

# PERSPEKTIV- NOTAT



## **Kan skolens kvalitet måles?** Om prøver i skolen og hva de kan brukes til



## Sammendrag

Skolen er vår viktigste samfunnsinstitusjon, og alle vil ha en god norsk skole. Den skal ruste barna vår for livet og for arbeidslivet, og «opne dører mot verda og framtida», som det står i skolens formålsparagraf. Årlig bruker vi 98 milliarder kroner på grunnopplæringen. Men lærer barna å bli «gagnlege menneskje»?

Folk har tillit til skolen i Norge og 77 prosent av befolkningen er godt fornøyd med grunnskolen, viser innbyggerundersøkelsen fra 2015<sup>1</sup>. Men skolen står likevel ovenfor store utfordringer vi i nær fremtid må finne løsninger på. Skolen skal favne et større mangfold, barna skal forberedes på en stadig mer kompleks verden og forventningene fra omgivelsene øker. Høye frafallstall, inkluderingsutfordringer, mobbing, ressursknapphet i tid og penger, rapporteringsbyråkrati og tillitsmangel er vanlige problembeskrivelser fra flere hold.

Skal vi bygge en enda bedre skole må vi først finne ut hva problemene egentlig er. Det kan være nevnte utfordringer, eller det kan være måling, styring og innhold. Eller helt andre ting.

I dette notatet ser vi kun se på det nasjonale prøveregimet: Et mye omdiskutert område som ofte trekkes frem som et av hovedproblemene i skolen. Fra både lærere, foreldre og elever uttrykkes det ofte misnøye med prøvene. De tar tid. De koster penger. De er belastende for barna. I den offentlige debatten om skolen er det ofte resultatene på prøvene som får mest plass. Når en skole gjør det bra skryter vi av både lærere og elever. Når en skole gjør det dårlig retter vi pekefingeren: Undervisningen er dårlig, skolen ledes dårlig, politikerne har ikke lykket.

Men hvor mye tid og plass tar prøveregimet? Og hva måler det egentlig? Kan prøvene si noe om skolens kvalitet? Kan de bedømme lærerens undervisning? Kan det si noe om hvor smarte barn vi har? Er prøveregimet egnet til å styre skolen i riktig retning?

Den offentlige debatten om prøver dreier seg mest om PISA-undersøkelsen, Nasjonale prøver og kartleggingsprøver. Derfor er det disse vi har tatt for oss i notatet. Vår konklusjon er at disse prøvene *i seg selv* ikke er noe grunnleggende problem for kvaliteten eller undervisningen i skolen i dag. Prøvene som gjennomføres på nasjonalt nivå er bare en liten del av prøvene de fleste norske skoleelever går igjennom i løpet av skolen, og utgjør ikke et problem i hverken tid eller kostnad. Omfanget av den enkelte skole og kommunes testregime utover disse prøvene har vi ikke sett på i dette notatet.

Og viktigere: Prøvene gir oss også mye interessant og viktig informasjon. PISA-undersøkelsen kan gi oss en vurdering av trender i norske elevers kompetanse. Nasjonale prøver gir lærere og skoler viktig informasjon om forskjeller mellom elever. Kartleggingsprøver identifiserer hvilke elever som trenger særskilt oppfølging og hjelp.

Prøvene er gjennomarbeidet og av høy kvalitet – og felles og tilgjengelig for alle skoler. Informasjonen de gir oss er valid. Men vi skal ikke overdrive prøvenes verdi eller betydning. Prøvene gir bare informasjon om et liten brøkdelen av skolens oppgave. De måler det som kan måles, det vil si en svært liten del av den kunnskapen og kompetansen elevene får, og som vi ønsker at de skal få. En annen begrensning med prøvene er at de i ulik grad kan brukes til å sammenlikne over tid og på tvers av skoler. En rekke andre faktorer påvirker resultater. Prøvene må dermed tolkes med varsomhet og alltid sammen med annen informasjon.

Problemet er dermed ikke prøvene i seg selv, men kunnskap og praksis for tolkning og anvendelse av resultatene. Prøvene er ikke et tilstrekkelig kunnskapsgrunnlag for strukturelle eller faglige endringer. De kan heller ikke brukes isolert til å vurdere kvaliteten på elever, lærere, skoleledere, skoler, eller gjennomførte politiske tiltak og reformer. Informasjonen vi får er et nyttig grunnlag for informasjon og debatt blant både forskere og lærere, men det trengs mer kunnskap og videreutviklet praksis for tolkning og anvendelse. Nasjonale myndigheter informerer godt om prøvenes muligheter og begrensninger. Det er kommunenes ansvar at prøvene ikke misbrukes.



## INNHOOLD

<b>1. Innledning</b> .....	<b>5</b>
<b>2. Det norske kvalitetsvurderingssystemet</b> .....	<b>5</b>
<b>2.1 PISA-undersøkelsen</b> .....	<b>7</b>
2.1.1 Utforming og gjennomføring .....	7
2.1.2 Testresultatene.....	9
2.1.3 Kritikken av PISA .....	9
2.1.4 Hva kan vi bruke PISA til? .....	11
<b>2.2 Nasjonale prøver</b> .....	<b>12</b>
2.2.1 Utforming og gjennomføring.....	12
2.2.2 Grunnleggende ferdigheter .....	13
2.2.3 Resultatene.....	13
2.2.4 Hva kan vi bruke nasjonale prøver til? .....	14
2.2.5 Kritikken av nasjonale prøver .....	15
<b>2.3 Utdanningsdirektoratets kartleggingsprøver</b> .....	<b>16</b>
2.3.1 Utforming og gjennomføring .....	16
2.3.2 Hva kan vi bruke kartleggingsprøvene til?.....	16
2.3.3 Kritikkk av kartleggingsprøver .....	17
<b>2.4 Tester som informasjonskilde om skolen</b> .....	<b>17</b>
2.4.1 Testinformasjonens funksjoner .....	18
<b>2.5 Testvariasjon</b> .....	<b>19</b>
<b>3. Gjør prøvene skolen bedre?</b> .....	<b>21</b>
3.1 Hva kan vi bruke prøvene til? .....	21
3.2 Hva mer trenger vi? .....	23
<b>4. Konklusjon</b> .....	<b>24</b>
4.1 Beholde eller forkaste?.....	24
4.2 Hva nå? .....	25
<b>5. Sluttnoter</b> .....	<b>27</b>



## 1. Innledning

Internasjonale undersøkelser, nasjonale prøver, kommunale kartleggingsverktøy, prøver på klasse/trinn, karakterstøttende prøver, eksamenskarakterer, standpunkt karakterer, brukerundersøkelser og forskning utgjør informasjons- styrings- og kunnskapsgrunnlaget i skolen. Alle inngår i kvalitetsvurderingssystemet i skolen.

Det vies ulik oppmerksomhet til delene av det norske kvalitetsvurderingssystemet. Selv om for eksempel standpunkt karakterer og eksamenskarakterer får store konsekvenser for elevers muligheter senere i livet, er det påfallende lite debatt om disse vurderingssystemene i skolen. Mer oppmerksomhet vies til for eksempel internasjonale undersøkelser og Nasjonale prøver.

Selv om sistnevnte undersøkelser er grundige, etterprøvbare og solide, belyser de en veldig liten del av skolens oppgaver og mål. Så er de heller verken dominerende i tid for skolene og avgjørende for elever.

Informasjon og kunnskap om skolen er bra. Det er ofte også bedre med *litt* viten, enn *ingen* viten. Men hvis vi ikke tolker og bruker informasjonen vi sitter på rett, tar vi ikke veloverveide beslutninger, veivalg og prioriteringer i skolen. Det er derfor viktig å forstå hva vi faktisk vet eller ikke vet på bakgrunn av ulike undersøkelser og prøver.

Dette notatet tar for seg de mest omdiskuterte undersøkelsene og prøvene i skoledebatten: PISA-undersøkelsen, Nasjonale prøver og kartleggingsprøver, og forsøker å klargjøre hva de enkelte prøvene og testene gir oss av informasjon, og hva vi kan bruke dem til.

## 2. Det norske kvalitetsvurderingssystemet

Norske barns prøveresultater er ofte utslagsgivende for hvordan vi vurderer barna, læreren, rektor, skolen, og politikere på både nasjonalt og lokalt nivå. I Stortingsmelding om kvalitet og mangfold i fellesskolen «[På rett vei](#)» fra 2013 heter det at «kunnskap om tilstanden i sektoren og i den enkelte kommune og skole er grunnlaget for å vurdere forholdet mellom tilstand og mål for opplæringen, og om det er behov for endringer og utviklingstiltak»<sup>2</sup>. Prøvene skal med andre ord både gi en situasjonsbeskrivelse – til lærer, foreldre, skoler og myndigheter, og styringsinformasjon.

Kvalitetsvurderingssystemet

Det norske kvalitetsvurderingssystemet ble innført i 2004, i kjølvannet av Kvalitetsutvalget, ledet av Astrid Sjøgnen, og het opprinnelig «Det nasjonale kvalitetsvurderingssystemet (NKVS)<sup>3</sup>». Ved innføringen var Norge et av få sammenlignbare land uten et kvalitetsvurderingssystem<sup>4</sup>, og det hadde i lengre tid vært diskutert hvordan vi kunne få et bedre kunnskapsgrunnlag om hva som foregår i skolen. Utvalget var opprinnelig opprettet av Stoltenberg I-regjeringen, men fikk sitt mandat og sammensetning endret etter regjeringsskiftet av daværende utdanningsminister Kristin Clemet. Kvalitetsvurderingssystemet ble vedtatt innført i forbindelse med behandlingen av Statsbudsjettet for 2004, og omfattet hele bredden av kvalitetssikring og vurdering, herunder omdannelsen av Læringscenteret til det som skulle bli Utdanningsdirektoratet<sup>5</sup>.

Kvalitetsvurderingssystemet består av nasjonale prøver, ulike brukerundersøkelser, kartleggingsprøver, eksamen og internasjonale undersøkelser. Tilsynsrapporter, lærings- og karakterstøttende prøver, forskning og evaluering er også regnet som en del av systemet<sup>6</sup>.

Kartleggings- og vurderingsmateriellet skolene har, er fordelt på fem nivå: internasjonalt (gjelder kun for et utvalg elever), nasjonalt, kommunalt, på den enkelte skole også lærer/teamnivå.

Kunnskapsdepartementet beskriver systemet som «det viktigste datagrunnlaget for å belyse elevers læringsresultater»<sup>7</sup>.

Men departementet tar også mange forbehold om at testene må gjennomføres og brukes riktig på alle nivåer, og tolkes og forstås rett av alle som leser dem. Departementet utdyper at «[kvalitetsvurderingssystemet] ikke er det eneste grunnlaget for å vurdere hvordan det står til med skolen og elevene», og at «[n]asjonale utdanningsmyndigheter har arbeidet for å tydeliggjøre formålet med kvalitetsvurderingssystemet og de ulike nivåenes roller»<sup>8</sup>.

«På rett vei» påpeker at «[kvalitetsvurderingssystemet] ikke [gir] et heldekkende kunnskapsgrunnlag»<sup>9</sup>. For det første er det ikke alle grunnleggende ferdigheter som inngår i prøvene. Skriveferdigheter og muntlige ferdigheter måles ikke med noen av prøvene i kvalitetsvurderingssystemet. For det andre er det naturligvis en lang rekke ferdigheter og kompetansemål som prøvene ikke kan måle. Som vi kommer tilbake til er det også vanskelig å måle hvorvidt skolen lykkes med det brede samfunnsoppdraget.

Så hva er det da vi egentlig forsøker å finne ut av? Hensikten med nasjonale prøver er først og fremst å «gi lærerne et godt verktøy for å gi bedre opplæring til den enkelte elev», uttalte Kunnskapsministeren til Dagsavisen i forbindelse med publisering av nasjonale prøveresultater i 2015. Han la til at «Nasjonale prøver er også et godt utgangspunkt for å sette skolekvalitet på dagsorden ute i kommunene»<sup>10</sup>.

Nasjonale prøvers funksjon kommer vi tilbake til i kapittel 1.2, men Kunnskapsministeren har rett i at prøveresultater setter skolepolitikk på dagsorden. Den offentlige skoledebatten preges ofte av tall fra både internasjonale og nasjonale tester, og resultatene får store oppslag i media. Ofte er det disse tallene som danner grunnlaget for offentlig debatt. Men hva kan faktisk egentlig prøvene måle?

Informasjon fra prøveresultater viser for eksempel at skoler i Oslo vest ofte presterer bedre på Nasjonale prøver enn skoler i Oslo øst. Urbane og tettbygde områder presterer bedre enn grisgrendte strøk, med Bærum, Oslo og Stavanger på topp. Betyr det at skolene på vestkanten er bedre enn skolene på østkanten? Ikke nødvendigvis. Som vi skal se er det flere veier til en god skole, og gode resultater på for eksempel Nasjonale prøver kan ha andre forklaringer enn at selve skolen er bedre enn snittet eller at undervisningen er i særklasse god.

Selv om prøvene på ulike nivåer og på ulike måter gir et godt grunnlag for både egenvurdering og sammenlikning, har de vært gjenstand for betydelig kritikk, både fra forskere, foreldre og politikere. Kritikken kan oppsummeres i tre punkter:

Prøvene flytter fokus i undervisningen. Det kan hevdes at stor oppmerksomhet om prøvene gjør at undervisningen innrettes mot å skåre bra på standardiserte tester, ikke å lære og utforske. Fokus på prøvene tar tid i undervisningen og konsentrasjonen fra andre deler av skolen. På den annen side kan man argumentere at øving på grunnleggende ferdigheter som lesing og skriving er bra og nettopp blant skolens kjerneoppgaver.

Vi måler bare det som kan måles – og får dermed for lite informasjon. Prøvene måler ikke skolens formålsparagraf og lærerplanens kompetansemål, og gir dermed et forvrengt bilde av måloppnåelse som overdriver verdien av kvantitative mål på læring. Fordelen med presise indikatorer er at de kan

sammenliknes mellom land, byer og skoler og (flere av dem) følges over tid. Ulempen er at de aldri gir hele bildet<sup>11</sup>.

Prøvene er villedende og ikke representative for det som er viktigst. De gir interessant informasjon innenfor enkelte fag og kompetanseområder, men kan vi derfra påvise en sammenheng mellom gode prestasjoner på standardiserte prøver, og gode prestasjoner på områder som kreativitet, kritisk tenking, skriving, muntlighet, samarbeid, standhaftighet, relasjonelle ferdigheter etc.? Er det slik at hvis elever skårer godt på Nasjonale prøver i regning og lesing, så gjør de det også godt i naturfag og historie?

I det følgende skal vi presentere ulike prøver og drøfte spørsmålene over. Analysen er avgrenset til PISA, nasjonale prøver og kartleggingsprøver. Andre deler av kvalitetsvurderingssystemet er viktige komponenter men mindre kontroversielle.

De ulike testene har ulike funksjoner og gir ulik informasjon. I dette notatet presenteres tre typer prøver: PISA-undersøkelsen, nasjonale prøver og kartleggingsprøver. For å svare på dette gjennomgår vi prøvenes utforming og formål, resultater, og styrker og svakheter, samt hva de kan brukes til.

## 2.1 PISA-undersøkelsen

Den mest kjente av de internasjonale undersøkelsene er PISA, men den er ikke den eneste. Norge deltar og har deltatt i flere ulike undersøkelser. Ettersom PISA-undersøkelsen er den mest kjente og omdiskuterte, og de internasjonale prøvene har mange fellesnevnerer, avgrenser vi oss til PISA i dette notatet.

Internasjonale undersøkelser forsøker å sammenligne resultater mellom land. Det har vært tverrpolitisk enighet i Norge om at deltakelse i slike tester er ønskelig<sup>12</sup>. Poenget med slike internasjonale undersøkelser er å kunne sammenlikne Norges nivå og utvikling med andre land og oss selv.

### 2.1.1 Utforming og gjennomføring

Norge deltar i følgende internasjonale tester:

Navn	Fagområde	Populasjon
PISA	Lesing, matematikk og naturfag	15-åringer
TIMSS	Matematikk og naturfag	4. klassetrinn
PIRLS	Lesing	4. klassetrinn

Tidligere har vi også deltatt i undersøkelser som CIVIC/ICCSV (demokrati og engasjement).

PISA er en internasjonal komparativ undersøkelse av skolesystemer i ulike land. Undersøkelsen gjennomføres hvert tredje år, og er i regi av OECD. Det opereres med *utvalg*, slik at ikke alle norske elever tar testene/prøvene. Målet med undersøkelsen er å evaluere utdanningssystemer ved å teste og sammenligne 15-åringers ferdigheter og kompetanse i følgende fagområder: Lesing (reading literacy), matematikk (mathematical literacy) og naturfag (scientific literacy)<sup>13</sup>. Problemløsning (problem solving) blir også testet<sup>14</sup>.

I følge OECD deltok en halv million elever på verdensbasis i 2012-testen, og til sammen nesten 70 land. Norge har deltatt i PISA siden oppstarten i 2000. Om lag tre prosent av norske skolelever deltar på PISA-testen.

Den faglige prøven i PISA varer i to timer, og består av åpne spørsmål og flervalgsoppgaver. Elevenes faglige oppgaver suppleres med et spørreskjema fra både elever og rektor om elevenes «backgrounds, schools and learning experiences and about the broader school system and learning environment»<sup>15</sup>.

Fagområdene er valgt ut fordi «lesing og matematikk er grunnleggende ferdigheter som er viktige i seg selv, og viktige for læring i andre fag. Naturfag i PISA inneholder temaer som er viktige og felles i mange land (helse, miljø, teknologi, etc.)»<sup>16</sup>.

*Boks 1: Hovedområder for PISA-undersøkelsene så langt*

2000 - Lesing  
 2003 - Matematikk  
 2006 - Naturfag  
 2009 - Lesing  
 2012 - Matematikk  
 2015 – Naturfag

*Kilde: Utdanningsdirektoratet<sup>17</sup>*

De tre fagområdene PISA forsøker å måle er med hver gang prøven gjennomføres, men ett av dem er hovedområde ved hver gjennomføring (se faktaboks).

PISA-testen forsøker å måle ferdigheter og kompetanser i lesing, matematikk og naturfag, og OECD går grundig til verks i utviklingen. Hver test lages av et bredt sammensatt internasjonalt team av forskere, og gjennomgår cirka tre år med utvikling og prøving før den lanseres. I Norge er det cirka 3.5 årsverk fordelt på flere forskere ved Institutt for lærerutdanning og skoleforskning (ILS) ved Universitetet i Oslo som arbeider med gjennomføring av undersøkelsen.

Det er et internasjonalt senter (Westat) som trekker utvalget av elever som skal gjennomføre testen. Utvalget blir trukket sentralt for å sørge for at alle land har like gode prosedyrer. Det norske utvalget trekkes basert på tall fra Grunnskolens informasjonssystem (GSI). Når utvalget og skolene er trukket blir de fulgt opp av norske PISA-forskere i forberedelses- og gjennomføringsfasen.

Ettersom flere av oppgavene brukes om igjen, er ikke alle offentlig tilgjengelig. Forskerne Hopmann, Brinek og Retzl mener at det «ser ut til at virksomhetens overveldende suksess har ført til at ethvert forsøk på å diskutere PISAs utforming, datainnsamling og analysemetoder blir betraktet som smålig og respektløst. PISAs egen strategi med ikke å gi tilgang til hele databasen, inkludert samtlige spørreskjema, bidrar til dette problemet<sup>18</sup>».

Flere forskere tilbakeviser denne påstanden<sup>19</sup>. Selv om ikke alle oppgavene som benyttes i PISA-testen er offentlige, betyr ikke det at de er hemmelige. Mange oppgaver er frigitt, og de som ikke er frigitt er tilgjengelig ved forespørsel. Både spørreskjemaene og datasettene med resultat, metode og manualer ligger tilgjengelig på <http://www.oecd.org/pisa/pisaproducts/>. Hvorvidt OECD er mottakelige for kritikk



og vitenskapelig vurdering av andre forskere skal ikke forfatteren av notatet vurdere, men datagrunnlaget for kritikk er tilgjengelig for alle som ønsker det.

I PISA 2015, da naturfag var hovedområde, var det med ca. 180 oppgaver i naturfag. Alle elevene løser imidlertid ikke alle oppgavene. På grunn av tidsbegrensninger løser elevene ulike deler av oppgavesettene. Det er likevel et stort nok antall elever som svarer på de ulike oppgavene til at de kan benyttes i PISAs analyse for å gi informasjon om hvordan norske elever gjør det på oppgavene.

I PISA brukes såkalt booklet-rotasjon (multiple matrix sampling – MMS) som er en metode for 1) å håndtere kravet til mange oppgaver for å gjøre gode målinger og 2) å oppnå en akseptabel prøvetid for elevene<sup>20</sup>. PISA 2009 besto for eksempel av totalt 510 minutters mulig testtid, og det ville vært en umulighet om alle elever skulle hatt alle oppgaver.<sup>21</sup>

Dette designet gjør at det er stor usikkerhet knyttet til målinger på elevnivå, og undersøkelsene kunne ikke gi informasjon om enkeltelevers kompetanse på en forsvarlig måte. Derimot er denne metoden og designet effektivt for storskala-undersøkelser av systemnivå, og elevbesvarelser og bakgrunnsdata anslår svært presist trekk ved populasjonen<sup>22</sup>.

«Det internasjonale samarbeidet om utvikling av oppgaver pågår i om lag tre år. I den prosessen prøves det ut mange flere oppgaver enn vi faktisk bruker i den endelige testen», sier stipendiat ved ILS Tove Stjern Frønes<sup>23</sup>. Oppgavene rettes og svarene vurderes av flere forskergrupper verden over.

### 2.1.2 Testresultatene

Resultatene fra PISA-undersøkelsene publiseres av OECD. Det lages også nasjonale hovedrapporter. I Norge er det forskere ved Institutt for lærerutdanning og skoleforskning ved Universitet i Oslo som lager hovedrapporten. Basert på resultatene fra PISA-testen er det skrevet og levert følgende rapporter om norsk skole:

Hovedrapporter<sup>24</sup>:

PISA 2012: [Fortsatt en vei å gå.](#)

PISA 2009: [På rett spor.](#)

PISA 2006: [Tid for tunge løft.](#)

PISA 2003: [På rett spor eller ville veier?](#)

PISA 2000: [Godt rustet for framtida?](#)

Andre publikasjoner:

PISA 2012: [Norske elevers kompetanse i problemløsning i PISA 2012.](#)

PISA 2009: [Elever på nett](#) (Temarapport om den digitale leseprøven i 2009).

Rapporten fra 2012 oppsummerer norske elevers testresultater på PISA-testen slik: Norske prestasjoner er stabile og ligger omtrent på OECD-gjennomsnittet. Flere av endringene i både positiv og negativ retning er imidlertid ikke statistisk signifikante<sup>25</sup>.

### 2.1.3 Kritikken av PISA

En sentral kritikk mot PISA-testen er at den ikke tar utgangspunkt de ulike lands lærerplaner. En prøve som forsøker å måle kompetanser som ikke har vært en del av undervisningen enda er følgelig ikke veldig nyttig. Andre vil hevde at utgangspunktet i PISA-testen i stor grad er overlappende med nasjonale lærerplaner, selv om utgangspunktet for PISAs rammeverk ikke er hentet derfra.

Rammeverkene er dokumentene som danner grunnlaget for utvikling av oppgavene. Rammeverket i naturfag beskriver for eksempel hvilke kompetanser som skal måles. Rammeverkene for 2015-prøven finner du her: <http://www.oecd.org/pisa/pisaproducts/pisa2015draftframeworks.htm>.

PISA har fått kritikk for at blant annet det kommersielle læremiddelselskapet Pearson Inc. har fått oppgaven med deler av denne utviklingen. Kritikerne sår tvil om motivasjonen bak testingen og testens samarbeidspartnere<sup>26</sup>. Pearsons rolle i PISA er å administrere en gruppe internasjonalt ledende forskere som står bak arbeidet med rammeverkene.

Begge innvendingene dreier seg om hvorvidt kunnskap kan måles og sammenliknes. Sosialøkonom, tidligere statssekretær på statsministerens kontor, og ekspedisjonssjef for det daværende utdanningsdepartementets internasjonale avdeling gjennom mange år, Kjell Eide, ledet OECDs arbeid med utdanningsplanlegging fra 1961 til 1964. I en artikkel i Norsk Pedagogisk Tidsskrift skriver han at «utdanningspolitikken ble interessant for OECD på grunn av dens antatte betydning for økonomisk utvikling»<sup>27</sup>. På slutten av seksti- og starten av syttitallet vokste interessen i OECD for «output» av medlemslandenes politikk<sup>28</sup>. Politikk og forvaltning skulle styres etter tydelige mål, med indikatorer på måloppnåelse.

Som en del av kartleggingsarbeidet over de ulike medlemslands arbeidskraft og utdanningsressurser utviklet organisasjonen ifølge Eide en «solid kompetanse i komparativ statistikk<sup>29</sup>», og en naturlig interesse i lands utdanningssystemer og prestasjoner. OECD har siden den tid hatt en rekke prosjekter med mål om å skaffe mer informasjon om, samt utvikle og forbedre ulike lands utdanningssystemer, deres mål og formål<sup>30</sup>. Andreas Schleicher, spesialrådgiver for utdanning i OECD og PISA-testens «far», siterer den amerikanske professoren W. Edwards Deming og sier det slik: «Uten data er du bare atter en person med en mening»<sup>31</sup>. Det må tilføyes at statistikeren Deming også var svært kritisk til hvordan slike data ble tolket og anvendt i beslutninger. Tilfeldig variasjon ble ifølge Deming fort feiltolket som årsakssammenhenger, og resultater belønnet når man egentlig ikke har innflytelse på dem<sup>32</sup>.

Komplekse mål er imidlertid vanskelige å måle. Det blir ikke lettere når mange land skal bli enige om få parametre. Problemet var ifølge Kjell Eide å finne indikatorer og mål det kunne være politisk enighet om. For hva er egentlig de gode mål og formål med for eksempel utdanningssystemet? I OECD fant man frem til «nærmere 150 ulike måldimensjoner (de fleste dimensjoner også sterkt avhengig av hverandre), og sto så overfor det umulige problem å tillegge hvert av målene en slags politisk vekt». Åpenbare og store uenigheter førte til at «løsningen ble her som ellers i slik sammenheng at man konsentrerte seg om et fåtall måldimensjoner, og lot som om resten ikke eksisterte. «Dermed forsvant [...] muligheten for å bruke denne typen data som styringsinstrument, eller som grunnlag for vurdering av et lands politikk»<sup>33</sup>.

Selv om indikatorarbeidet i stor grad ble lagt dødt på 70-tallet, blusset debatten opp igjen på 80-tallet, da kvalitetsbegrepet for alvor fikk gjennomslag i den internasjonale politiske debatten som skolen. Da, som nå, ønsket politikere og forvaltere målinger på skolens «kvalitet». I OECDs egne fagmiljøer på den tiden var det ifølge Eide «klart for de fleste at kvalitet ikke representerer noen objektiv målestokk», og at kvalitetsbegrepet må sees i sammenheng med utdanningspolitikkenes mål. Dersom sosial utjevning er det viktigste, påvirker det hva man mener med «kvalitet i skolen». Hvis kvalitet er skolens produksjon av arbeidskraft, så er det hvor mange kvalifiserte arbeidere skolen produserer som blir kvalitetsmålet.

PISA-testen er derfor ikke et prosjekt som skal måle hvorvidt skolen lykkes med sitt brede samfunnsmandat, men en test som skal måle elevers ferdigheter og kompetanser på noen områder, og undersøke utviklingen av disse kompetansene over tid.

Når man måler det samme i mange land, er det nettopp for å sammenlikne mellom land. Er det mulig? Prosjektleder for PISA og førsteamanuensis Marit Kjærnsli ved Universitetet i Oslo, sa det slik til Aftenposten: «Vi må ikke overdrive funnene, men det er viktig å se hva norsk skole gjør bra og hvilke utfordringer vi har. Rangeringen er ikke vesentlig, men vi bør se på hva vi eventuelt kan lære av andre land»<sup>34</sup>.

#### 2.1.4 Hva kan vi bruke PISA til?

Hvis testene gir litt men begrenset kunnskap fra mange land, hva kan vi bruke dem til?

De norske PISA-forskerne Kjærnsli, Jensen og Frønes mener trendmålingen PISA gir oss er viktig. Før vi hadde PISA hadde vi heller ingen mulighet til å følge resultater på disse områdene slik vi kan gjøre nå. Sammenligningsmuligheten med andre land er også interessant, men da først og fremst med land som ligner oss selv kulturelt og språklig. Selv om det er spennende å se på for eksempel Sør-Korea og Polen er det først og fremst interessant for oss å se på resultatene i Norge, og de nordiske landene.

Informasjonen vi får fra PISA (og mange andre tester), er imidlertid ingen fasit eller måling på «kvaliteten» i skolen. PISA, som resten av kvalitetsvurderingssystemet, er bare en av mange indikasjoner på hvilket nivå elever holder på deler av kompetansene vi ønsker at de skal besitte. Prøvenes hovedrolle kan dermed sies å være todelt: For det første er trendmålingen sentral. Det at Norges resultater holdes stabile trekkes ofte frem som en styrke, og som noe positivt for skolen. Resultatene i Norge er oppsiktsvekkende stabile over tid.

For det andre skaper prøveresultatene debatt om skolen, og hvordan innholdet læres bort. «Det er riktig at PISA ikke kan si noe sikkert om årsak og virkning, men undersøkelsen gir oss informasjon om sammenhenger som kan være spennende, blant annet som utgangspunkt for viktige diskusjoner», sier Fredrik Jensen, forsker ved Institutt for lærerutdanning og skoleforskning ved UiO<sup>35</sup>. Etter PISA 2001 ble det for eksempel påpekt at Norge manglet en systematisk opplæring i lesing. Den påfølgende faglige debatten om leseundervisning er noe flere forskere trekker frem som positivt for måten vi underviser lesing på i dag. Debatt førte til mer bevissthet om undervisningspraksis. Det er likevel alltid nyttig å stille spørsmål ved hvorvidt all debatt basert på prøveresultater er god debatt. Dersom forståelsen av prøvenes informasjon og funksjon er begrenset, blir også det etterfølgende debatten mindre fruktbar.

PISA gir oss bare informasjon på nasjonalt nivå, ikke fordelt på skoler. «Vi kan ikke si noe om den enkelte skole fordi utvalget av elever ikke er representative for hver enkelt skole, men for hele populasjonen av norske elever», sier Marit Kjærnsli. Forskerne understreker at resultatene dermed ikke brukes som kvalitetsmål på hver enkelt skole, men at indikasjonene vi får fra norske elevers prestasjoner sett under ett er interessant.

Den viktigste innvendingen mot bruken av testresultatene til rangering og utvikling av skolesystemer er at testen ikke kan si noe om hvilke læringsprosesser som er gode, hva slags undervisning som «fungerer», hvilke pedagogiske metoder som er brukt, eller skolens betydning for samfunnet. Det gjør det enda viktigere at hvert land må tolke resultatene i en nasjonal kontekst. Hvert enkelt lands læreplaner og ulik undervisning er derfor viktige premisser og forklaringsvariabler<sup>36</sup>.

Resultatene kan dermed ikke brukes til å gi normative råd til politikere om hvilken skolepolitikk som skal føres. Derfor er det viktig å skille mellom funnene PISA-forskerne drøfter, og rådene<sup>37</sup> OECD gir land basert på datasettene.

Eleven, læreren og skolen får ikke vite hvordan de gjorde det på prøven. Det betyr at prøven ikke kan brukes av skoleledere og lærere i en formativ læringsprosess.

## 2.2 Nasjonale prøver

De første nasjonale prøvene ble innført i 2004 etter enstemmig vedtak på Stortinget<sup>38</sup>, og ble obligatoriske i 2005. Prøvene ble innført på bakgrunn av Søgnerutvalgets delinnstilling NOU 2002: 10 *Førsteklasses fra første klasse— Forslag til rammeverk for et nasjonalt kvalitetsvurderingssystem av norsk grunnopplæring*<sup>39</sup> og hovedinnstilling NOU 2003: 16 *I første rekke*<sup>40</sup>.

Begrunnelsen for innføringen av prøvene var blant annet at:

Det manglet systematiske data på resultater i opplæringen som skoleeiere og skoler kunne benytte seg av.

Norge var ett av få land i verden uten et nasjonalt kvalitetsvurderingssystem.

Skoler og skoleeiere tok ikke systematisk i bruk kartleggings- og læringsstøttende prøver.

Et nasjonalt system for kvalitetsvurdering må ansvarliggjøre skoleeiere som den fremste garantist for gode skoler.<sup>41</sup>

Utvalget foreslo tre kvalitetsbegreper: Strukturkvalitet, prosesskvalitet og resultatkvalitet. På bakgrunn av dette foreslo utvalget:

Et nasjonalt system for kvalitetsvurdering [som] skal legge hovedvekten på resultatkvaliteten og således på det helhetlige læringsutbyttet elever og læringer får i grunnopplæringen.

Det helhetlige læringsutbyttet forankres i opplæringslov § 1-2 og er nærmere beskrevet i læreplanens generelle del. Det helhetlige læringsutbyttet omfatter kunnskaper, ferdigheter og holdninger.

Søgnerutvalgets forslag møtte kritikk fra mange skole- og lærermiljøer. Utdanningsforbundet mente blant annet at «dette forslaget var for snevert og i for stor grad var bygget på måling og kontroll»<sup>42</sup>.

### 2.2.1 Utforming og gjennomføring

De første prøvene ble gjennomført i 2003. Etter kritikk<sup>43</sup> av den opprinnelige prøveutformingen ble formatet endret noe i 2007, men har vært det samme siden. Prøven er utvidet med ankerprøver i 2014. Ankerprøver er, i motsetning til den andre delen av nasjonale prøver, like fra år til år, slik at de kan være grunnlag for sammenligning over tid.

Nasjonale prøver utvikles i samarbeid med lærere og fagpersoner i Utdanningsdirektoratet, fagfolk ved høyskoler, universiteter og nasjonale sentere<sup>44</sup>.

Dagens prøver, som gjennomføres på 5., 8., og 9. trinn av til sammen 360.000 elever<sup>45</sup>, har som formål å vurdere og utvikle elevens grunnleggende ferdigheter i lesing, regning og i deler av faget engelsk. I motsetning til for eksempel PISA-testen, har nasjonale prøver derfor et mål om at prøven skal brukes både som en del av elevens læringsprosess og i skolens/skoleeiers utviklingsarbeid (formativ vurdering) og til måling av ferdigheter mellom skoler og over tid (summativ vurdering).

## Boks 2: Nasjonale prøver

**5. trinn:** Engelsk, regning, lesing  
**8. trinn:** Engelsk, regning, lesing  
**9. trinn:** Regning og lesing

### 2.2.2 Grunnleggende ferdigheter

I Kunnskapsløftet<sup>46</sup> er det definert fem ferdigheter som utgjør grunnleggende forutsetninger for læring og utvikling i skole, arbeid og samfunnsliv: digitale ferdigheter, muntlige ferdigheter, å kunne lese, å kunne regne og å kunne skrive. Alle ferdighetene inngår i læreplanenes kompetansemål, og anses som viktig for å bidra til at elever skal utvikle fagspesifikk kompetanse. Skriveferdigheter og muntlighet inngår ikke i kvalitetsvurderingssystemet.

De grunnleggende ferdighetene nasjonale prøver er ment å undersøke defineres i Utdanningsdirektoratets «Rammeverk for grunnleggende ferdigheter»<sup>47</sup>. I rammeverket skisseres ferdighetenes funksjon og beskriver progresjonen av hver av dem på fem nivåer. De nasjonale prøvene bygger på dette rammeverket.

Som eksempel innebærer grunnleggende ferdigheter i lesing at elever skal kunne forberede, utføre og bearbeide strategier for å forstå tekster, finne informasjon, tolke og sammenholde, reflektere og vurdere kritisk. Den grunnleggende leseopplæringen innebærer for eksempel å mestre lesing av papirtekster og skjermbaserte tekster med konsentrasjon, utholdenhet, flyt og sammenheng<sup>48</sup>.

Grunnleggende ferdigheter deles i Nasjonale prøver igjen inn i fem nivåer. Nivå 1 i lesing når det gjelder informasjonsletting betyr at eleven finner eksplisitt uttrykt informasjon som er sentralt plassert i enkle tekster. På nivå 5 evner eleven å finne detaljert og implisitt informasjon i tekster med ukjent form eller innhold.

### 2.2.3 Resultatene

Resultatene fra prøvene er et *øyeblikksbilde* av ferdighetene elevene har på høsten når de går i 5., 8., og 9. klasse. Prøveutformingen er ikke lik fra år til år, avlegges kun én gang, og vanskelighetsgraden på prøvene kan variere – ellers måtte de vært identiske mellom år. Som sammenligningsgrunnlag kan resultatene fra nasjonale prøver dermed kun brukes til å sammenligne resultatene mellom skoler, klasser og elever det året prøven ble tatt. Man kan for eksempel ikke sammenligne prøveresultatene til klasse 8 B på for eksempel Radøy ungdomsskole i regning fra år 2014 med resultatene til det samme trinnet (eller andre) i 2015.

Det prøven kan fortelle oss er at skole A presterte bedre enn skole B i regning i 2012. Det er imidlertid noen skoler og områder som alltid gjør det godt på Nasjonale prøver. Man kan hevde det viser at kvalitet på skolene er veldig bra. Det kan også være sant, men det er ikke nødvendigvis slik. En skole kan for eksempel ha fantastiske lærere og Norges beste undervisning, men et mer utfordrende elevgrunnlag enn for eksempel naboskolen. Resultatene på prøver behøver ikke da å gjenspeile kvaliteten på selve skolen. En rapport fra NIFU påpeker det samme: «Nasjonale prøver måler elevenes grunnleggende ferdigheter, men de måler ikke kvaliteten på den undervisningen skolen gir»<sup>49</sup>.

Mange faktorer kan forklare variasjon. Elevenes sosioøkonomiske bakgrunn, prøvenes vanskelighetsgrad, læringssituasjonen, eller tilfeldig variasjon. Barn som voksne har gode og dårlige dager og dette påvirker resultatene.

En skole med et svakt elevgrunnlag kan gjøre det middels og det kan tyde på at skolen har evnet å løfte elevene mange hakk. Eller det kan være andre forklaringer. En skole med et veldig godt elevgrunnlag kan også gjøre det middels på nasjonale prøver. Det at de gjør det middels betyr heller ikke at undervisningen eller skolen er middels. Resultatene kan være tilfeldig variasjon, av en særlig vanskelig prøve, eller andre faktorer. Men resultatene kan likevel være et interessant informasjonsgrunnlag.

#### 2.2.4 Hva kan vi bruke nasjonale prøver til?

Nasjonale prøver kan si noe om forskjell mellom skoler og kommuner i et gitt år. Det er verdifullt for å vite om alle barn har likeverdig utdanningstilbud på tvers av bydeler, kommuner og landsdeler.

Nasjonale prøver har frem til 2014 ikke kunnet gi en trendmåling over tid. Nasjonale prøver kan, på samme måte som PISA, heller ikke si noe om er *hvorfor* resultatet er som det er. Mange faktorer kan påvirke resultatet: ressursinnsats i kroner og øre, kvalifikasjonene til lærer og ledelse, elevenes forutsetninger det året, foreldrene til barna på trinnet og skolen det året, pedagogiske opplegg, fravær den uken prøvene ble tatt, og så videre.

Det er også viktig å skille mellom prøvers formative og summative funksjon. En *summativ* prøve legger til grunn at du skal måle det samlede læringsutbytte en elev har i en ferdighet eller fag etter avsluttet periode. Eksamen er en slik summativ vurdering. En *formativ* vurdering har som mål å vurdere elevens ferdighet og kunnskap før læringsperioden er avsluttet, for så å justere læringsprosessen slik at eleven nærmer seg målet. Eleven skal lære av tilbakemeldingene som læreren kan gi på bakgrunn av prøveresultatet. Å ta en prøve gir isolert sett eleven veldig lite informasjon, annet enn at eleven skjønner hva hen ikke får til. Prøvene har en formativ funksjon hvis de gir læreren informasjonen som trengs for å gi framoverretta tilbakemeldinger og tilpasset opplæring.

Professor i pedagogikk ved Universitetet i Bergen, Olga Dyste, beskriver forskjellen på summativ og formativ vurdering slik:

*«Det er vanleg å setje vurdering for læring og vurdering av læring opp mot kvarandre, og likeeins formativ og summativ vurdering. Det er viktig å vere klar over at desse distinksjonane handlar om intensjonen bak vurderinga og ikkje om ulike former for vurdering. Vurdering for læring betyr at vurderinga har som intensjonen å skaffe informasjon som gjer undervisninga og rettleiinga betre, og som fremjar læring. Vurdering av læring (summativ vurdering) har som formål å gi ein karakter eller ein skåre og å rangere eller kvalifisere»<sup>50</sup>.*

Nasjonale prøver er ikke en summativ prøve i tradisjonell forstand, ettersom den ikke tar hensyn til hvor klassene som avlegger prøven er i læringsprosessen i ulike fag. Den forsøker å summere og måle ferdigheter, men tar ikke høyde for hvor barna er i læringsprosessen. Noen barn kan derfor ha kommet lenger i undervisningen, eller kan ha startet i en annen ende med opplæringen. Likevel har prøvene en viss summativ verdi: som øyeblikksbilde og til vektet og balansert sammenlikning mellom skoler.

Nasjonale prøvers funksjon må derfor sies å hovedsakelig være *formativ*: Læringsfremmende prøver på lærer-elev-nivå som brukes til å forbedre elevens ferdigheter. Det har, som vi kommer tilbake til,



betydning for hvordan vi leser og bruker resultatene fra nasjonale prøver for å vurdere skoler og deres kvalitet.

En lærer med ansvar for en klasse på 25 elever, vil ha god oversikt over elevenes ferdigheter sammenliknet med hverandre. Men hvor store forskjellene er kan være nyttig å få et enda bedre bilde av. I skriving har forskning avdekket at elevers ferdighetsnivå kan variere med hele seks år i en enkelt klasse<sup>51</sup>. Et verktøy som gir nivåoversikt kan derfor hjelpe læreren å planlegge og gjennomføre undervisningsopplegg som er tilpasset hver enkelt elevs nivå.

Prøvenes potensielle formative og begrensede summative mål kan imidlertid også dekkes inn av andre prøver, som lærerne i større grad kan styre selv. Utdanningsdirektoratet tilbyr også elektroniske læringsstøttende<sup>52</sup> og karakterstøttende<sup>53</sup> prøver, hvor resultatene ikke rapporteres inn til direktoratet. Resultatene i henholdsvis de læringsstøttende og karakterstøttende prøvene er ment å gi læreren informasjon om elevenes nivå underveis i undervisningsåret, og gi læreren et bedre grunnlag for å gi sluttvurdering. Lærerprofesjonen benytter seg også i stor grad av sine egne utformede formative prøver. Nivåvurdering med følgende tilpasset undervisning er for mange lærere en del av hverdagen.

### 2.2.5 Kritikken av nasjonale prøver

Hovedinnvendingene mot Nasjonale prøver handler mer om prøvens politiske betydning i styringsøyemed, enn om prøvens indikasjon på elevens ferdigheter. Utdanningsforbundet påpeker blant annet at «prøver og tester bare er én av mange [lærings]metoder, og dersom [standardiserte prøver] får for stor vekt vil det undergrave lærerens faglige skjønn under observasjon av elevenes læring»<sup>54</sup>.

Lærere benytter en rekke formative prøver, og Nasjonale prøver er blant dem. Det er når prøvene ilegges for stor vekt på den samlede vurderingen av skolen, og når de brukes i styringsøyemed, at problemene oppstår.

I en evaluering av NKVS gjort av Agderforskning og Danmarks Pædagogiske Universitet (DPU) i 2009<sup>55</sup> kommer det også frem at det finnes systematiske forskjeller i vurderingen av systemet. Evalueringen, som omfatter Nasjonale prøver, er viktig når vi diskuterer verdien av Nasjonale prøver som en summativ og formativ prøve.

I evalueringen av systemet kommer det frem at «kommunene vurderte NKVS mer positivt enn rektorene, som igjen vurderte NKVS mer positivt enn lærerne. Mønsteret som avtegner seg er at jo lengre vekk man er fra den daglige undervisningen, jo mer positivt vurderer man NKVS». Selv om evalueringen omfatter hele NKVS er dette en relevant kritikk av Nasjonale prøver. Hvis prøvens viktigste funksjon er formativ, er det alvorlig at rektorer, ledere og lærere vurderer prøvene som mer relevant som styringsredskap enn som pedagogisk redskap<sup>56</sup>.

Et av hovedproblemene med Nasjonale prøver kan hevdes å være at styresmakter ofte tar prøveresultater som uttrykk for skolens samlede kvalitet. OECD-rapporten «Reviews of evaluation and assessment in education: Norway» peker på det samme, og anbefaler tydeligere kommunikasjon fra myndighetene:

«[...] The national authorities should continue to be clear in their communication that raw national test results are not fulsome measures of student achievement or progress, and even less so of teacher or school quality»<sup>57</sup>.

Evalueringen fra Agderforskning og DPU «viser at den pedagogisk-diagnostiske bruk<sup>1</sup> overfor de enkelte elever i undervisningen er liten sett i forhold til den styringsmessige og kontrollerende bruk. Der er dessuten funnet en tydelig wash-back effekt av disse prøvene, hvilket betyr at bare tilstedeværelsen av prøvene har medført, at det i undervisningen er blitt lagt større vekt på det prøvene måler»<sup>58</sup>.

En annen innvendig er at prøvene er satt til å måle relativt kompliserte kompetanser. Ifølge Ludvigsenutvalget (NOU 2015: 8 *Fremtidens skole*) bør læring i skolen dreie i retning av bredere kompetansemål som å lære å lære, kreativ og sosial kompetanse – på tvers av ulike fag og ferdigheter<sup>59</sup>. Med bredere kompetansemål vil vurdering og måling av kompetansene også bli vanskeligere<sup>60</sup>.

Bruk av prøveresultatene til rangering av skoler er også mye kritisert. Tilhengerne av rangering vil si at skolene som skårer bra er *bedre* skoler, og mener derfor at resultatene bør benyttes til rangering av skoler, slik at foreldre kan se hvilke skoler som er gode og hvilke som er dårlige. Det er imidlertid, som vi har sett, lite som tyder på at man kan benytte prøveresultatene til en slik rangering. Ettersom prøvene ikke kan gi fasitsvar om skolers kvalitet eller informasjon om *hvorfor* skoler gjør det som de gjør, er det ingen grunn til at rangering gir oss relevant eller god informasjon. Snarere tvert om: Det villeder mer enn det opplyser.

## 2.3 Utdanningsdirektoratets kartleggingsprøver

Utdanningsdirektoratet utvikler *kartleggingsprøver*<sup>61</sup> for både grunnskolen og videregående opplæring<sup>62</sup>. Prøvene er ifølge direktoratet utviklet spesielt for å finne de elevene som trenger ekstra oppfølging<sup>63</sup>: De er diagnostiske prøver som skal finne elevene som ligger under, eller er i nærheten av, et visst nivå. Det er skoleeiers ansvar at prøveresultatene følges opp, og at elever som havner under grensen får nødvendig og tilpasset oppfølging<sup>64</sup>.

### 2.3.1 Utforming og gjennomføring

Kartleggingsprøvene utformes av Utdanningsdirektoratet og gjennomføres årlig på de laveste trinnene og på VG1. Noen av kartleggingsprøvene er *obligatoriske* og andre er *frivillige*.

**Obligatoriske kartleggingsprøver** i grunnskolen som skal gjennomføres i alle skoler er:

1. trinn: Lesing
2. trinn: Lesing og regning
3. trinn: Lesing
1. trinn VGO: Lesing og regning

**Frivillige** kartleggingsprøver fra Utdanningsdirektoratet er:

1. trinn: Regning
3. trinn: Engelsk og Regning
4. trinn: Engelsk og Digitale ferdigheter
1. trinn VGO: Engelsk

### 2.3.2 Hva kan vi bruke kartleggingsprøvene til?

Kartleggingsprøvene på 1.- 4. trinn skal bidra til at elevene med særlige behov blir identifisert tidlig, slik at de kan få hjelp til å komme videre i opplæringen. Informasjonen er kun regnet for elever som befinner seg rundt eller under en bekymringsgrense. Kartleggingsprøvene differensierer dermed ikke mellom de

<sup>1</sup> Vurdering som fremmer læring.



sterkeste elevene, slik nasjonale prøver gjør. Prøvenes eneste funksjon er å finne de svake elevene som trenger ekstra oppfølging og hjelp.

Kartleggingsprøvene måler ikke kvaliteten ved skoler. Kunnskapsdepartementet understreker at «det vil være uheldig å bruke resultater fra kartleggingsprøver som styringsinformasjon»<sup>65</sup>.

Innføringen av nasjonalt utformede og felles kartleggingsprøver har erstattet ulike og til dels kostbare prøver av varierende kvalitet.

### 2.3.3 Kritikk av kartleggingsprøver

Kartleggingsprøver er de minst kontroversielle prøvene i kvalitetsvurderingssystemet. Som nevnt over er prøvene designet for å identifisere de svakeste elevene slik at de kan få den oppfølgingen de behøver og har krav på. Det er stor enighet om at vi trenger prøver som kan hjelpe lærerne å identifisere elevene som trenger nettopp slik særskilt oppfølging. Den mest sentrale kritikken som omhandler kartleggingsprøver gjelder derfor mer prøvens bruk og rammene rundt, enn selve prøven. Gjennom en artikkelserie i Klassekampen høsten 2015 har for eksempel Osloskolen fått hard kritikk for å øve på kartleggingsprøver før prøvene avholdes<sup>66</sup>.

Forfatteren av boken «Kunnskapsbløffen», Magnus Marsdal, uttalte til Klassekampen at «[d]et perverse med dette eksempelet med kartleggingsprøvene, er at hele poenget, som er å kartlegge, forsvinner».

Etter at VG i oktober 2015 skrev at de minste elevene på Jeriko skole i Oslo har egne tester før kartleggingsprøver, for å få ned antall elever under den kritiske grensen kommenterte forsker Sture Nome ved Lesesenteret ved Universitetet i Stavanger til samme avis at det er «virkelig kritisk og katastrofalt både for kartleggingsprøvene og flere av elevene at kartleggingsprøvene misbrukes. Det kan lett føre til at de svakeste elevene, de som trenger å bli identifisert for å få spesielt tilrettelagt undervisning, ikke blir fanget opp»<sup>67</sup>.

## 2.4 Tester som informasjonskilde om skolen

Tabellen under gir en oppsummering av vår gjennomgang av PISA-undersøkelsen, Nasjonale prøver og kartleggingsprøver, hvilket nivå de tas på, hvilken måleform de benytter, hvilket formål de har, samt hvem de er nyttige for. Selv om flere av prøvene kan gi interessant informasjon for flere grupper er det først og fremst forskere og lærere/skoler som i størst grad kan benytte seg av informasjonen fra prøvene.

Test	PISA	Nasjonale prøver	Kartleggingsprøver
Nivå	Internasjonal	Nasjonal	Lokal
Måleform	Summativ	Formativ	Formativ
Formål	Trend, kompetansenivå, sammenligning av land.	Kompetansenivå, sammenligning.	Identifisere de svakeste elevene.
Nyttig for	Forskere, myndigheter	Lærere, foreldre, myndigheter	Lærere, foreldre
Kritikk	Måler «smale» ferdigheter, for dominerende i skoledebatten.	Brukes til styring og rangering. «Teach to test».	Det øves på prøvene.

### 2.4.1 Testinformasjonens funksjoner

Det er naturlig å ville vite hva vi får igjen for pengene vi investerer i utdanning. Derfor tester vi – for å vite om og forstå skolen. Et optimalt testsystem ville gitt oss både perfekt styringsinformasjon, og læringsfremmende informasjon. Men får vi nyttig informasjon fra testene som inngår i kvalitetsvurderingssystemet?

Det finnes det ikke entydig og enkel forskning på. Det vil også alltid være en risiko for at vi måler det som kan måles i stedet for det vi lurer på.

Det er her spenningen mellom ulike alternativer oppstår når vi skal finne frem til god styringsinformasjon. Uten god forskning vet vi ikke, og uten kunnskap om skolen kan vi ikke utvikle den til å bli bedre. Men faren er alltid tilstede for at selve forskningen korrupperer det vi prøver å undersøke. Forsker ved Institutt for lærerutdanning og skoleforskning på UiO, Rolf Vegar Olsen, sier det slik:

«Jo mer sentral en kvalitetsindikator blir i de politiske/byråkratiske beslutningsprosessene, og jo større konsekvensene er for de involverte partene, jo større er faren for usunne tilpasningsprosesser som «ødelegger» målingene av dette kvalitetsområdet – og det kan til og med føre til praksiser som heller fører til kvalitetssenking»<sup>68</sup>.

Demings innvendinger om feiltolking av tilfeldige variasjoner understreker at det vil være meningsløst å belønne lærere og rektorer for resultater på for eksempel nasjonale prøver, uten å kontrollere for blant annet sosioøkonomisk bakgrunn og for tilfeldig variasjon<sup>69</sup>. Demings poeng er også at resultater blir svekket hvis en prøver å justere prosesser med utgangspunkt i tilfeldige utslag.

Testinformasjon er komplisert materie. Nyttiggjøring av denne informasjonen krever derfor god innsikt og forståelse av metode, funn og testintensjon. All statistikk og alle testresultater er meningsløse som styringsinformasjon, læringsfremming, og kvalitetssikring hvis ikke forståelsen av *hva* vi faktisk har målt er tilstede.

Nyttiggjøring av testinformasjon kan deles inn i fire: Forskning, styring, læring, foreldreinformasjon.

#### Forskning

Det meste vi kan få av informasjon kan være interessant for forskere. Det samme gjelder informasjonen fra kvalitetsvurderingssystemet. Testresultater og informasjon, kombinert med annen statistikk, undersøkelser og data kan gi oss andre interessante innsikter i skolens virke og resultater.

Et eksempel på dette er PISA+prosjektet:

<http://www.uv.uio.no/ils/forskning/prosjekter/pisa-pluss/index.html>

#### Styring

Ifølge både myndigheter og forskere kan ikke dagens testinformasjon brukes som eneste grunnlag for styring. Det betyr ikke at informasjonen ikke kan være interessant for beslutningstakere, men de må være innforstått med hva resultatene faktisk forteller oss. Ingen av prøvene vi har sett på sier noe om hva som forårsaker resultatene, som er grunnen til at de er så vanskelige å bruke som styringsindikator. Men i kombinasjon med annen informasjon (som elevgrunnlaget) kan de gi oss et varsel om hvilke skoler som gjør det uvanlig bra eller dårlig, men som nevnt ikke hvorfor. Det er i stor grad kommunene som skoleeier som har ansvaret for at informasjonen ikke misbrukes til mål- og resultatstyring.



### Læring

Som drøftet over fungerer tester bare læringsfremmende hvis de har en formativ funksjon. Prøver som tolkes og vektet mot annen informasjon kan brukes i en læringsprosess som ikke er avsluttet, til å tilpasse og forbedre undervisningsopplegget. Flere av prøvene vi har drøftet kan fungere slik. Det er imidlertid behov for mer kunnskap om hvordan prøveresultater skal tolkes og i hvilke prosesser skolene og lærerne kan bruke dem.

### Foresatte

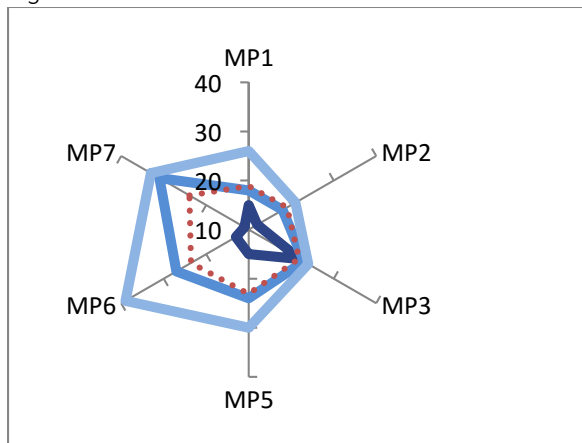
Foreldre ønsker informasjon om hvordan barna gjør det på skolen. Derfor er Nasjonale prøver, PISA-prøven og andre prøver som får stor medieoppmerksomhet blitt et viktig referansepunkt for foreldre når de skal vite hvordan barna gjør det på skolen. Problemet er at de får ikke vite alt det de *vil* vite. Mye av informasjonen fra kvalitetsvurderingssystemet blir misbrukt og mistolket, og egner seg dårlig til å trekke konklusjoner om en gitt skoles kvalitet.

## 2.5 Testvariasjon

Felles for alle prøvene er elevers testvariasjon. Barn presterer ikke jevnt når de testes. Den samme eleven kan derfor ha varierende resultater på prøver med lik vanskelighetsgrad, noe erfaringer fra blant annet skriveprosjektet *Normprosjektet*<sup>70</sup> viser. Prosjektet varte i to år, og hadde 20 skoler, 3088 elever, og 500 lærere som deltakere.

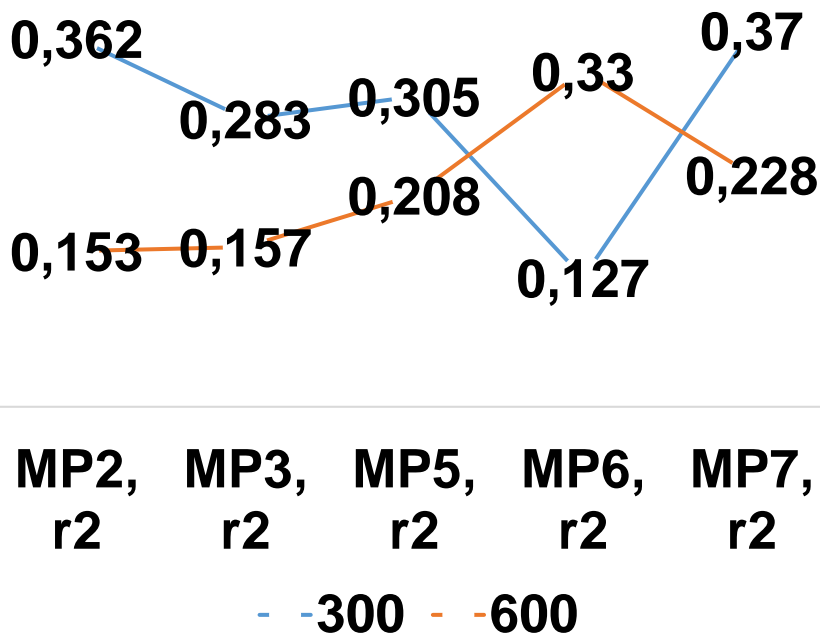
Formålet med prosjektet var å utvikle og ta i bruk en nasjonal forståelse av skrijving, samt utvikle og ta i bruk nasjonale forventningsnormer for skrijving etter fire og syv års opplæring. Erfaringene fra prosjektet viser blant annet at elevers prestasjon på ulike skriveoppgaver varierer veldig.

Figur 1



Figur 1 (over) viser tre elever i 7.klasse, og deres variasjon i prestasjoner på gjentatte skriveprøver. Den innerste sirkelen viser den svakestestende elevens prestasjoner, og den ytterste sirkelen viser den bestestende. Selv om dette bare er eksempler på tre elever, forteller det oss at elevers prestasjoner kan variere. Den svakestestende eleven scorer ved et tilfelle like bra som resten av elevene, men jevnt over svakt ellers. Den bestestende eleven scorer jevnt over godt, men varierer fra å skrive på snittet, til å skrive best.

Figur 2



Figur 2 (over) viser to elevgrupper (300 = elever som gikk i 3.klasse da prosjektet startet, 600 = elever som gikk i 6. klasse da prosjektet startet) som har deltatt i *Normprosjektet*, og de to elevgruppernes (mangel på) stabilitet i skriveferdigheter. "MP" står for målepunkt, dvs. der elevene blir vurdert etter å ha skrevet en ny tekst.

Grafen viser forutsigbarheten i elevenes skriveferdigheter ved seks anledninger eleven skriver tekster og blir vurdert. I tabellen betyr 0,362=36,2%. Det vil si at man ved f.eks. målepunkt 2 (MP2) bare kunne predikere nivået til om lag en tredjedel av elevene. Ved målepunkt 6 var det hele 90 prosent av elevene som presterte på et annet nivå enn pretest (MP1=nivået elevene skrev på før de ble en del av prosjektet).

Med andre ord varierer elevenes skriveferdigheter veldig med undervisning, tilrettelegging, oppgaver, motivasjon og så videre. Elevenes skriveferdigheter er altså ikke stabile og kan ikke generaliseres fra én skrivehendelse (og ett målepunkt).

Erfaringene fra *Normprosjektet* forteller oss at et *øyeblikksbilde* av elevers ferdigheter ikke nødvendigvis gjenspeiler den samlede, generelle skrivekompetansen. En elev kan score dårlig på mandag, bedre på tirsdag, og veldig bra på onsdag. Treff på oppgaven, dagsform, innsatsvilje, elevens situasjon hjemme, brudd med kjæresten og så videre: Det er lett å se for seg hundre forskjellige grunner til at en arbeidstaker kan ha en dårlig dag på jobben – det samme gjelder skoleelever. Alle slike faktorer kan påvirke elevens prestasjon på prøvedagen.

Forekomster av testvariasjon på alle prøver gjør at vi ikke kan bruke resultatene fra Nasjonale prøver som et *dekkende* bilde på elevers ferdigheter gjennom hele året, og i alle lese- og skrivesituasjoner. De kan gi en indikasjon på nivået, men det kan ikke si noe sikkert om enkeltelevens ferdigheter uavhengig av oppgave.

### 3. Gjør prøvene skolen bedre?

Hvilke samfunnsfunksjoner skal skolen ha, og på hvilke måter skal den bidra til enkeltelevens vekst og utvikling? Skal for eksempel skolen bidra til å skape fellesskap, eller skal den først og fremst fremme interessene til hver enkelt elev?

Professor i pedagogikk ved NTNU, Gunn Imsen, kaller forholdet mellom skolen og samfunnet et av de mest sentrale spørsmålene i pedagogikkfaget<sup>71</sup>. For Imsen har skolen fire funksjoner: En *reproduktiv* funksjon, en *produktiv* funksjon, en *identitetsskapende* funksjon, og en *likhetsskapende* funksjon. Kort fortalt går det ut på at skolen skal bringe videre det vi anser å være vår felles kulturarv (reproduktiv), utdanne kandidater til å fylle viktige funksjoner i næringsliv, forvaltning og øvrige samfunnsoppgaver (produktiv), legge til rette for at hver enkelt elev kan vokse og utvikle seg ut fra egne forutsetninger (identitetsskapende), samt skape fellesskap (likhet). Alle fire funksjoner finner vi igjen i opplæringslovens formålsparagraf<sup>72</sup>.

Oppsummert beskrev den gamle formålsparagrafen<sup>73</sup> det godt: «skolen skal gi elevene «god allmennkunnskap, slik at dei kan bli gagnlege og sjølvstendige menneske i heim og samfunn». For en mer utdypende og grundigere redegjørelse er lærerplanens generelle del<sup>74</sup> god lesning.

Et kvalitetsvurderingssystem kan ha flere funksjoner, og ideelt sett kan vi se for oss at prøver i et slik system dekker fire behov:

For det første læreren og skolens informasjonsbehov om elevers progresjon for å erverve seg kunnskapene og egenskapene formulert i lærerplanens generelle del og de fagspesifikke lærerplanene. Resultatene brukes da av lærer og skole for å tilpasse undervisningen og opplæringen til den enkelte elev og klasse slik at de når målene.

For det andre er det et ønske om at prøvene fungerer som et verktøy for politiske myndigheter, som gir kunnskap om hvor det er behov for ressurser, hvordan og hvor reformer og politikk har effekt, og hvordan skolen kan følges opp og styres på en bedre måte.

For det tredje er det ønskelig å kunne gi foreldre dekkende informasjon om skolen er god og om barna lærer det vi har bestemt at de skal lære.

Sist, er det ønskelig at prøveresultater kan gi informasjon som kan benyttes i viktig skoleforskning.

#### 3.1 Hva kan vi bruke prøvene til?

Som vi har sett kan internasjonale og nasjonale prøver gi forskere og fagmiljøer interessante data om både trender og nivå på norske elever innenfor noen få områder. Det er grunn til å tro at både PISA og Nasjonale prøver har ført til gode og viktige diskusjoner i både forskningsmiljøer, lærerutdanningen og på skolene.

Den offentlige debatten om skolen kan sies å ha to formål: For det første har den en demokratisk funksjon: Folk skal kunne mene noe om skolen og bli hørt, og de skal kunne stille politikere til ansvar og kreve handling hvis det går dårlig i skolen. For det andre kan den avdekke forhold om skolen vi ellers ikke kunne fått vite om, ettersom det ikke er mulig å lage systemer som fanger opp all verdifull informasjon. I mangel av slik informasjon kan debatt derfor bidra til å gi myndighetene bedre beslutningsgrunnlag.

I så måte gjør prøver skolen bedre: Informasjon innhentes, og informasjon brukes til å føre nyttige debatter om barns læring. De internasjonale undersøkelsene gjør også at vi retter øynene utover til andre land og andre skolesystemer, bringer interessante idéer hjem, og drøfter dem i en nasjonal og lokal kontekst. Det er likevel ikke slik at det automatisk følger gode løsninger fra debatt. Hvis debatten ikke er opplyst, og hvis informasjonen debatten baserer seg på er feil, kan også løsningene som følger en feilaktig virkelighetsbeskrivelse bli feil. Det understreker behovet for en nøktern debatt om resultatene fra kvalitetsvurderingssystemet.

Prøver med en formativ funksjon bidrar også til at lærere kan legge opp en undervisning som er tilpasset hver enkelt elev. Selv om Nasjonale prøver på langt nær gir et øyeblikksbilde av alle ferdighetene vi skulle ønske at vi kunne måle, er indikasjonen av nivå og ferdigheter god informasjon for lærerkollegiet og skoleledelsen når de skal legge opp undervisningen.

Prøver er derfor viktige av to grunner: For det første bidrar de til mer og viktig debatt om skolens innhold og barns læring. For det andre kan det hjelpe læreren å hjelpe eleven i læringsprosessen.

Undersøkelsene er imidlertid ikke egnet til å si noe om for eksempel skolepolitikken er vellykket, hvilke pedagogiske virkemidler som er gode, hvorvidt flere eller færre lærere er gunstig eller om skolen lykkes med å oppfylle et langt bredere samfunnsmandat enn grunnleggende ferdigheter på noen områder.

Nasjonale prøver kan for eksempel bare gi oss en indikasjon på nivået elevene holder i noen grunnleggende ferdigheter (ikke resten av skolens samfunnsoppdrag) på et gitt tidspunkt, og er best egnet til 1) å stimulere lærerkollegiet og skoleledelsen til refleksjon og diskusjon rundt forskjeller i resultater, elevgruppens egenart og planleggingen av den videre undervisningen og 2) som et læringsfremmende verktøy for elevene. Hovedformålet med prøvene blir derfor læring, fremfor kontrollering. Prøvenes utforming gjør dem ikke egnet til et kvalitetssikringsverktøy for skolesystemet.

Sentrale styresmakter ser ut til å være enige. I Stortingsmeldingen «*På rett vei*» fra 2013 uttrykkes det slik:

«Det nasjonale kvalitetsvurderingssystemet slik det er beskrevet i denne meldingen, gir ikke et heldekkende kunnskapsgrunnlag. Det kan derfor være misvisende å omtale vurderingssystemet som et nasjonalt kvalitetsvurderingssystem, slik det hittil har vært navngitt. Regjeringen vil derfor heretter bruke betegnelsen kvalitetsvurderingssystemet om det kunnskapsgrunnlaget som produseres fra nasjonale myndigheters side, og forutsette at dette suppleres med annet kvalitativt og kvantitativt materiale for vurderings- og utviklingsformål, på både nasjonalt og lokalt nivå. Systemet er primært etablert for å fremme kvalitetsutvikling på den enkelte skole, selv om også nasjonalt nivå bruker systemet.»<sup>75</sup>

Selv om kvalitetsvurderingssystemet ble innført i fravær av en samling av testresultater og statistikk betyr ikke det at det har vært et fravær av kvalitetssikring i skolen før. Førsteamanuensis ved HiOA og Senter for profesjonsstudier Sølvi Mausethagen sier det slik til Viten og Praxis:

«Det har alltid vært intern kontroll i skolen. Vurderinger blir gjort i samråd med læreplan og kolleger, og for eksempel har karakterer vært en form for vurdering som eies av lærerprofesjonen. Det nye på 2000-tallet er den eksterne kontrollen, der lærerne i større grad enn tidligere holdes ansvarlige for elevenes prestasjoner<sup>76</sup>».

I studien «Practices of Data Use in Municipalities and Schools (PraDa)» prøver Maustehagen og hennes kolleger å finne ut «hvordan informasjon om elevenes resultater fra nasjonale prøver og karakterer brukes av kommuner og skoler»<sup>77</sup>. Prosjektet går fra høsten 2014 til våren 2017.

I en artikkel i Utdanningsnytt<sup>78</sup> skriver forskerne Guri Skedsmo, Sølvi Mausethagen, Tine S. Prøitz og Therese N. Hopfenbeck at:

«Evalueringer av implementeringen av Kunnskapsløftet og det nasjonale kvalitetsvurderings-systemet har vist at det er store variasjoner mellom kommuner med tanke på i hvilken grad og hvordan resultatbruk inngår i det lokale kvalitetssystemet, samt at skolene har svært ulik kapasitet og kompetanse til å ta resultatene i bruk». Som uttrykt tidligere er det kommunens ansvar at kvalitetsvurderingssystemet ikke brukes som dekkende styringsinformasjon.

Svaret på om prøvene fører til en bedre skole er derfor ja, gitt det store forbeholdet at de ikke misbrukes som dekkende styringsinformasjon, og at de ikke fører til større prøveregimer på skolene som tar oppmerksomheten vekk fra andre store og viktige fagområder og oppgaver.

### 3.2 Hva mer trenger vi?

Prøvene gir oss nyttig, men ikke tilstrekkelig informasjon. Det betyr ikke at vi trenger flere prøver. Kvalitetsvurderingssystemet består av flere deler, og mange av dem får langt mindre oppmerksomhet i offentlig debatt. Ett eksempel er standpunktkarakter, som settes av lærer i ungdomsskole og videregående skole. Standpunktkarakterene har langt større betydning for den enkelte elevs videre studier og muligheter enn for eksempel nasjonale prøver. Standpunktkarakter er dessuten en mulighet for nettopp å vurdere en større bredde i kompetanse enn det vi klarer med standardiserte prøver. Likevel stiller forskere spørsmål ved om lærere i tilstrekkelig grad veiledes i karaktersettingen. En studie fra NIFU viser stor variasjon i karaktersettingen og bemerker at «dersom grad av statlig regulering kan tolkes som et signal om samfunnets anseelse av et felt, synes eksamen å være oppfattet som svært viktig sammenliknet med standpunktkaraktersetting»<sup>79</sup>. En annen og senere studie fra NIFU problematiserer oppfatningen om at eksamenskarakter og standpunktkarakter skal likne og spør om eksamenskarakter kan og bør brukes som kalibrering av standpunktkarakter – når den siste nettopp er ment å gi en bredere vurdering over lengre tid<sup>80</sup>.

Det er også behov for mer kunnskap og rutiner om tolkning og anvendelse av ulike prøveresultater i skolen og hos myndigheter. Politikere, skoleledere, lærere og foreldre har ikke nødvendigvis tilstrekkelig forståelse av hva prøvene sier og ikke sier til å sette sammen og tolke informasjon på en fornuftig måte. Som en del av forskningsprosjektet PraDa kartlegges slike rutiner ved ulike skoler. Foreløpige resultater viser til begrenset kunnskap om hva prøveresultater kan si noe om og brukes til, og varierende rutiner for hvordan informasjonen anvendes<sup>81</sup>.

## 4. Konklusjon

Skolen er vår mest formende samfunnsinstitusjon. Den skal sikre like muligheter for alle elever på tvers av økonomisk og kulturell bakgrunn, og skape gode mennesker. Fellesskolen gjør derfor det motsatte av å gjøre alle barn like: Den sørger for at vi kan skape vår egen identitet og egenart, samt leve den ut i et samfunn hvor individ og fellesskap lever og arbeider sammen.

Et av spørsmålene vi stilte oss innledningsvis er om kvalitetsvurderingssystemet er i stand til å gi oss svaret på om skolen lykkes. Svaret på det er nei, ikke hele. Ta for eksempel Imsens fire funksjoner eller skolens detaljerte samfunnsmandat i formålsparagrafen og lærerplanens generelle del: kvalitetsvurderingssystemet gir oss ikke svaret på skolens suksess i noen av dem.

Gode resultater på prøvene alene kan gi en pekepinn men vil ikke alene svare på om skolen lykkes med å utvikle «gagnlege og sjølvstendige menneske». Systemet måler for eksempel ikke ferdigheter i skriving og muntlighet. Langt mindre brede kompetanse mål som kreativitet, sosial kompetanse og evne til å lære.

Det er ikke nødvendigvis slik at elever som lykkes i det nasjonale kvalitetssikringssystemet lykkes med resten.

### 4.1 Beholde eller forkaste?

Gir prøvene likevel nyttig informasjon? Ja. De gir oss en pekepinn på læring av grunnleggende ferdigheter. Naturligvis er det en sammenheng mellom å mestre lesing og regning, og det å mestre andre deler av skolen.

Hvis vi frykter at prøvene brukes feil og at det fører til at skolen ikke oppfyller sitt samfunnsoppdrag, kunne en nærliggende politisk konsekvens være å kutte ut hele kvalitetsvurderingssystemet. Selv om de fleste er enige i at informasjonen ikke er dekkende for å måle skolens samfunnsoppdrag, brukes resultatene til å vinne valg, sette dagsorden om skolens kvalitet og iverksette reformer. Det er vanskelig å la være å slå politisk mynt på PISA-resultatene og Nasjonale prøver.

Som vi har sett kan for mye fokus på smale indikatorer 1) gi et forvrengt bilde av måloppnåelse, og 2) vri oppmerksomheten bort fra andre viktige oppgaver. Kutt av hele kvalitetsvurderingssystemet kunne derfor fungert som en god påminnelse til både velgere, skoleledere, kommuner, pressen og elever om at prøvene *alene* overhodet ikke er et interessant kvalitetsmål for skolen, og at vi fint kan lykkes med skolen uten dem.

Vi mener likevel at vi bør beholde systemet. PISA-undersøkelsen, Nasjonale prøver og kartleggingsprøver kan være gode verktøy i norsk skole. Prøvene gir en indikasjon på kvalitet, men viktigst er at de fører til debatt og at lærerne kan dra nytte av verktøyene i sitt daglige virke. Kartleggingsprøvene er et ubetinget gode. En prøve som har som eneste funksjon å identifisere elever som trenger ekstra hjelp er bare bra.

Kvalitetsvurderingssystemet kan også friskmeldes fra mye av den vanlige kritikken: Det tar ikke for mye tid, det koster ikke for mye penger. I debatten om tidstyver i skolen finner ikke vi noe som tyder på at det er prøvene i kvalitetsvurderingssystemet som tar tid. Det kan likevel ikke sies for ofte: Prøvene er bare en indikasjon som lærere og ledere må drøfte i en lokal skole- og elevkontekst. Det kan være et godt supplement til viktigere faktorer som gode lærers utdannings- og fagkompetanse. Til syvende og sist er det lærere og skoleledere i fellesskap som må finne ut av hvordan informasjonen skal benyttes





og forfølges, og det er kommunenes ansvar å sørge for at informasjonen ikke misbrukes til styring og ressursfordeling.

Svaret er altså ikke avvikling av systemet, men bedre kunnskap om hva det faktisk kan måle og hva det kan brukes til. Eller sagt på en annen måte: Enighet om prøvenes begrensninger.

## 4.2 Hva nå?

Det er avgjørende av prøvene brukes fornuftig både som verktøy i skolen og som informasjonsgrunnlag ellers.

For det første, god skår på prøvene må ikke bli et mål i seg selv. Prøvene kan ikke måle alt. Dersom prøveresultater tillegges for mye vekt sammenliknet med annen informasjon, kan lærerne stå i skvis mellom å velge undervisningsmetoder og -ressurser som er best egnet til å nå læreplanmålene eller å gjøre det godt på prøvene (målforskyvning). Campbells lov<sup>82</sup>, etter professor Donald T. Campbell, understøtter dette: Jo mer en kvalitativ indikator brukes som grunnlag for beslutninger, jo mer vil den forvrengte og korrumpere de prosessene den var ment å måle<sup>83</sup>. Det kan for det første sette lærernes profesjonelle faglighet under press, men kanskje viktigere: Prøvene kan da, ironisk nok, gjøre det enda vanskeligere å implementere læreplaner, altså å realisere politikken i skolen.

For det andre må kunnskap og rutiner i skolen om hva prøvene kan brukes til og hvordan manges steder styrkes. Det innebærer også behov for en mer opplyst offentlig debatt om hva prøvene sier og ikke sier oss. I forlengelsen av dette kan det være nyttig å følge med flere typer data når prøveresultatene legges frem: Når sosial bakgrunn fremdeles er den viktigste forklaringsvariabelen for skoleresultater, er det eksempelvis interessant å se resultater fra nasjonale prøver vektet mot sosioøkonomisk bakgrunn – som vi vet er en av de viktigste forklaringsvariablene for resultater i skolen.

For det tredje må vi sørge for at annen informasjon også anvendes og styrkes, som brukerundersøkelser, standpunkt karakterer og andre deler av kvalitetsvurderingen. Det er påfallende hvor stor oppmerksomhet PISA-undersøkelsen, Nasjonale prøver og kartleggingsprøver får sammenliknet med andre deler av kvalitetsvurderingssystemet. Først og fremst fordi nevnte prøver er en langt mindre grundig vurdering av elever enn andre vurderinger lærere gjør kontinuerlig, men også fordi de får langt mindre betydningsfulle konsekvenser for elever enn for eksempel eksamen og standpunkt karakterer. Flere lærere etterspør tydeligere retningslinjer for fastsettelse av standpunkt karakterer<sup>84</sup>.

Fra et styringsperspektiv er det ønskelig med mer og bedre styringsinformasjon enn den vi har i dag. Norge bruker 98 milliarder på utdanning<sup>85</sup> årlig, og skolen er kontinuerlig gjenstand for politisk debatt. Behovet for kunnskap, innsikt og sikker viten om skolen kommer derfor ikke til å bli mindre.

Mer både kvalitativ og kvantitativ følgeforskning av enkelt skoler, elever og klasser, kombinert med data fra annen forskning fra alle samfunnsområder kan gi oss en bedre og annen innsikt enn det kvalitetsvurderingssystemet forsøker å gi oss.

Videre er skolen som alle andre offentlige etater dårlig tjent med instrumentell målstyring. Mål- og resultatstyring har på mange samfunnsområder gitt oss uløste utfordringer. Vi måler for mye og ikke de riktige tingene, vi tenker for lite nytt og vi skaper unødig byråkrati, er blant innvendingene. Endringer i statlige styringsprinsipper over tid er analysert av mange forskere<sup>86,87</sup>.

Gudmund Hernes stilte spørsmål i 2007 om gjeldende styringsprinsipper og minnet om betydningen av tillit og sosial kapital i den norske modellens høye produktivitet. Det er et paradoks at mens både ledelsesfaget og internasjonal forskning på offentlig forvaltning ser til nye løsninger, er norsk forvaltning påfallende lojal til hierarkiske strukturer og tanken om det rasjonelle økonomiske individ. I Danmark har man lansert begrepet *Tillitsreform* som operasjonaliseringen av en revidert styringsform, innenfor rammen av mål- og resultatstyring. Tillitsreform er hverken frislipp av kontroll eller en enhetlig modell som kan vedtas innført. Det er et sett med prinsipper som flytter makt og kontroll. Det kan for eksempel bety at vi i større grad må stole på, og legge til rette for, at lærerprofesjonen drøfter, diskuterer og bruker den kunnskapen de har tilgjengelig til å ta gode undervisningsvalg for barna våre. Det krever at vi har lærere som kan drive sitt virke under et regime med tillit og muligheter fra skolens ledelse. I en slik verden lever kvalitetsvurderingssystemet godt.

I neste notat skal vi se på styring av skolen.

*Notatet er skrevet av rådgiver Ola Magnussen Rydje høsten 2015. Forfatteren står ansvarlig for alle eventuelle feil og mangler i dokumentet. Ta gjerne kontakt dersom du finner slike.*

*Tankesmien Agenda vil rette en stor takk til flere som har tatt seg tid til å svare på faglige spørsmål og utfordringer underveis:*

*Professor Gunn Imsen, Institutt for pedagogikk og livslang læring, NTNU.*

*Førsteamanuensis og prosjektleder Marit Kjærnsli, rådgiver Tove Frønes, forsker Fredrik Jensen, forsker Rolf Vegar Olsen, ved Institutt for lærerutdanning og skoleforskning Universitetet i Oslo.*

*Professor Kjell Lars Berge, Universitetet i Oslo.*

*Professor emeritus Svein Sjøberg, Universitet i Oslo.*

*Arnhild Bie-Larsen, Trond Harsvik og Harald Skulberg, Utdanningsforbundet.*

*Anne Finborud og Terje Moen, Skolenes Landsforbund.*

*Simon Malkenes, lektor og forfatter.*

## 5. Sluttnoter

- <sup>1</sup> <https://www.difi.no/rapporter-og-undersokelser/statistikk-og-undersokelser/innbyggerundersokelsen-2015/hva-mener-innbyggerne/kommunale-og-fylkeskommunale-tjenester>
- <sup>2</sup> Meld. St. 20 (2012-2013). *På rett vei*. Kunnskapsdepartementet. S. 146
- <sup>3</sup> For mer bakgrunn og informasjon om både norsk og internasjonal kontekst, se f.eks. forsker Tine S. Prøitz rapport (2015) *Formål med vurdering og vurdering av fagovergripende kompetanser*: <http://nettsteder.regjeringen.no/fremtidensskole/files/2014/05/Form%C3%A5l-med-vurdering-og-vurdering-av.pdf> [Lesedato 17.12.2015]
- <sup>4</sup> Blogg, Kristin Clemet. [http://clemet.blogg.no/1315665932\\_stoltenbergs\\_kunnskap.html](http://clemet.blogg.no/1315665932_stoltenbergs_kunnskap.html) [Lesedato: 17.12.2015]
- <sup>5</sup> <https://www.stortinget.no/no/Saker-og-publikasjoner/Publikasjoner/Innstillinger/Budsjett/2003-2004/innb-200304-012/4/#a2.1.1>
- <sup>6</sup> Meld. St. 20 (2012-2013). *På rett vei*. Kunnskapsdepartementet. S. 148 (tabell)
- <sup>7</sup> Ibid, s. 150.
- <sup>8</sup> Ibid, s. 147.
- <sup>9</sup> Ibid, s. 147.
- <sup>10</sup> Dagsavisen (2015). *Oslo-elevene langt foran på nasjonale prøver*. <http://www.dagsavisen.no/oslo/oslo-elevene-langt-foran-p%C3%A5-nasjonale-pr%C3%B8ver-1.437273>. [Lesedato: 17.12.2015]
- <sup>11</sup> NIFU, rapport 4/2013 Evaluering av nasjonale prøver som system, s. 160.
- <sup>12</sup> Meld. St. 20 (2012-2013). *På rett vei*. Kunnskapsdepartementet. S. 157.
- <sup>13</sup> <http://www.oecd.org/pisa/aboutpisa/>
- <sup>14</sup> <https://snl.no/PISA>
- <sup>15</sup> <http://www.oecd.org/pisa/aboutpisa/>
- <sup>16</sup> <http://www.uv.uio.no/ils/forskning/prosjekt-sider/pisa/vanlige-sporsmal/index.html>
- <sup>17</sup> <http://www.udir.no/Tilstand/Internasjonale-studier-/PISA---Programme-for-International-Student-Assessment/>
- <sup>18</sup> Hopmann, Stefan T.; Brinek G.; Retzl, M. (eds.) (2007). *PISA zuzolge PISA – PISA According to PISA*. Wien: Lit.
- <sup>19</sup> Oppgitt av forskerne Marit Kjærnsli og Fredrik Jensen i samtale med forfatter av notatet, Ola Magnussen Rydje, 11.11.2015.
- <sup>20</sup> Rutkowski mfl 2014b:85.
- <sup>21</sup> Rutkowski mfl 2014b:80.
- <sup>22</sup> Rutkowski mfl 2014b:79.
- <sup>23</sup> Samtale med forfatter av notatet, 11.11.2015.
- <sup>24</sup> Alle rapportene ligger her: <http://www.uv.uio.no/ils/forskning/prosjekt-sider/pisa/publikasjoner/>
- <sup>25</sup> Lesing og naturfag siden 2009.
- <sup>26</sup> Dagbladet (2014). Kommentator Gudleiv Forr, *Penger og PISA*. [http://www.dagbladet.no/2014/07/23/kultur/meninger/hovedkommentar/kommentar/pisa/3446816\\_1/](http://www.dagbladet.no/2014/07/23/kultur/meninger/hovedkommentar/kommentar/pisa/3446816_1/) [Lesedato 17.12.2015]
- <sup>27</sup> Norsk Pedagogisk Tidsskrift, 4/1990: s. 220.
- <sup>28</sup> Norsk Pedagogisk Tidsskrift, 5/1990: s. 256.
- <sup>29</sup> Norsk Pedagogisk Tidsskrift, 4/ 1990: s. 219.
- <sup>30</sup> Norsk Pedagogisk Tidsskrift, 5/1990: s. 256.
- <sup>31</sup> The Guardian (2013). *The OECD's Pisa delivery man*. <http://www.theguardian.com/education/2013/nov/26/pisa-international-student-tests-oecd>
- <sup>32</sup> Deming, W. Edwards (1986): *Out of the Crisis*. Cambridge, Mass: Massachusetts Institute of Technology, Center for Advanced Engineering Study
- <sup>33</sup> Norsk Pedagogisk Tidsskrift, 5/1990: s. 256.
- <sup>34</sup> Aftenposten (2013). *Forsker: PISA har endret det norske skolesystemet*. <http://www.aftenposten.no/nyheter/Forsker-PISA-har-endret-det-norske-skolesystemet-7392319.html> [Lesedato: 17.12.2015]
- <sup>35</sup> Oppgitt i samtale med forfatter av notatet, 11.11.2015.

- <sup>36</sup> Utdanningsnytt (2005). *Skjeve uttalelser om Pisa og Timss*. <http://utdanningsnytt.no/debatt/2005/januar/skjeve-uttalelser-om-pisa-og-timss/> [Lesedato 17.12.2015]
- <sup>37</sup> <http://www.udir.no/Tilstand/Internasjonale-studier-/PISA---Programme-for-International-Student-Assessment/>
- <sup>38</sup> <https://www.regjeringen.no/no/dokumenter/stprp-nr-1-tillegg-nr-3-2002-2003-/id435850/?q=&ch=2>
- <sup>39</sup> NOU 2002: 10. *Førsteklasses fra første klasse*. <https://www.regjeringen.no/no/dokumenter/nou-2002-10/id145378/?ch=1&q=>
- <sup>40</sup> NOU 2003: 16. *I første rekke*. <https://www.regjeringen.no/contentassets/37a02a7bd6d94f5aacd8b477a3a956f3/no/pdfs/nou200320030016000dddpdfs.pdf>
- <sup>41</sup> NOU 2002: 10 *Førsteklasses fra første klasse*.
- <sup>42</sup> <https://www.utdanningsforbundet.no/Hovedmeny/Vi-mener/Nasjonale-prover/> [Lesedato 17.12.2015]
- <sup>43</sup> Meld. St. 20 (2012-2013). *På rett vei*. Kunnskapsdepartementet. S. 150.
- <sup>44</sup> <http://www.udir.no/Vurdering/Nasjonale-prover/#Hva-er-nasjonale-prover> [Lesedato 17.12.2015]
- <sup>45</sup> <https://www.regjeringen.no/no/aktuelt/Bruk-resultatene-til-elevenes-beste-/id2342784/> [Lesedato 17.12.2015]
- <sup>46</sup> <http://www.udir.no/Lareplaner/Kunnskapsloftet/> [Lesedato 17.12.2015]
- <sup>47</sup> [http://www.udir.no/globalassets/upload/larerplaner/lareplangrupper/rammeverk\\_grf\\_2012.pdf](http://www.udir.no/globalassets/upload/larerplaner/lareplangrupper/rammeverk_grf_2012.pdf) [Lesedato 17.12.2015]
- <sup>48</sup> [http://www.udir.no/globalassets/upload/larerplaner/lareplangrupper/rammeverk\\_grf\\_2012.pdf](http://www.udir.no/globalassets/upload/larerplaner/lareplangrupper/rammeverk_grf_2012.pdf), s.10 [Lesedato 17.12.2015]
- <sup>49</sup> NIFU, rapport 4/2013 Evaluering av nasjonale prøver som system, s. 160.
- <sup>50</sup> Bedre skole nr 4/2008. Professor i pedagogikk, Olga Dyste. *Klasseromsvurdering og læring*.
- <sup>51</sup> Larsson, Skrivförmåga.
- <sup>52</sup> <http://www.udir.no/Vurdering/Laringsstottende-prover/> [Lesedato 17.12.2015]
- <sup>53</sup> <http://www.udir.no/Vurdering/Karakterstottende-prover/> [Lesedato 17.12.2015]
- <sup>54</sup> Temanotat 9/2011: Systemvurdering av norsk utdanning – hvordan kan vurdering forbedre grunopplæringen?, s. 22.
- <sup>55</sup> <http://www.agderforskning.no/reports/fou82009nkvsrapport.pdf> [Lesedato 17.12.2015]
- <sup>56</sup> Ibid, s. 16.
- <sup>57</sup> Nusche, L., Earl, L., Maxwell, W and Schrewbridge, C.: OECD REVIEWS OF EVALUATION AND ASSESSMENT IN EDUCATION: NORWAY © OECD 2011 s. 63.
- <sup>58</sup> <http://www.agderforskning.no/reports/fou82009nkvsrapport.pdf> s. 11. [Lesedato 17.12.2015]
- <sup>59</sup> NOU 2015: 8 *Fremtidens skole* Kunnskapsdepartementet. S. 18.
- <sup>60</sup> Ibid, s. 78.
- <sup>61</sup> Eksempel på oppgaver i kartleggingsprøver: <http://www.udir.no/globalassets/filer/vurdering/kartlegging/eksempeloppgaver-kpregning1-trinn.pdf> [Lesedato 17.12.2015]
- <sup>62</sup> Temanotat 9/2011, Utdanningsforbundet: Systemvurdering av norsk utdanning, s. 13.
- <sup>63</sup> <http://www.udir.no/Vurdering/Kartlegging-gs/#Oppfolging-av-resultater-pa-skoleniva> [Lesedato 17.12.2015]
- <sup>64</sup> Ibid.
- <sup>65</sup> Meld. St. 20 (2012-2013). *På rett vei*. Kunnskapsdepartementet. S. 152.
- <sup>66</sup> Klassekampen (2015). *Øver på kartleggingsprøve*. <http://www.klassekampen.no/article/20150929/PLUSS/150929792> [Lesedato 17.12.2015]
- <sup>67</sup> Verdens Gang (2015). *Forsker: - Kritisk og katastrofalt*. <http://www.vg.no/nyheter/innenriks/skole-og-utdanning/forsker-kritisk-og-katastrofalt/a/23541808/> [Lesedato 17.12.2015]
- <sup>68</sup> Utdanningspolitisk forum, Oslo Arbeiderparti.
- <sup>69</sup> Se for eksempel <http://www.vg.no/nyheter/innenriks/skole-og-utdanning/her-er-de-nye-skjemaene-som-erstatte-rektor-kontraktene/a/23413594/> og <http://www.vg.no/nyheter/innenriks/skole-og-utdanning/oslo-rektorer-beloennes-for-elevresultater/a/10107214/>
- <sup>70</sup> <http://norm.skriresenteret.no/> [Lesedato 17.12.2015]
- <sup>71</sup> Hva er PEDAGOGIKK, Imsen, s. 127.

- <sup>72</sup> § 1-1, Lov om grunnskolen og den videregående opplæringen (opplæringsloven): [https://lovdata.no/dokument/NL/lov/1998-07-17-61/KAPITTEDeL\\_1#KAPITTEL\\_1](https://lovdata.no/dokument/NL/lov/1998-07-17-61/KAPITTEDeL_1#KAPITTEL_1). [Lesedato 17.12.2015]
- <sup>73</sup> Regjeringen.no: [Formålsparagrafens historie](#). [Lesedato 17.12.2015]
- <sup>74</sup> Lærerplanens generelle del. [http://www.udir.no/upload/larerplaner/generell\\_del/generell\\_del\\_lareplanen\\_bm.pdf](http://www.udir.no/upload/larerplaner/generell_del/generell_del_lareplanen_bm.pdf) [Lesedato 17.12.2015]
- <sup>75</sup> Meld. St. 20 (2012-2013). *På rett vei*. Kunnskapsdepartementet. S. 147.
- <sup>76</sup> Viten og praksis. *NASJONALE PRØVER HAR ENDRET NORSK SKOLE*. [http://vitenogpraksis.no/Skole-og-utdanning/Nasjonale-proever-har-endret-norsk-skole?utm\\_source=hoot&utm\\_medium=some&utm\\_campaign=gensome](http://vitenogpraksis.no/Skole-og-utdanning/Nasjonale-proever-har-endret-norsk-skole?utm_source=hoot&utm_medium=some&utm_campaign=gensome) [Lesedato 17.12.2015]
- <sup>77</sup> <http://www.hioa.no/Forskning-og-utvikling/Hva-forsker-HiOA-paa/FoU-SPS/prosjekter/Bruk-av-elevresultater-data-i-norske-skoler-og-kommuner-PraDa> [Lesedato 17.12.2015]
- <sup>78</sup> Utdanningsnytt. *Nye verktøy og rutiner for resultatoppfølging*. Guri Skedsmo, Sølvi Mausethagen, Tine S. Prøitz og Therese N. Hopfenbeck <http://www.utdanningsnytt.no/debatt/2015/september/nye-verktoy-og-rutiner-for-resultatoppfolging/>
- <sup>79</sup> Borgen, Jorunn Spord og Tine Prøitz (2010): Rettferdig standpunkt vurdering – det (u)muliges kunst?. NIFU rappr 16/2010
- <sup>80</sup> Hovdhaugen, Elisabeth, Idunn Seland, Berit Lødding, Tine S. Prøitz og Nils Vibe (2014): Karakterer i offentlige og private videregående skoler. En analyse av eksamens- og standpunkt karakter i norsk og matematikk og rutiner for standpunkt vurdering i offentlige og private videregående skoler. NIFU rapport 24/2014.
- <sup>81</sup> <http://www.utdanningsnytt.no/debatt/2015/september/nye-verktoy-og-rutiner-for-resultatoppfolging/>
- <sup>82</sup> <http://dianeravitch.net/2012/05/25/what-is-campbells-law/> [Lesedato 17.12.2015]
- <sup>83</sup> Donald T. Campbell: "assessing the impact of planned social change" The Public Affair Center, Dartmouth College, Hanover New Hampshire, USA, des. 1976.
- <sup>84</sup> Bedre Skole (2011), nr. 1, Tine S. Prøitz og Jorunn S. Borgen: *Rettferdig standpunkt vurdering – på grensen til det umulige?* [https://www.utdanningsforbundet.no/upload/Tidsskrifter/Bedre%20Skole/BS\\_nr\\_1\\_11/BS\\_1-11\\_Pr%C3%B8itz\\_og\\_Borgen.pdf](https://www.utdanningsforbundet.no/upload/Tidsskrifter/Bedre%20Skole/BS_nr_1_11/BS_1-11_Pr%C3%B8itz_og_Borgen.pdf) [Lesedato 17.12.2015]
- <sup>85</sup> Grunnopplæringen.
- <sup>86</sup> Benington og Hartley (2001):
- <sup>87</sup> Osborne, Stephen P. (Ed). (2010). *The New Public Governance?: Emerging Perspectives on the Theory and Practice of Public Governance*. Routledge