, 2010; Nilsen, 2014; Olsen, 2005; Pedersen, 2014; Solheim, 2010; Turmo, 2003), og datasettene brukes aktivt i utdanningen på flere læresteder i forbindelse med masteroppgaver. Det kan også tilføyes at organisasjonene bak undersøkelsene tilbyr skreddersydde metodekurs, de publiserer manualer som dokumenterer dataene og hva man må ta hensyn til når man skal bruke dem for analyse (se f.eks. Foy, 2017; OECD, 2009) og de bidrar også til å utvikle fritt tilgjengelig programvare og makroer som gjør det lettere å gjennomføre analyser. Her er programvaren IDB Analyser, som IEA har utviklet, et spesielt nyttig verktøy. I en systematisk kunnskapsoppsummering viser eksempelvis Hopfenbeck mfl. (2018) den omfattende rollen som data fra PISA har i den internasjonale forskningslitteraturen.

Datasettene er selvsagt ikke en Sareptas krukke, verken for den som ønsker å skaffe til veie et informert og utvetydig grunnlag for politikkutforming, eller for forskeren som søker ny innsikt og kunnskap til sitt forskningsfelt. De må brukes med fornuft og med oppmerksomhet om begrensninger knyttet til undersøkelsesnes utforming og data. Eksempelvis er dette tverrsnittstudier som vanskelig kan lede til sikre slutninger om årsaker,⁹ og studiene inkluderer kun noen utvalgte

8. Se <http://www.oecd.org/pisa/data/> og <http://www.iea.nl/data>

9. Men potensialet for forsiktige kausale tolkninger finnes også for denne typen studier, se for eksempel Strietholt, Bos, Gustafsson, og Rosén (2014).

utfallsmål for utdanningssystemer. Videre inkluderer datasettene beskrivelser av klasserom og skoler basert på elevers og læreres selvrappoteringer, og det er opplagt utfordringer knyttet til det å sammenlikne en del av resultatene på tvers av veldig ulike kultur- og språkområder.

NORSK DELTAKELSE I DE INTERNASJONALE UNDERSØKELSENE

En oversikt over alle de internasjonale storskalaundersøkelsene med norsk deltakelse er oppsummert i tabell 1.1.

TABELL 1.1 Oversikt over internasjonale undersøkelser med norsk deltakelse i perioden 1995–2015

Undersøkelse	Faglig område	Årstrinn/ alder	Når	Organi- sasjon
CIVED/ ICCS	Samfunnsfag: Demokratiforståelse og medborgerskap	8./9. trinn	1999, 2009 og 2016	IEA
ICILS	Grunnleggende digital kompetanse	9. trinn	2013 og 2018 (Norge deltok ikke i 2018)	IEA
PIRLS	Lesing	4./5. trinn	Hvert 5. år siden 2001	IEA
TEDS-M	Matematikklærerstudenters pedagogiske, didaktiske og matematiske kompetanse		2008	IEA
TIMSS	Matematikk og naturfag	4./5. trinn og 8./9. trinn	Hvert 4. år siden 1995 (Norge deltok ikke i 1999)	IEA
TIMSS Advanced	Matematikk og fysikk	13. trinn (vg3)	1995, 2008 og 2015	IEA
IALS/ALL/PIAAC	Voksnes kompetanse (lesing, regning og IKT-basert problemløsning)	16–65 år	1998, 2003 og 2012	OECD
PISA	Lesing, matematikk og naturfag	15-åring	Hvert 3. år siden 2000	OECD
TALIS	Læreres og skolelederes arbeidssituasjon	Ungdomstrinnet	Hvert 5. år siden 2008	OECD
TALIS Starting Strong Survey	Barnehageansattes arbeidssituasjon	Barnehager	2018	OECD

Tabellen viser at Norge er sterkt delaktig i denne typen internasjonale studier. Ungdomstrinnet er representert ved flest studier, og det er matematikk og lesing som dominerer. Grunnen til dette er nok at disse områdene er sentrale skolefaglige kompetanser i alle land, og de representerer to grunnleggende ferdigheter som er avgjørende for individers muligheter til å tilegne seg ny kunnskap gjennom hele livet. I tillegg er nok en bakenforliggende årsak at fagområdene representerer testkonstrukturer som det relativt sett er lettere å bli enige om hvordan man kan måle – på tvers av ulike kulturer og land – både fordi fagområdene har preg av å være universelle,¹⁰ og fordi det finnes god kunnskap og rik erfaring med å utvikle prøver for disse fagområdene.

Videre viser tabellen at IEA-studiene definerer populasjoner som kohorter av årstrinn. Dette valget kan knyttes til det generiske rammeverket for IEA-studiene. Her presenteres den grunnleggende modellen for IEA, hvor læringsutbytte blir sett på som en effekt av undervisning og andre påvirkninger fra skolesystemet.¹¹ Av samme grunn består utvalgene i de fleste IEA-studiene av hele klasser, og man legger stor vekt på å undersøke prosessene som skjer i klasserommet. I utgangspunktet er populasjonene for majoriteten av IEA-studiene (TIMSS, PIRLS, ICCS og ICILS) definert å gjelde for henholdsvis 4. og/eller 8. trinn. Imidlertid ble det etter hvert tydelig at Norge på denne måten ble representert ved elever som var omtrent ett år yngre enn elevene i de fleste andre land som det er naturlig å sammenlikne med (deriblant de andre nordiske deltakerlandene). Utdanningsdirektoratet, som representerer Norge i IEA, førte derfor over flere år en argumentasjon for at vi heller bør være representert med elever fra 5. og 9. trinn (Kavli & Thorsen, 2014). Dette ble akseptert, men for at det fortsatt skulle være mulig å følge utviklingen over tid, måtte Norge delta med populasjoner som representerte to tilgrensende årstrinn i noen år (noe som vises i tabellen ved at det står henholdsvis 4./5. trinn og 8./9. trinn for disse studiene). Dette gir unike datasett for Norge i disse årene med to tilgrensende klassetrinnkohorter. Imidlertid gir denne endringen også tolkningsutfordringer knyttet til betydningen av lengden av skolegangen, og det gir ikke minst utfordringer knyttet til det å studere utvikling over tid for norske elever i disse studiene.

PISA definerer populasjonen som en bestemt aldersgruppe; det vil si at det er barn født i en bestemt periode som er inkludert i utvalget. For Norge vil ikke denne definisjonen skille seg vesentlig fra IEA-populasjonene med klassetrinn. I vårt

10. Det finnes imidlertid opplagt kulturelt sensitive deler også for fag som matematikk og naturfag.

11. I engelskspråklig drakt benytter IEA en såkalt curriculum-modell, som litt upresist kan oversettes til læreplanmodell, hvor man ser på effekter av intendert læreplan (systemnivå), implementert læreplan (undervisningsnivået) og en oppnådd læreplan (læringsresultater).

system holdes elever stort sett i sammenholdte klasser med hele alderskohorter. For andre land gir imidlertid dette ganske store utslag. 15-åringer i Tyskland er eksempelvis spredt over flere ulike klassetrinn og ulike skoleslag. Dette reflekterer at PISA definerer de faglige utbyttene som en samlet effekt av læringsaktiviteter i både skole og hjem. PISA har derfor også et annet utvalgsdesign, hvor det innenfor de deltakende skolene trekkes ut 30 elever på tvers av klasser. PISA legger dermed hovedvekt på å studere prosesser på skolenivå.

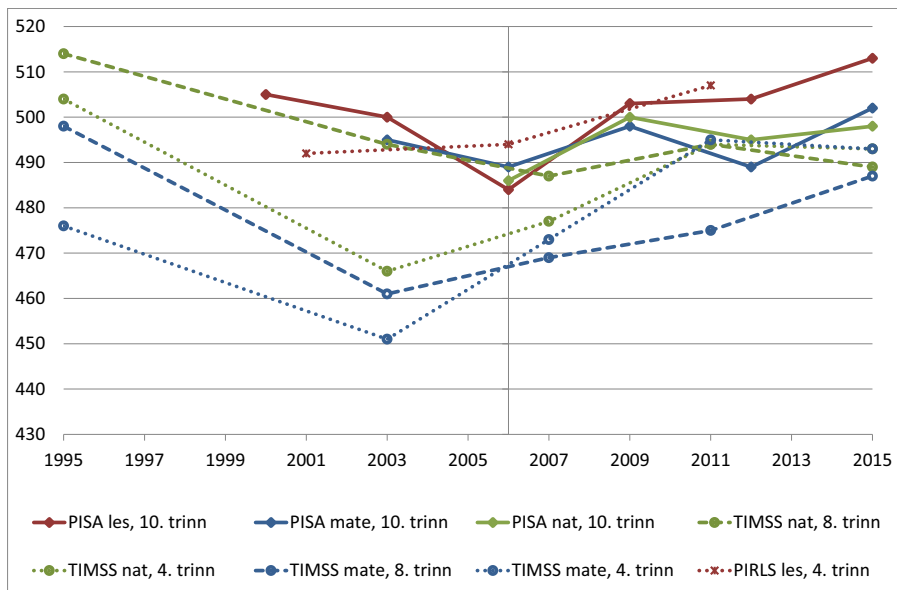
NOEN SENTRALE FUNN FRA DE INTERNASJONALE UNDERSØKELSENE I PERIODEN 1995–2015

Det norske skolesystemet og hverdagen i skolene har opplagt endret seg mye i denne tjuårsperioden: Det er innført skolestart for 6-åringer, to store utdanningsreformer har vært gjennomført, lærerutdanningene har blitt reformert flere ganger, opplæringsloven og mange forskrifter har endret seg betydelig, det er innført nasjonale prøver, og skolens hverdag er preget av IKT på en helt annen måte enn før, for kun å nevne noen viktige endringer. Samtidig ville nok en elev fra 2015 som ble sendt tilbake i tid til 1995, ikke føle at hun hadde kommet til en fremmed planet – mye består og er som før: Elevene sitter fortsatt som regel bak hver sin pult i klasserom med 15–30 elever, og læreren har sitt kateter. Det gis undervisning i stort sett de samme fagene, læreboka består som en sentral og definerende undervisningsressurs, elevene får karakterer først i ungdomsskolen, det er eksamener i siste år på ungdomstrinnet og i videregående skole, nesten alle elever går i offentlig eid skole osv.

Det er opplagt ikke mulig å knytte noen tydelig kausal forbindelse mellom norsk deltakelse i de internasjonale undersøkelsene og de endringene som har funnet sted i den samme perioden. I en nylig publisert analyse av denne sammenhengen beskriver Nortvedt (2018) hvordan politikktutforming skjer i en kompleks kulturell kontekst, hvor anbefalinger til norsk skole fra de internasjonale organisasjonene ikke nødvendigvis blir fulgt opp. Hun konkluderer med at resultater og analyser fra de internasjonale undersøkelsene i stedet blir brukt til å bekrefte de eksisterende retningene i politikktutforming, et funn som resonnerer godt også med analyser gjort i andre kontekster (se f.eks. Takayama, 2008).

I stedet for å gi en komplett beskrivelse av slike faktiske endringer av skolens innhold og organisering vil vi i det følgende presentere noen utviklingstrekk for norsk skole slik dette kommer til uttrykk i data fra de internasjonale undersøkelsene – og med vekt på PISA og TIMSS.

Figur 1.1 viser utviklingen i faglige prestasjoner for tre årstrinn (4., 8. og 10. trinn) i tre sentrale fagområder (lesing, matematikk og naturfag) i norsk skole i denne tjuårsperioden.¹² Figuren gir gjennomsnittresultatene for samtlige av studiene som inngår i PIRLS, PISA og TIMSS. Disse tre undersøkelsene har alle et internasjonalt gjennomsnitt som er satt til 500 det første året undersøkelsene ble gjennomført. Basert på erfaringer fra mange undersøkelser er effekten av ett års skolegang omtrent 40–50 poeng på barnetrinnet og 25–35 poeng på ungdomstrinnet. (Se for øvrig Olsen & Bjørnsson i kapittel 4.) Hver enkelt linje i figuren knytter sammen flere målepunkter som er direkte sammenliknbare. For PISA starter trendlinjen for matematikk i 2003 og for naturfag i 2006. Til tross for at disse fagområdene også var med i de tidligere PISA-studiene, var det i disse årene fagområdene var fullt utviklet og implementert for første gang (Kjærnsli & Jensen, 2016a).



FIGUR 1.1 Skår for norske elever over tid for faglige prestasjoner for alle internasjonale storskalaundersøkelser i perioden 1995–2015. Se tekst for forklaring om hvordan figuren kan tolkes

Det er viktig å påpeke at selv om skalaene i de tre ulike studiene er definert ut fra liknende prinsipper, så kan de ulike linjene i figur 1.1 *ikke* sammenliknes direkte.

12. Denne figuren er en revidert versjon av en tilsvarende figur i Olsen, Hopfenbeck, og Lillejord (2013).

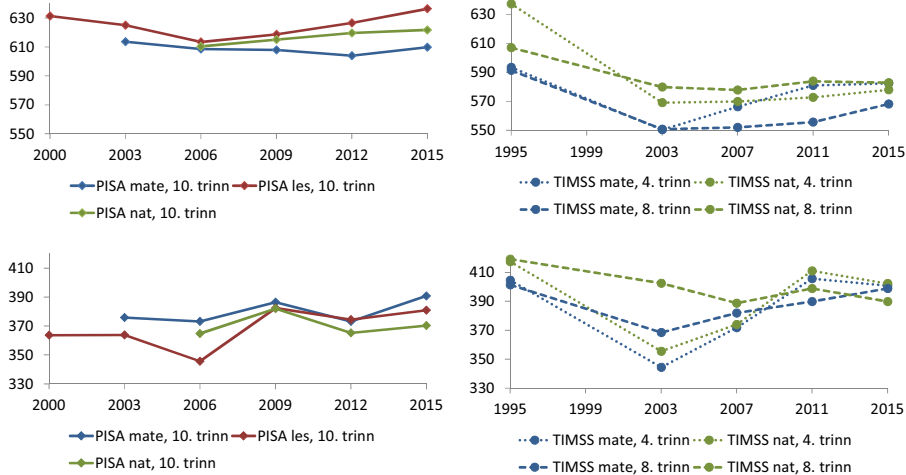
Dette er uavhengige måleserier som ikke er formelt lenket til hverandre. Det gir eksempelvis ingen mening å gjøre en direkte sammenlikning av prestasjoner for 4. og 8. trinn i TIMSS. Når vi likevel har valgt å plassere alle linjene i en felles figur, er det for å få fram et mer helhetlig bilde av utviklingstrekk i norsk skole i denne perioden.

Hovedfunn 1: En stor nedgang tidlig i perioden. En første observasjon er at prestasjonene til elevene som begynte på skolen i siste del av 1990-tallet, var svake. Elevene som deltok i TIMSS-undersøkelsen i 2003 og PISA i 2006, begynte sin skolegang i perioden 1995–1999. Nedgangen i resultatene i TIMSS fra 1995 til 2003 er på omtrent 40 poeng for noen fag og trinn. Denne utviklingen støttes også av andre funn som ikke er synlige i figur 1.1. Norge deltok i en internasjonal lesestudie i 1991. Dette var en forløper til PIRLS-studiene, som begynte i 2001, men resultatene kan ikke linkes direkte til samme skala. I lesestudien i 1991 presterte de norske elevene betydelig bedre enn det internasjonale gjennomsnittet, mens vi av figur 1.1 ser at de like gamle elevene som deltok i PIRLS 2001, presterte lavere enn det internasjonale gjennomsnittet. Resultatene fra PIAAC-undersøkelsen som ble gjennomført i 2012, peker i samme retning. I denne studien var det mulig å lenke tilbake til tidligere tilsvarende undersøkelser (se tabell 1.1), og man kan derfor slå fast at den yngste aldersgruppen i PIAAC (16–24 år) skåret betydelig svakere enn de samme alderskohortene i de tidligere undersøkelsene av voksnes lese- og regneferdigheter. Store deler av 16–24-åringene som deltok i PIAAC i 2012, begynte også sin skolegang i siste halvdel av 1990-tallet.

Hovedfunn 2: En (nesten) like stor framgang i siste halvdel i perioden. Den andre observasjonen fra figuren er at prestasjonene til elever som begynte på skolen etter årtusenskiftet, er betydelig bedre. Figuren viser at den tidligere nedgangen i prestasjoner i stor grad følges av en markert framgang for de senere alderskohortene. Framgangen er mest markant for naturfag og matematikk på 4. trinn, mens naturfag på 8. trinn fortsetter noe ned og flater ut. Jevnt over viser de siste undersøkelsene at de norske prestasjonene i matematikk og naturfag er på eller litt under det internasjonale gjennomsnittet, mens prestasjonene i lesing er signifikant høyere enn det internasjonale gjennomsnittet, både for 4. og 10. trinn.

For å belyse utviklingen i gjennomsnittlige prestasjoner i perioden mer i detalj viser de fire framstillingene i figur 1.2 hvordan prestasjonene for elever i de to endene av fordelingen har utviklet seg i perioden, uttrykt som henholdsvis 10. prosentil (nedre halvdel av figur 1.2) og 90. prosentil (øvre halvdel av figur 1.2) for de faglige skårene i PISA og TIMSS. Feilmarginene til de ulike punktene i figur

1.2 er noe større enn for figur 1.1, og de varierer en del. En god tilnærming til å vurdere usikkerheter i figuren vil være å betrakte endringer på omtrent 15 poeng fra ett punkt til det neste som statistisk signifikante endringer.



FIGUR 1.2 Utvikling over tid for 10. (nedre del) og 90. (øvre del) prosentilene for de norske elevenes skårer i PISA og TIMSS. Alle figurene er holdt i samme målestokk for å muliggjøre relative sammenlikninger.

Hovedfunn 3: Framgangen i siste halvdel er i hovedsak knyttet til framgang for elevene nederst i skårfordelingen. I likhet med figur 1.1 viser figur 1.2 at det er en større variasjon over tid for TIMSS-resultatene. Det er altså tydeligere profiler for endringer i norske elevers skårer i TIMSS-undersøkelsen enn for PISA-undersøkelsen. En måte å tolke dette på er at TIMSS-skåren synes å være mer sensitiv for de endringene som har skjedd i elevers faglige dyktigheter i matematikk og naturfag i perioden. En alternativ tolkning er at elevenes faglige prestasjoner i matematikk og naturfag er mer stabile for 10. trinn enn for de lavere trinnene. En første overordnet lesing av figur 1.2 er at endringene i begge ender i stor grad følger utviklingen for norske elevers gjennomsnitt i perioden, men utslagene er større for 10. enn 90. percentil. Hvis vi ser spesielt på TIMSS-figurene i høyre del av figur 1.2, legger vi merke til at nedgangen i den første delen av perioden ser ut til å være ganske likt representert i begge endene av fordelingen. Imidlertid er den senere framgangen i langt sterkere grad knyttet til framgang for elevene i nedre del av fordelingen (med naturfag for 8. trinn som et unntak). For PISA er det en liknende tendens, hvor den gjennomsnittlige profilen har et omtrent perfekt sam-

menfall med profilen for 10. prosentil. Helt overordnet er det altså slik at «nedgangstiden» er et allment fenomen, mens «oppgangstiden» først og fremst gjen-speiler et løft i bunnen av fordelingen.

Hovedfunn 4: Det er små kjønnsforskjeller i realfagene, men jentene leser mye bedre enn guttene. Over tid er det også etablert at kjønnsforskjeller i faglige prestasjoner endrer seg lite over tid. Det funnet som med rette får størst oppmerk-somhet, er at jentene leser betydelig bedre enn guttene på ungdomstrinnet. I PISA-undersøkelsen skårer jentene i Norge stabilt 40–50 poeng høyere enn guttene, og dette er gjennomgående en av de høyeste kjønnsforskjellene i lesing i hele OECD-området (Frønes, 2016). Mange forventet nok at disse kjønnsforskjellene skulle bli mindre når prøvene ble elektroniske, men de holdt seg på omtrent det samme nivået (Eriksen & Narvhus, 2013; Støle & Schwippert, 2017). Kjønnsforskjellen kan i hovedsak knyttes til at om lag 2/3 av de aller svakeste leserne er gutter (Frønes, 2016; Solheim & Gourvennec, 2017). Kjønnsforskjellene i lesing er også betydelige på barnetrinnet, men de tilsvarende forskjellene i PIRLS-studien har holdt seg stabilt på 15–20 poeng (Gabrielsen & Hovig, 2017). Gjennom analyser av tekstene og oppgavene i prøvene har det blant annet blitt pekt på at kjønnsforskjellene kan knyttes til lange sammenhengende tekster og tekster som kan karakteriseres som tekster for voksne (Roe, 2013), men forskjellene knyttet til teksttyper forsvinner når man kontrollerer for elevenes lesedyktighet (Solheim & Gourvennec, 2017). Uansett hva som er årsak bak kjønnsforskjellene i lesing, må det etter 20 år med internasjonale undersøkelser slås fast at norsk skole fortsatt ikke lykkes spesielt godt med å utjevne denne forskjellen. Kjønnsforskjellene i prestasjoner i de realfaglige områdene er en helt annen sak: Slike forskjeller finnes konsekvent ikke når det gjelder faglige prestasjoner i matematikk og naturfag i TIMSS og PISA (Bergem, 2016a, 2016b; Kjærnsli & Jensen, 2016b; Nortvedt & Pettersen, 2016).

Hovedfunn 5: Det er relativt små forskjeller mellom klasser og skoler i Norge. Gjennom alle de internasjonale undersøkelsene som Norge har deltatt i, er det også et tydelig funn at det gjennomgående er små forskjeller mellom klasse-rom og mellom skoler i Norge. De internasjonale undersøkelsene har et utvalg av elever som gjør det mulig å si noe om forskjellene til elevene kan knyttes til hvilken skole de går på. Jevnt over viser resultatene at 10 ± 2 % av forskjellene i pre-stasjoner kan knyttes til hvilke skoler elevene går på.¹³ 10 % er imidlertid ikke en

13. En mer teknisk korrekt måte å uttrykke dette på er at om lag 10 % av den totale variansen kan knyttes til skolenivået.

ubetydelig andel. Det betyr eksempelvis at skoler i «toppen av listen» presterer svært mye bedre enn de lavest-presterende skolene, men i en internasjonal sammenheng framstår skoleforskjellene likevel som små.

Hovedfunn 6: Hjemmebakgrunn er sterkt relatert til elevers prestasjoner også i Norge, men sammenhengen er svakere enn i de fleste andre land. Effekten av elevenes sosioøkonomiske hjemmebakgrunn (SES) er betydelig også i Norge, men lavere enn i så å si alle andre land som det er naturlig å sammenlikne seg med. Her er det også viktig å få fram at norske elever viser seg å ha svært høye verdier for de målene på SES som finnes i studiene. Med unntak av de store kjønnsforskjellene i lesing framstår derfor Norge som et land hvor en relativt mindre andel av forskjellene kan knyttes til hjemmebakgrunn eller hvilken skole eleven går på. Dette fenomenet analyseres og omtales grundig av Nilsen, Olsen og Bjørnsson i kapitel 7 i denne boka.

Hovedfunn 7: Norske lærere deltar lite i faglig relevant kompetanseheving. Et ikke like flatterende trekk ved norsk skole er manglende volum og systematikk i den kontinuerlige kompetansehevingen eller etterutdanningen av lærere – i alle fall når det gjelder tiltak rettet mot spesifikke fag. Dette er dokumentert gjennom alle de internasjonale studiene som har inkludert spørreskjemaer til lærerne. I tillegg viser TIMSS-studien at deltakelsen i faglig relevant etterutdanning har sunket betydelig fra 2007 til 2015 (Kaarstein, Nilsen, & Blömeke, 2016). Det er imidlertid rimelig å forvente at de siste årenes endringer i både organisering og volum av etter- og videreutdanningstilbudet vil føre til en framgang neste gang dette blir kartlagt i TALIS 2018 og TIMSS 2019.

Hovedfunn 8: Norsk skole er bygget på allmennlærere med middels lang utdanning. Relativt få norske lærere har en mastergrad (ca. 10 % og 20 % på hhv. barne- og ungdomstrinnet), og relativt få lærere rapporterer at de har en utdanning med spesialisering i det faget de underviser i (ibid.). I stor grad gjenspeiler dette at det norske utdanningssystemet har hegnet om allmennlæreren. Dette vil opplagt på lang sikt endre seg ved at lærerutdanningene i dag leder til mastergrad med sterkere faglig fordyping,

Hovedfunn 9: Elevene rapporterer om bedring i læringsmiljøet i perioden. Som et siste moment i denne oppsummering av sentrale funn, basert på resultater over de siste 20 årene, velger vi å inkludere at norsk skoleklima og læringsmiljø rapporteres å være til dels betydelig bedre i 2015 enn i tidligere år i undersøkelsen.

sene (Kjærnsli & Rohatgi, 2016; Nilsen, 2016). De ulike studiene har inkludert litt ulike aspekter ved skoleklima og læringsmiljø (god lærer–elev-relasjon, orden og arbeidsro, fravær av mobbing mm.), og innenfor hver av studiene har det også vært gjort endringer i de konkrete spørsmålene som inngår. Likevel er det et konsistent funn at for enkeltspørsmål og konstrukter som er beholdt uendret over tid, rapporteres det om en forbedring. For indikatorer knyttet til ro og orden har eksempelvis andelen elever, lærere og skoleledere som rapporterer at dette er problem, blitt omtrent halvert i de siste 10–20 årene. Det er imidlertid et funn som peker i motsatt retning på et sett med spørsmål som berører tilhørighet på skolen. Selv om de aller fleste elevene fortsatt rapporterer om god tilhørighet på skolen (for eksempel at de lett får venner), så er det en økning av andelen elever som gir negative tilbakemeldinger på disse spørsmålene (Kjærnsli & Rohatgi, 2016). Litt enkelt kan dette derfor oppsummeres som at de fleste, og stadig flere, opplever skolen som et sted med et godt psykososialt miljø, men at man samtidig må ha oppmerksomhet knyttet til økningen i den gruppen av elever som opplever utenforskap.

DE NESTE 20 ÅRENE MED INTERNASJONALE UNDERSØKELSER

Det er ingen grunn til å tro at de internasjonale undersøkelsene vil få mindre betydning eller synlighet i løpet av de neste 20 årene. Undersøkelsene har i mange land blitt en integrert del av de nasjonale systemene for å overvåke kvaliteten i utdanningen.¹⁴ Det er grunn til å tro at undersøkelsene gradvis vil endre seg over tid, både som en følge av at teknologi, design og metoder er i endring, men også fordi skolens innhold og forventninger til hva som er viktige utbytter av skolegang, konstant diskuteres. Det er derfor relevant å identifisere to hovedutfordringer som vi tror de internasjonale undersøkelsene i økende grad må forholde seg til i den videre utviklingen. Disse to hovedutfordringene er knyttet til styrkene som vi har beskrevet for de internasjonale undersøkelsene ovenfor, nemlig muligheten for å skape mening gjennom to typer sammenlikninger: sammenlikninger med seg selv over tid og sammenlikninger med andre land. Hovedutfordringen med den første typen sammenlikninger diskuterer vi nedenfor i lys spenningen mellom det

14. I Norge er de internasjonale undersøkelsene definert som en del av det nasjonale kvalitetsvurderingssystemet, og deltakelse for skoler/klasser/elever som trekkes ut, er derfor obligatorisk – på samme måte som deltakelse i nasjonale prøver og eksamen er obligatorisk. (Se <https://www.udir.no/tall-og-forskning/internasjonale-studier/obligatorisk-a-delta-i-internasjonale-undersokelser/>)

å bevare vs. å fornye. Den andre typen sammenlikninger drøfter vi ved å se på spenningen mellom det internasjonale vs. det nasjonale

BEVARE VS. FORNYE

De internasjonale undersøkelsene er konservative i sin natur. Det er minst to grunner til dette: For det første er rapporteringer av trender blitt den viktigste funksjonen for disse undersøkelsene. I mange land, inkludert vårt eget, er det – som tidligere nevnt – få nasjonale datakilder som kan brukes for å følge utviklingen over tid. Store endringer i utformingen av studiene er derfor en trussel mot dette viktige formålet. For det andre er dette prosjekter som involverer et stort antall land, skoler og elever. Det er derfor spesielt viktig at man i disse undersøkelsene benytter godt utprøvde instrumenter for å måle relativt godt forståtte begreper. Det vil alltid være større risiko for å mislykkes når man i stedet forsøker å måle nye og innovative begreper, hvor det ikke finnes mye tidligere kunnskap og erfaringer å bygge på.

Samtidig finnes det grunner for å revidere eller inkludere nye elementer i studiene. For det første er det slik at metodene som brukes i studiene, hele tiden er i utvikling. Hele historien med de internasjonale undersøkelsene siden oppstarten tidlig på 1960-tallet har vært preget av at metoder, og ikke minst regnekraft som trengs i forbindelse med å kunne implementere disse metodene, konstant har vært i utvikling. For det andre endrer jo verden utenfor seg, og med den også skolen og utdanningssystemet. Noen endringer er raske. I tjuårsperioden som omtales i denne boka, har det eksempelvis vært en eksplosiv teknologisk utvikling, og det er vel ingen grunn til å tro at denne utviklingen vil stoppe opp – snarere tvert om. Helt siden TIMSS 1995 ble planlagt, har det vært ytret ønsker om at man skal gå over til elektroniske prøver, men det var først med PISA i 2015 at dette ble implementert i fullskala.¹⁵ I dag gjennomføres undersøkelsene gjennom bruk av stasjonære og/eller bærbare pc-er med dedikert programvare, men vi ser nå at skoler i mange land er i ferd med å fase ut datamaskiner og i stedet går over til å bruke et mylder av ulike bærbare teknologier som Ipad, Chromebooks og liknende. Dermed vil det presse seg fram behov for nye løsninger som er plattformuavhengige og basert på internett. Andre endringer er langsomme og gradvise, og det er først over lengre tid at slike endringer blir synlige. Man vil for eksempel over en periode på noen tiår kunne se at skolens innhold, format og samfunnsmessige funk-

15. Den første småskalautprøvingen skjedde i forbindelse med PISA 2006 (Turmo & Lie, 2006), hvor tre land i PISA gjennomførte den såkalte CBAS- (Computer Based Assessment of Science) undersøkelsen elektronisk (OECD, 2010).

sjon og rolle stadig endrer seg, og med det må også de internasjonale undersøkelsene fornye seg.

Derfor kan det altså finnes svært gode argumenter for å ha stabile og uendrede målinger i de internasjonale undersøkelsene, men samtidig må studiene også respondere på endringer i samfunnet for øvrig. Det er imidlertid mulig å bevare samtidig som man endrer! Det vil kreve at man utvikler et underliggende rammeverk og en enighet om noen prinsipper for hvordan slike endringer skal implementeres. Nedenfor nevner vi noen mulige prinsipper som kan nedfelles i et slikt rammeverk:

1. Man kan åpne for revisjon av enkeltspørsmål i spørreskjemaet dersom man samtidig har et prinsipp om å bevare en uendret kjerne i den gruppen av spørsmål som til sammen utgjør et konstrukt. Dette gir mulighet for lenking over tid etter samme prinsipper som man bruker for de faglige prøvene (ankerspørsmål som er uendret over tid).
2. Man bør bli enige om en kjerne i spørreskjemaene som uansett ikke tillates å bli endret, kombinert med en definert (liten) andel spørsmål/konstrukter som kan endres fra én undersøkelse til den neste.
3. Prinsippet ovenfor kan kombineres med et eksplisitt design som beskriver hvordan noen få nye konstrukt kan inkluderes hvert år, uten at andre konstrukt tas helt ut, og uten at lengden på spørreskjemaet øker. Dette kan man oppnå dersom man på en systematisk måte roterer konstrukt over tid. Man kan eksempelvis ha med et overordnet motivasjonskonstrukt i kjernen hver gang – og samtidig ha mål for mer spissede motivasjonsbegreper annenhver gang.
4. Man bør ikke tillate endringer i prøveplattformen før man har gjennomført nøye planlagte lenkestudier som dokumenterer hvordan endringen påvirker målingen over tid.
5. Og aller viktigst: Man bør nedfelle prinsipper som beskriver et noe mer nyanisert syn på muligheten for å bevare lenkede mål uendret over tid. Det er sannsynlig at enkeltspørsmål og skalaer over tid endrer seg på måter som vi ikke kan forutsi. En slik mer realistisk oppfatning vil medføre at man innfører rapporteringer av flere delvis overlappende trendserier, eller såkalte «årgangstrender». Et eksempel på en slik type årganger eller generasjoner av trender er konsumprisindekser eller andre tilsvarende standardiserte økonomiske indekser. Konsumprisindeksen er basert på en standard handlekurv, men denne endrer seg jo med den økonomiske utviklingen. Når man må gjøre endringer i den standardiserte handlekurven, vil man fortsette å rapportere trenden for den gamle handlekurven i en periode, samtidig som man faser inn den nye.

DET NASJONALE VS. DET INTERNASJONALE

De internasjonale undersøkelsene vokser både i form av antallet og heterogeniteten i land som deltar i undersøkelsene. Undersøkelsene er i stor grad drevet fram gjennom konsensusprosesser, og når antallet interessenter eller partnere som skal bli enige, øker, vil et lite land få redusert innflytelse og påvirkningsmulighet på den konkrete innretningen av undersøkelsene. I tråd med at undersøkelsene nettopp skal legge til rette for internasjonal sammenliknbarhet, er det også nødvendig å sikre at undersøkelsene gjennomføres likt, og det er et behov for standardisering – noe som tradisjonelt har vært oppfattet som at alt må gjennomføres likt. I de siste årene har man i stadig større grad erkjent at gyldigheten av internasjonale sammenlikninger kan være begrenset. Dette blir det det eksempelvis lagt vekt på gjennom såkalte multigruppeanalyser, hvor man undersøker såkalt måleinvarians på tvers av grupper. Kaarstein og Nilsen gir et eksempel på en slik type analyse i kapittel 2 i denne boka.

Dette betyr ikke at internasjonale sammenlikninger ikke gir mening, men det bør bety at man i framtida legger til rette for å styrke analytiske koplinger mellom de internasjonale undersøkelsene og den nasjonale konteksten. En slik forsterket kopling kan man oppnå på minst to ulike måter:

For det første kan det være grunn til å inkludere *nasjonalt spesifikke komponenter* i datainnsamlingene. På den ene siden kan dette gjøres gjennom å ha enkelte spørsmål i spørreskjemaene som er nasjonalt spesifikke. Dette har allerede blitt forsøkt implementert i spørreskjemaene til elevene i både PISA og TIMSS knyttet til spørsmål om forhold ved hjemmet. Noen av disse spørsmålene blir brukt for å lage en indikator for *hjemmets økonomi*. De obligatoriske spørsmålene i dette konstruktet (eksempelvis antallet TV-apparater) fungerer dårlig i land som Norge fordi «alle har alt», for å si det litt enkelt. Dette fører til at indikatoren får store såkalte takeffekter, med det resultatet at indikatoren ikke evner å skille mellom elever som kommer fra hjem med reelt svært ulik økonomi. Rutkowski og Rutkowski (2013; 2018) har vist hvordan det å inkludere noen nasjonale eller regionale spørsmål (i tillegg til et felles sett med kjernes spørsmål) ikke bare gir bedre nasjonale mål, men også kan være en effektiv løsning for å forbedre den internasjonale sammenliknbarheten. En annen, og litt mer gjennomgripende, nasjonal tilpasning vil være å utvikle flere såkalte *nasjonale opsjoner*. I de internasjonale undersøkelsene har man gjennom tiden hatt flere tilleggskomponenter som land har kunnet velge om de vil delta i, eller ikke. Prøven i «husholdningsøkonomi» (financial literacy) i PISA er en slik opsjon som for eksempel Norge har valgt ikke å delta i. Ved å bygge denne ideen videre ut kan man tenke seg at de internasjonale undersøkelsene videreutvikles til å bli mer modulbaserte. I tillegg til kjerneområ-

der som alle må delta i, kan man inkludere et begrenset sett med mer spesifikke moduler som kan passe i undergrupper av land. Gitt den nasjonale konteksten kan det være gode grunner for at land velger å delta i litt ulike deler av studiene.¹⁶

For det andre kan det nasjonale analytiske potensialet i undersøkelsene forsterkes gjennom en *faktisk kopling med nasjonale data*. Dette er spesielt relevant i land som har gode nasjonale registre, (eksempelvis de nordiske landene). Dette vil kreve at data fra de internasjonale undersøkelsene har personidentifiserbar informasjon, noe de typisk ikke har i dag.¹⁷ Dersom vi antar at en slik kopling er gjennomførbar, kan et eksempel synliggjøre potensialet i denne typen informasjon. Nilsen, Olsen og Bjørnsson diskuterer i kapittel 7 hvordan mål for elevers sosio-økonomiske hjemmebakgrunn er svært sentrale i de internasjonale undersøkelsene. I dag er disse indikatorene basert på selvrappotering av eiendeler i hjemmet og spørsmål til elevene om foreldrenes utdanning og yrke. Ut fra disse blir det utviklet samleindikatorer eller konstrukter. Dette er variabler av stor betydning i de internasjonale undersøkelsene, enten som en substansiell variabel i seg selv eller som en kontrollvariabel for andre analyser. Ovenfor pekte vi blant annet på hvordan takeffektene for eiendeler som skal fange inn hjemmets økonomi, gir en sterk analytisk begrensning. I tillegg kan det nevnes at mange elever ikke vet hva de skal svare på spørsmål om foreldrenes utdanning og yrker. Ved å kople til nasjonale registre kan man få det som kan regnes som fakta eller «harddata» for slike forhold ved hjemmet. Med en slik kopling oppnår man både å få langt bedre variabler i analysene, og man kan studere validiteten til de selvrappoterte målene som brukes i de internasjonale undersøkelsene i dag. Tilsvarende ville man kunne gjøre koplinger mot nasjonalt forankrede målinger av elevers prestasjoner, eksempelvis de nasjonale prøvene – noe som også ville vært viktig for å etablere validiteten til målingene i de internasjonale undersøkelsene. En annen opplagt fordel som en kopling mot registerdata vil kunne gi, er muligheten for å kunne følge elever både bakover og framover i tid. De koplede registerdataene kan imidlertid uansett ikke erstatte de målene man samler i de internasjonale undersøkelsene. Man vil fortsatt ha behov for mål som er standardiserte og sammenliknbare på tvers av land. Men en slik kopling gir data som er unike i forskersammenheng, og det vil gi langt bedre informasjonsgrunnlag for nasjonal politikktutvikling.

16. Dette er heller ikke en helt ny idé. PISA har eksempelvis inkludert noen ekstra lette hefter til bruk i land med svært lav gjennomsnittskår, og TIMSS har en europeisk modul i spørreskjemaene.

17. Dette vil opplagt kreve en omfattende juridisk og personvernetisk vurdering, uten at vi her følger den tråden videre.

REFERANSER

- Angell, Carl. (1996). *Elevers fysikkforståelse. En studie basert på utvalgte fysikkoppgaver i TIMSS*. (dr.scient.), Universitetet i Oslo, Norge.
- Bergem, Ole Kristian. (2016a). Hovedresultater i matematikk. I Ole Kristian Bergem, Hege Kaarstein & Trude Nilsen (red.), *Vi kan lykkes i realfag. Resultater og analyser fra TIMSS 2015* (s. 22–43). Oslo: Universitetsforlaget.
- Bergem, Ole Kristian. (2016b). Hovedresultater i naturfag. I Ole Kristian Bergem, Hege Kaarstein & Trude Nilsen (red.), *Vi kan lykkes i realfag. Resultater og analyser fra TIMSS 2015* (s. 44–63). Oslo: Universitetsforlaget.
- Eriksen, Anna, & Narvhus, Eva Kristin. (2013). Digital lesing. I Marit Kjærnsli & Rolf Vegar Olsen (red.), *Fortsatt en vei å gå. Norske elevers kompetanse i matematikk, naturfag og lesing i PISA 2012* (s. 201–218). Oslo: Universitetsforlaget.
- Foy, P. (2017). *TIMSS 2015 User Guide for the International Database*. Boston: TIMSS & PIRLS International Study Center.
- Frønes. (2016). Resultater i lesing. I Marit Kjærnsli & Fredrik Jensen (red.), *Stø kurs. Norske elevers kompetanse i naturfag, matematikk og lesing i PISA 2015* (s. 136–171). Oslo: Universitetsforlaget.
- Gabrielsen, Egil. (2002). *Lese for livet. Lesekompetanse i den norske voksenbefolkningen sett i lys av visjonen om en enhetsskole*. (Dr. Philos.), Universitetet i Bergen, Norge.
- Gabrielsen, Egil, & Hovig, Johanne B. (2017). Hovedresultater fra PIRLS 2016 i Norden. I Egil Gabrielsen (red.), *Klar framgang! Leseferdighet på 4. og 5. trinn i et femtenårsperspektiv* (s. 32–49). Oslo: Universitetsforlaget.
- Gonzalez, E. J., & Smith, T. A. (red.). (1997). *User Guide for the TIMSS International Database – Primary and Middle School Years 1995 Assessment*. Chestnut Hill, MA: TIMSS International Study Center, Boston College.
- Hanushek, Eric A., & Wössmann, Ludger. (2011). The Economics of International Differences in Educational Achievement. I Eric A. Hanushek & Ludger Wössmann (red.), *Handbook of the Economics of Education* (Vol. 3, s. 89–200). Amsterdam: Elsevier.
- Hopfenbeck, Therese N., Lenkeit, Jenny, El Masri, Yasmine, Cantrell, Kate, Ryan, Jeanne, & Baird, Jo-Anne. (2018). Lessons Learned from PISA: A Systematic Review of Peer-Reviewed Articles on the Programme for International Student Assessment. *Scandinavian Journal of Educational Research*, 62(3), 333–353. doi: [10.1080/00313831.2016.1258726](https://doi.org/10.1080/00313831.2016.1258726)
- Hopfenbeck, Therese Nerheim. (2009). *Learning about Students' Learning Strategies. An empirical and theoretical investigation of self-regulation and learning strategy questionnaires in PISA*. (PhD), University of Oslo, Norway. (Series of dissertations submitted to the Faculty of Education, University of Oslo, No 117)
- Husén, Torsten, & Tuijnman, Albert. (1994). Monitoring Standards in Education: Why and How it Came About. I Albert C. Tuijnman & T. Neville Postlethwaite (red.), *Monitoring the standards of education. Papers in honor of John P. Keeves* (s. 1–21). Oxford: Pergamon.
- Kavli, Anne Berit, & Thorsen, Hallvard. (2014). Introduction: Northern Lights Report on TIMSS and PIRLS 2011. I Nordisk ministerråd (red.), *Northern Lights on TIMSS and PIRLS 2011: Differences and similarities in the Nordic countries* (s. 11–17). København: Nordisk ministerråd.

- Kind, Per Morten. (1996). *Exploring Performance Assessment in Science*. (Dr. Scient.), University of Oslo, Norway.
- Kjærnsli, Marit, & Jensen, Fredrik. (2016a). PISA 2015 – Gjennomføring og noen sentrale resultater. I Marit Kjærnsli & Fredrik Jensen (red.), *Stø kurs. Norske elevers kompetanse i naturfag, matematikk og lesing i PISA 2015* (s. 11–31). Oslo: Universitetsforlaget.
- Kjærnsli, Marit, & Jensen, Fredrik. (2016b). Resultater i naturfag. I Marit Kjærnsli & Fredrik Jensen (red.), *Stø kurs. Norske elevers kompetanse i naturfag, matematikk og lesing i PISA 2015* (s. 49–71). Oslo: Universitetsforlaget.
- Kjærnsli, Marit, & Rohatgi, Anubha. (2016). Læringsmiljøet i skolen. I Marit Kjærnsli & Fredrik Jensen (red.), *Stø kurs. Norske elevers kompetanse i naturfag, matematikk og lesing i PISA 2015* (s. 172–190). Oslo: Universitetsforlaget.
- Kaarstein, Hege. (2015). *Modelling, operationalising and measuring mathematics pedagogical content knowledge: threats to construct validity*. (PhD), University of Oslo, Norway. (Series of dissertations submitted to the Faculty of Education, University of Oslo, No 222)
- Kaarstein, Hege, Nilsen, Trude, & Blömeke, Sigrid. (2016). Lærerkompetanse. I Ole Kristian Bergem, Hege Kaarstein & Trude Nilsen (red.), *Vi kan lykkes i realfag. Resultater og analyser fra TIMSS 2015* (s. 97–119). Oslo: Universitetsforlaget.
- Lie, Svein, Kjærnsli, Marit, & Brekke, Gard. (1997). *Hva i all verden skjer i realfagene? Internasjonalt lys på trettenåringers kunnskaper, holdninger og undervisning i norsk skole*. Oslo: Institutt for lærerutdanning og skoleutvikling, UiO.
- Lundetræ, Kjersti. (2010). *16–24-åringers basisferdigheter. En studie av basisferdigheter relatert til selvpoppfatning, frafall i videregående opplæring og arbeidsledighet*. (PhD), Universitetet i Stavanger, Norge.
- Martin, Michael O., & Kelly, Dana L. (red.). (1997a). *Technical Report Volume I: Design and Development*. Chestnut Hill, MA: Center for the Study of Testing, Evaluation, and Educational Policy, Boston College.
- Martin, Michael O., & Kelly, Dana L. (red.). (1997b). *Technical Report Volume II: Implementation and Analysis*. Chestnut Hill, MA: Center for the Study of Testing, Evaluation, and Educational Policy, Boston College.
- Nilsen, Trude. (2014). *Trends in Physics Competence Using TIMSS for in-depth studies to characterize Norwegian students' physics competence and factors influencing and explaining changes in this competence*. (PhD), University of Oslo, Norway. (Series of dissertations submitted to the Faculty of Education, University of Oslo, No 202)
- Nilsen, Trude. (2016). Skolemiljø. I Ole Kristian Bergem, Hege Kaarstein & Trude Nilsen (red.), *Vi kan lykkes i realfag. Resultater og analyser fra TIMSS 2015* (s. 78–96). Oslo: Universitetsforlaget.
- Nortvedt, Guri A. (2018). Policy impact of PISA on mathematics education: the case of Norway. *European Journal of Psychology of Education*. doi: [10.1007/s10212-018-0378-9](https://doi.org/10.1007/s10212-018-0378-9)
- Nortvedt, Guri A., & Pettersen, Andreas. (2016). Matematikk. I Marit Kjærnsli & Fredrik Jensen (red.), *Stø kurs. Norske elevers kompetanse i naturfag, matematikk og lesing i PISA 2015* (s. 107–135). Oslo: Universitetsforlaget.
- OECD. (1997). *Education at a Glance*. Paris: OECD Publications.

- OECD. (2009). *PISA data analysis manual: SPSS and SAS, second edition*. Paris: OECD Publications.
- OECD. (2010). *PISA Computer-Based Assessment of Student Skills in Science*. Paris: OECD Publications.
- Olsen, Rolf Vegar. (2005). *Achievement tests from an item perspective. An exploration of single item data from the PISA and TIMSS studies, and how such data can inform us about students' knowledge and thinking in science*. (Dr. Scient.), University of Oslo, Norway. (Series of dissertations submitted to the Faculty of Education, University of Oslo, No 48)
- Olsen, Rolf Vegar, Hopfenbeck, Therese Nerheim, & Lillejord, Sølvi. (2013). Elevenes lærings-situasjon etter Kunnskapsløftet. *Norsk pedagogisk tidsskrift*, 97(6), 355–369.
- Pedersen, Ida Friestad. (2014). *Insights from TIMSS Advanced on critical aspects of the advanced mathematics program in Norwegian upper secondary school*. (PhD), University of Oslo, Norway. (Series of dissertations submitted to the Faculty of Education, University of Oslo, No 213)
- Robitaille, David F., & Garden, Robert A. (red.). (1996). *Research Questions & Study Design* (Vol. 2). Vancouver: Pacific Educational Press.
- Robitaille, David F., & Maxwell, Beverley. (1996). The Conceptual Framework and Research Questions for TIMSS. I David F. Robitaille & Robert A. Garden (red.), *Research Questions & Study Design* (Vol. 2, s. 34–43). Vancouver: Pacific Educational Press.
- Robitaille, David F., Schmidt, William H., Raizen, Senta, Mc Knight, Curtis, Britton, Edward, & Nicol, Cynthia. (1993). *Curriculum Frameworks for Mathematics and Science*. Vancouver: Pacific Educational Press.
- Roe, Astrid. (2013). Norske gutters resultater på nasjonale og internasjonale leseprøver. I Kåre Kverndokken (red.), *Gutter og lesing* (s. 13–33). Bergen: Fagbokforlaget.
- Rutkowski, David, & Rutkowski, Leslie. (2013). Measuring Socioeconomic Background in PISA: One Size Might not Fit all. *Research in Comparative and International Education*, 8(3), 259–278. doi: [10.2304/rcie.2013.8.3.259](https://doi.org/10.2304/rcie.2013.8.3.259)
- Rutkowski, Leslie, & Rutkowski, David. (2018). Improving the Comparability and Local Usefulness of International Assessments: A Look Back and A Way Forward. *Scandinavian Journal of Educational Research*, 62(3), 354–367. doi: [10.1080/00313831.2016.1261044](https://doi.org/10.1080/00313831.2016.1261044)
- Solheim, Oddny Judith. (2010). *Vurdering av leseprestasjoner. Grunnleggende forutsetninger for måling av leseferdighet*. (PhD), Universitetet i Stavanger, Norge.
- Solheim, Oddny Judith, & Gourvennec, Aslaug Fodstad. (2017). Gutte- og jentelesere i topp og bunn. En utforskning av leserprofiler. I Egil Gabrielsen (red.), *Klar framgang! Leseferdighet på 4. og 5. trinn i et femtenårsperspektiv* (s. 186–203). Oslo: Universitetsforlaget.
- Strietholt, Rolf, Bos, Wilfried, Gustafsson, Jan Eric, & Rosén, Monica (red.). (2014). *Educational policy evaluation through international comparative assessments*. Münster/New York: Waxmann.
- Strietholt, Rolf, & Rosén, Monica. (2016). Linking Large-Scale Reading Assessments: Measuring International Trends Over 40 Years. *Measurement: Interdisciplinary Research and Perspectives*, 14(1), 1–26. doi: [10.1080/15366367.2015.1112711](https://doi.org/10.1080/15366367.2015.1112711)

- Støle, Hildegunn, & Schwippert, Knut. (2017). Norske resultater fra ePIRLS – Online Informational Reading. I Egil Gabrielsen (red.), *Klar framgang! Leseferdighet på 4. og 5. trinn i et femtenårsperspektiv* (s. 50–74). Oslo: Universitetsforlaget.
- Takayama, Keita. (2008). The politics of international league tables: PISA in Japan's achievement crisis debate. *Comparative Education*, 44(4), 387–407.
doi: [10.1080/03050060802481413](https://doi.org/10.1080/03050060802481413)
- Turmo, Are. (2003). *Naturfagdidaktikk og internasjonale studier. Store internasjonale studier som ramme for naturfagdidaktisk forskning: En drøfting med eksempler på hvordan data fra PISA 2000 kan belyse sider ved begrepet naturfaglig allmenndannelse*. (Dr. Scient.), University of Oslo, Norway. (Series of dissertations submitted to the Faculty of Education, Universitetet i Oslo, No 23)
- Turmo, Are, & Lie, Svein. (2006). PISA's Computer-based Assessment of Science (CBAS): Gjennomføring og norske resultater våren 2005. *Acta Didactica*(2).