

De nationale test – mellem performancemåling og pædagogisk evaluering

De senere års debat om de nationale test har fokuseret på metoderne og kvaliteten i målingerne. I denne artikel fokuserer vi i stedet fokusere på testenes formål. Vi hævder, at de nationale test er blevet tildelt to formål, for det første at måle folkeskolens og de enkelte skolars performance med hensyn til elevernes færdigheder, for det andet at give input til pædagogisk evaluering af elevernes læring. Disse to formål har i lovgivning og debat været flettet sammen på uigennemskuelige måder, som samtidig har været forbundet med interesserne hos forskellige grupper, herunder lærere, skoleledere, kommuner, stat og forskere. I artiklen søger vi at efterspore disse formål i beslutningerne om de nationale test og vurdere opfyldelsen af dem.

Nøgleord: test; evaluering; pædagogik; uddannelsespolitik; skolepolitik

I foråret 2020 besluttede den socialdemokratiske regering at begrænse anvendelsen af de nationale test. Gennemførelsen af test er blevet frivillig for flertallet af skoler, men der skal stadig gennemføres test ved et repræsentativt udsnit af skoler, og ligesom de 20 procent dårligt præsterende skoler skal teste. Beslutningen kom efter en evaluering af de nationale test, som viste betydelig usikkerhed, hvad angår metoder og resultater (BUVM, 2020, s. 24), og efter en valgkamp året før, hvor en række partier – herunder regeringens støttepartier – havde krævet testene afskaffet eller gjort frivillige (Rasmussen, 2019).

Indførelsen af de nationale test, som skulle gennemføres for alle elever og på alle skoler, blev besluttet af Folketinget i 2006. Vedtagelsen skete på baggrund af flere års politisk debat præget af bekymring over danske

skoleelevers resultater i OECD's PISA-undersøgelse og med afsæt i OECD's evaluering af den danske grundskole et par år tidligere (Uddannelsesstyrelsen, 2004). Efter et intensivt og problemfyldt udviklingsarbejde blev testene første gang gennemført i skoleåret 2009-10.

Den pædagogiske og uddannelsespolitiske debat om de danske nationale test har handlet om to forbundne problemstillinger. Den ene gælder de valgte målemetoders validitet og pålidelighed, altså hvorvidt testene faktisk måler elevernes faglige viden og kunnen, og med hvilken grad af sikkerhed de måler. Den anden problemstilling gælder formål og anvendelse, altså med hvilke formål de nationale test blev indført, forholdet mellem disse formål og den faktiske anvendelse af testsystemet. I denne artikel fokuserer vi på den sidstnævnte problemstilling. Vi søger at belyse to spørgsmål: 1) Hvordan blev formålene med de nationale test formuleret i beslutningsprocesserne frem mod vedtagelsen og i lovgivningen?



2) Hvordan er de nationale test blevet anvendt, og hvilke virkninger har de haft?

Vi går ud fra, at de nationale test blev indført med to formål; dels at måle folkeskolens og de enkelte skolers performance i forhold til elevernes viden og færdigheder, dels at give input til formativ pædagogisk evaluering af elevernes læring. Som flere forskere har gjort opmærksom på (se eksempelvis Bundsgaard og Kreiner, 2019), har disse to formål været flettet sammen på ofte uigennemskuelige måder i den uddannelsespolitiske debat, i lovgivningen og i implementeringen af testsystemet.

I artiklen giver vi først en evalueringsteoretisk præcisering af performancemåling og formativ pædagogisk evaluering som to forskellige modeller med hver deres vidensformer og anvendelsesmuligheder. Derefter diskuterer vi den uddannelsespolitiske proces, hvorigennem ideen om nationale test blev fremsat, debatteret, legitimeret og implementeret. En hovedpointe er, at testenes formative pædagogiske formål var afgørende for legitimering og vedtagelse. På grundlag af eksisterende empirisk forskning belyser vi så, hvordan de pædagogiske test har fungeret som henholdsvis performancemåling og formativ pædagogisk evaluering. Vi trækker især på en grundig undersøgelse af anvendelsen af de nationale test, som er gennemført af forskere ved VIVE (Bjørnholt m.fl., 2020), men vi inddrager også andre undersøgelser.

Performancemåling og formativ pædagogisk evaluering

Forskellige typer af evaluering har hver deres karakteristiske genstandsfelter, værdier, metoder og anvendelser (Krogstrup, 2016; Andersen, Wahlgren og Wandel, 2017). Det gælder også de to typer evaluering i det danske skolesystem, som de nationale test har skullet varetage, nemlig performancemåling og formativ pædagogisk evaluering. Som grundlag for at vurdere erfaringerne med testsystemet vil vi kort redegøre for disse to begreber og metoder.

Performancemåling, som også kaldes præstationsmåling, har til formål at måle resultaterne af en indsats ved hjælp af et batteri af præstationskriterier og tilknyttede indikatorer. Ved performanceevaluering er man interesseret i at monitorere

en indsats (her skolers undervisning) og skabe viden om, hvad der kommer ud af indsatsen – forstået som det, der i sammenhængen kan måles og vejes (Krogstrup, 2016, s. 137ff.). I forbindelse med de nationale test er tankegangen, at kvaliteten af undervisning i den danske grundskole kan måles, ved at man lader elever på udvalgte klassetrin gennemføre en standardiseret faglig test, og at elevernes resultater heraf indikerer kvaliteten af undervisning og læring. Monitoreringen sker således ved centralt fastlagte procedurer for gennemførelse af test – ved angivelse af klassetrin, tidspunkter, vejledninger og fagrække. En generel kritik af performanceevaluering er usikkerheden omkring validitet og dermed om sikkerheden i opgjorte resultater. En anden generel kritik af performanceevaluering har fokus på den forenkling, der sker af et socialt felt, når indsatsen underlægges kravet om, at de skal kunne måles. Inden for det sociale felt uddannelse og læring kan det fænomen, som kaldes ”teaching to for the test”, indebære, at lærere indretter deres undervisning og materialer, så de matcher kravene i de nationale test (Bundsgaard, 2016).

I transformationen fra det levede liv i klasseværelset til opstillingen af hovedsageligt faglige målepunkter kan der potentielt ske en forenkling i et sådant omfang, at der kan sættes spørgsmålstegn ved nytteværdien og validiteten af de informationer, som evalueringen tilvejebringer.

Performanceevalueringens formål på skoleområdet er at skabe viden om skolepræstationer på alle offentlige skoler (og så vidt muligt også på fri- eller privatskoler). Det sker med henblik på 1) at kunne vurdere, om skolepræstationerne på de enkelte skoler, fag og klassetrin står mål med de krav og forventninger, der er formuleret i folkeskoleloven, og 2) at kunne sammenligne resultaterne og på den baggrund kunne udpege stærke og svage skoler/kommuner. Performanceevaluering i form af nationale test lanceres som et styringsredskab for primært skoleforvaltninger og skoleledelse, men også som et styringsredskab for Undervisningsministeriet.

Den er dermed en vigtig informationskilde for politiske beslutningsprocesser. Performanceevaluering har derfor i vores kontekst både et eksternt og internt formål – set fra skolens perspektiv.

Formativ pædagogisk evaluering har til formål til at danne et vidensgrundlag for lærere, elever og forældre, der kan kvalificere undervisning og feedback til elever med den hensigt at effektivisere og understøtte den enkelte elevs læreprocesser. Men den formative pædagogiske evaluering har også til formål at give læreren mulighed for justering af egne didaktiske valg, således at undervisningens form og indhold matcher elevers læringsforudsætninger bedst muligt. Den formative evaluering (Miller, 2003) har som fænomen været en del af folkeskoleloven siden afskaffelse af karakterer til alle elever på alle klassetrin i 1975 og med indførelse af § 13 om den løbende evaluering.

Når formativ pædagogisk evaluering nævnes i forbindelse med indførelse af de nationale test, handler det om, at testresultater er tiltænkt en rolle i lærerens pædagogiske arbejde med eleven. Det er imidlertid problematisk at lade summative evalueringresultater danne grundlag for en formativ evaluering, da summarisk evaluering har større legitimitet. Hvis sameksistens af summative og formative evalueringer skal danne grundlag for elevers læring, er det derfor en forudsætning, at den formative evaluering er elevinvolverende, og at der er en fælles fortolkning af testresultater (Andreasen, 2011; Miller, 2003).

I forbindelse med de nationale test har der været ønske om at styrke skolernes evalueringsskiltur med en vifte af evalueringsmetoder. Det betød, at der i evalueringssamtaler på baggrund af testresultater skulle lægges vægt på formidling af testresultaterne til eleven og forældre, så resultaterne kunne anvendes i elevens og klassens videre arbejde med stoffet. Den formative evaluering er ikke dømmende, men karakteriseret ved dialog. Målet er, at lærer og elev forstår elevens læringsvanskeligheder, og at det kan befordre, at læreren bliver bedre i stand til at understøtte og styrke elevens faglige arbejde. Formativ pædagogisk evaluering har derfor primært et internt formål – set fra skolens perspektiv.

Der er således principielle forskelle på de to typer evaluering, som de nationale test har skullet varetage. Disse forskelle var ikke tydeligt markeret i den uddannelsespolitiske proces omkring indførelsen af testsystemet, som vi belyser på de følgende sider. Vi trækker her på den samfundsvidenskabelige tradition for policy-analyse, som interesserer sig for, hvordan politikker formes i samspillet mellem aktører, beslutningsarenaer og bredere magtforhold, og hvilke behov og interesser der imødekommes eller tilsidesættes i disse processer (Ball, 2017; Mitra, 2018).

Fra PISA-chok til nationale test

De nationale test blev indført i foråret 2006 som led i en revision af folkeskoleloven. Den vedtagne lovtekst havde undertitlen *"Styrket evaluering og anvendelse af nationale test som pædagogisk redskab samt obligatoriske prøver m.v."*. Dermed blev der samtidig indført andre evalueringstiltag: Præcisering af, at elevernes udbytte af undervisningen skulle vurderes i forhold til trin- og slutmål for de enkelte fag i folkeskolen, obligatoriske afgangsprøver og udvidelse af kredsen af prøvefag. En yderligere lovændring et par måneder senere indførte kommunale kvalitetsrapporter og regler for, hvordan kommunerne skulle bruge resultaterne fra de nationale test til at kontrollere kvaliteten på de enkelte folkeskoler.

Denne vifte af evalueringstiltag blev vedtaget efter nogle års intens offentlig og politisk debat om kvaliteten og udbyttet af folkeskolens undervisning. Debatten var især blevet affødt af resultater fra OECD's komparative undersøgelser af skoleelevers færdigheder, de såkaldte PISA-undersøgelser, som Danmark indgik i (Jensen, 2011). Den første runde af PISA-resultater, som fremkom i 2000, viste, at de danske elevers færdigheder i læsning og matematik var dårligere end færdighederne i en række andre europæiske lande. Og resultaterne fra næste undersøgelsesrunde i 2004 var ikke bedre. Ligesom en række andre lande oplevede Danmark et "PISA-chok", hvor tilliden til det nationale skolesystem blev rystet. Inden for en snævrere kreds af fagfolk og skolepolitiske aktører var debatten allerede startet i 1994, hvor en international undersøgelse af elevers læsefærdigheder, gennemført af konsortiet IEA (International Evaluation of Achievement), også havde peget på,

at danske skoleelevers læsefærdigheder langt fra lå i top.

En konsekvens af PISA-resultaterne var, at Undervisningsministeriet inviterede OECD til at gennemføre en evaluering af den danske folkeskole. Evalueringen blev gennemført af et internationalt panel ledet af den engelske professor Peter Mortimore. Panelet tog blandt andet udgangspunkt i en analyse af de danske PISA-resultater, og dets anbefalinger lagde stærkt vægt på at styrke evalueringsskulturen i skolen.

Begrebet evalueringsskulturen omfattede en vifte af tiltag på forskellige niveauer og hos forskellige aktører i skolen (Uddannelsesstyrelsen, 2004; Dahler-Larsen, 2006). Lærere og skoler skulle evaluere deres egen indsats, den løbende evaluering af eleverne skulle styrkes og præciseres, feedback skulle baseres på objektive kriterier, og der skulle være en konstant, landsdækkende overvågning af landets skoler – udført af Undervisningsministeriet. Rapporten diskuterede nogle forskellige muligheder for denne nationale overvågning, blandt andet et system med obligatoriske test i 3., 6. og 9. klasse, som var blevet indført i den canadiske provins Alberta. Rapporten pegede også på, at det ville være godt med et nationalt overvågningsprogram ved siden af de internationale undersøgelser, idet "(...) den danske undersøgelse og mulighederne for hurtigt at gribe ind ikke ville være underkastet tidsrammen i det internationale program" (Uddannelsesstyrelsen, 2004, s. 135). Rapporten nævnte ikke direkte begrebet nationale test, men dens vægt på national overvågning pegede i den retning.

Legitimering af testning

PISA-resultaterne og debatten om dem førte også til, at de politiske partier overvejede og i nogle tilfælde reformulerede deres holdninger til evalueringer, test og central styring af folkeskolen. Debatten er blevet undersøgt grundigt af Line Renate Hansen (tidligere Line Renate Gustafsson), som især peger på markant ændrede holdninger hos Socialdemokratiet, men også på ændringer hos andre partier (Gustafsson, 2011; Hansen, 2015). I årene omkring år 2000, hvor de første PISA-resultater fremkom, lå regeringsmagten hos Socialdemokratiet i koalition med Det Radikale Venstre. De

skolepolitiske ideer hos begge partier var præget af reformpædagogik og afstandtagen til eksaminer og test: I forbindelse med reform af folkeskolen i 1993 havde partierne seriøst diskuteret afskaffelse af karakterer. Den radikale Magrethe Vestager, som var undervisningsminister, fulgte dog en mere styringsorienteret linje end tidligere radikale undervisningsministre, og hun fik blandt andet etableret Danmarks Evalueringsinstitut med henblik på en landsdækkende kvalitetsovervågning.

Ved folketingsvalget i 2001 gik Socialdemokratiet tilbage, og Venstre dannede regering med Anders Fogh Rasmussen som statsminister og Ulla Tørnæs som undervisningsminister. Den nye regering lagde stærk vægt på at styrke fagligheden og elevernes færdigheder. Det kunne blandt andet ske gennem nationale test, som de borgerlige partier tidligere havde foreslået; men den nye regering måtte have Socialdemokratiets opbakning på grund af traditionen for brede forlig på skoleområdet. Hansen (2015) peger på, at de borgerlige partier i de næste par år gennemførte en såkaldt delegitimering, hvor de gennem argumenter søgte at svække Socialdemokratiets skolepolitiske position. Et eksempel var statsministerens åbningstale i oktober 2002, hvor han gik til kamp mod den såkaldte "rundkredspædagogik" og hævdede, at den lod de svageste elever i stikken. Senere brugte regeringen de PISA-resultater, som blev offentliggjort i 2004, og hvor særligt børn af enlige forældre og tosprogede børn klarede sig dårligt, til at fremhæve, at den tidligere regerings politik ikke havde hjulpet elever med svag familiebaggrund til at klare sig bedre i skolen.

De borgerlige politikere fremførte således en kausal sammenhæng, hvor test blev forstået som et pædagogisk instrument, der kunne hjælpe de børn og unge, som havde mest behov for det. Tilsvarende argumenter blev fremført af andre aktører, og OECD-evalueringen pegede også på, at en bedre evalueringsskulturen kunne styrke den sociale lighed i skoleresultater. Eftersom social lighed er en grundlæggende værdi i Socialdemokratiets politik, havde partiets repræsentanter svært ved at håndtere dette angreb på deres legitimitet. Resultatet var en gradvis tilbagetrækning fra partiets tidligere position og en accept af, at nationale test kunne være nødvendige.

Det udtalte den daværende politiske ordfører, Frank Jensen, i januar 2005 (Gustafsson 2011, s. 197). Og selv om der var uenigheder internt i Socialdemokratiet, blev accepten af nationale test som led i en evalueringskultur efterhånden dominerende. Modstanden blev i stedet rettet mod potentiel offentliggørelse af testresultater og brugen af dem til at rangordne skoler. Dermed var vejen banet for forlig om de ændringer af folkeskoleloven, som indførte evalueringskultur og nationale test. Lovændringerne blev vedtaget af regeringen og Socialdemokratiet, men uden deltagelse af Det Radikale Venstre og venstrefløjsparterne.

Disse politiske positioner kan tydeligt iagttages i Folketingets behandling af forslaget om ændring af folkeskoleloven. Ved førstebehandlingen den 17. januar 2006 udtalte den socialdemokratiske ordfører Kim Mortensen, at uddannelsessystemet skal nedbryde den sociale arv, og at:

Det kræver mere fokus på og mere målrettethed over for den enkelte elev, og det kræver yderligere fokus på faglighed og indlæring specielt for den alt for store gruppe af elever, der forlader folkeskolen uden tilstrækkelige færdigheder (...). For Socialdemokratiet har det været afgørende for vores stillingtagen til indførelse af nationale test, at de for det første er et pædagogisk redskab og for det andet ikke skal stå alene, men være en del af en styrket evalueringskultur, der ikke kun måler det faglige indhold, men bredt måler alle aspekter af undervisningen.

Det er slående, at stort set alle indlæg i debatten opfatter de nationale test som et pædagogisk redskab, som lærerne kan og skal tage i brug. Ordføreren for Venstre siger for eksempel:

Testene skal udgøre et værktøj, der kan medvirke til, at læreren får et godt overblik over den enkelte elevs faglige niveau, og derigennem bidrager de til grundlaget for det faglige fremskridt for den enkelte elev ved at målrette undervisningen til netop denne elevs behov og forudsætninger.

Den eneste taler, som antyder et andet perspektiv, er ordføreren for Det Konservative Folkeparti, som beklager, at *"vi ikke har en stærk evalueringskultur i vores folkeskole, og det betyder, at vi har svært ved at følge og måle resultaterne af skolernes undervisning"*.

Efter førstebehandlingen i Folketinget blev lovforslaget sendt til behandling i Uddannelsesudvalget, og et flertal i udvalget (Venstre, Socialdemokratiet, Dansk Folkeparti og Konservative) afgav den 31. marts 2006 en betænkning, som indledte med følgende formulering:

Formålet med testene er at skabe et pædagogisk evalueringsredskab, der kan bidrage til en nuanceret vurdering af den enkelte elevs udbytte af forskellige undervisningsemner inden for udvalgte fag på bestemte klassetrin. Testene tænkes anvendt sammen med andre evalueringsredskaber som led i den løbende evaluering.

I OECD-panelets rapport var et nationalt overvågningsprogram og løbende pædagogisk evaluering af eleverne to forskellige elementer i en evalueringskultur. Det er rimeligt konsistent med en evalueringsfaglig forståelse. Forslaget om overvågning lægger op til en performancemåling på nationalt og lokalt plan, mens forslagene om løbende evaluering og feedback lægger op til formativ pædagogisk evaluering. Men i den politiske debat om lovforslaget blev de kortsluttet med hinanden. De nationale test blev præsenteret som et pædagogisk redskab, som lærerne skulle bruge den enkelte elev. Set ud fra den evalueringsfaglige forståelse må man være skeptisk over for at tildele et overvågningsprogram denne rolle, og den senere udvikling har bekræftet denne skepsis.

Udvikling og implementering

I Folketingets beslutning om de nationale test blev det understreget, at resultaterne ikke skulle være offentligt tilgængelige i en form, der kunne bruges til at rangliste skoler. Lærere skulle have adgang til deres elevers resultater, mens de samlede resultater på landsplan skulle offentliggøres, og kommunalbestyrelser og skoleledere skulle have oplyst de enkelte skolers testresultater.

Udviklingen og iværksættelsen af de nationale test var en stor og kompleks opgave. Der var tale om ti obligatoriske test i læsning, matematik, engelsk, fysik/kemi, geografi og biologi på tværs af alle klassetrin fra og med 2. klasse. Det blev besluttet, at testene skulle gennemføres via nettet, og at de skulle være adaptive, således at testen tilpassede sig niveauet hos den enkelte elev. Samtidig med det tekniske design skulle der udvikles et batteri af opgaver til de enkelte test. Udviklingen af testsystemet blev udbudt i licitation, men først efter to udbudsrunder blev der indgået kontrakt med en privat leverandør. Det viste sig, at leverandøren havde svært ved at løfte opgaven, især udviklingen af testopgaver – med den fornødne kvalitet og spredning – og indførelsen af systemet blev stærkt forsinket. Der var stor opmærksomhed og debat om sagen, og Rigsrevisionen gennemførte en undersøgelse af processen (Rigsrevisionen, 2010). Først i 2010 blev de ti obligatoriske nationale test gennemført første gang.

I forbindelse med folkeskolereformen i 2013 blev det besluttet, at de nationale test skulle anvendes til at vurdere udviklingen frem mod reformens overordnede målsætninger. To af de tre målsætninger blev operationaliseret med henvisning til de nationale test. Målet om, at "folkeskolen skal udfordre alle elever, så de bliver så dygtige, de kan", blev konkretiseret til, at mindst 80 procent af eleverne skulle være gode til at læse og regne i de nationale test. Målet om, at "Folkeskolen skal mindske betydningen af social baggrund i forhold til faglige resultater", blev konkretiseret til, at "andelen af elever med dårlige resultater i de nationale test for læsning og matematik uanset social baggrund skal reduceres år for år" (Regeringen, 2013, bilag 1). Hermed blev testsystemets funktion som performancemåling cementeret. I forlængelse af reformen foregik der en videreudvikling af de nationale test, blandt andet blev der i 2017 indført nye test i matematik og engelsk.

I det følgende belyser vi, hvordan de nationale test i praksis har fungeret som henholdsvis performancemåling og formativ pædagogisk evaluering. Som grundlag for at vurdere dette har vi anvendt en del foreliggende empirisk forskning, som er gennemført i løbet af det tiår, hvor de nationale test var i funktion for alle folkeskoler.

Der foreligger kvalitativ forskning, som på grundlag af interviews og observationer især har set på testsituationerne – og på elevers og læreres oplevelse og håndtering af dem (Andreasen, 2011; Andreasen og Kousholt, 2015; Holm, 2015; Kousholt, 2015). Og der foreligger spørgeskemaundersøgelser, som har fokuseret på læreres og elevers erfaringer med og vurderinger af testene (Puck og Bundsgaard, 2016; Eriksen, Lindholst og Knudsen, 2020).

Vi inddrager denne forskning, men vi bruger især den nyeste undersøgelse af anvendelsen af de nationale test, som er udarbejdet i forbindelse med evalueringen af testene (Bjørnholt m.fl., 2020). Det er en omfattende undersøgelse, som har indhentet data fra både elever, lærere, skoleforvaltninger, skolebestyrelser og politikere. Den bygger på en alsidig kombination af metoder, herunder spørgeskemaundersøgelser, kvalitative interviews, fokusgrupper og videoobservation af testsituationer. Undersøgelsen giver efter vores opfattelse et solidt grundlag for at vurdere anvendelsen af de nationale test, men vi inddrager som nævnt også andre undersøgelser.

Nationale test som performancemåling

Selv om de politiske aktører bag indførelsen af de nationale test helt overvejende betonedede testenes formative pædagogiske funktion, blev testenes funktion som performancemåling dog også nævnt. Testsystemet skulle dokumentere udviklingen af danske skoleelevers faglige niveau på nationalt plan, og det skulle give kommuner og skoleledere grundlag for at følge elevernes resultater – og på det grundlag styrke kvaliteten af undervisningen. I forlængelse af skolereformen i 2013 blev de nationale test som nævnt også brugt til at monitorere opfyldelsen af folkeskolens overordnede mål. De nationale test skulle således stille ledelsesrelevant information til rådighed for stat, kommuner og skoler.

På det nationale niveau kan man skelne mellem den administrative og den politiske anvendelse. Administrativt bruges de nationale test på forskellige måder til at understøtte styring og ledelse.

Data fra de nationale test anvendes som led i Børne- og Undervisningsministeriets generelle overvågning af skoleområdet, og mere specifikt anvendes de i forbindelse med det nationale tilsyn med folkeskolens faglige kvalitet. Der er tale om et såkaldt risikobaseret tilsyn, hvor ministeriet foretager foreløbig screening af skolerne ud fra fem kvalitetsindikatorer og på det grundlag vurderer, om skoler har vedvarende udfordringer med kvaliteten. Resultater fra nationale test i dansk og matematik udgør den ene af disse indikatorer.

På grundlag af interviews med udvalgte politikere finder forskerne, at den direkte politiske anvendelse af data fra de nationale test begrænset. En af politikerne udtaler: "Der har aldrig været en debat i min tid, hvor man har brugt de nationale test eller den faglige progression i et argument" (Bjørnholt m.fl., 2020, s. 106). Eksempelvis forelægges data fra testene kun forligskredsen bag folkeskolereformen ved en årlig statusreddegørelse for folkeskolens udvikling og kun på et generelt niveau. Derimod spiller de nationale test en betydelig indirekte rolle i politikudvikling, idet de bidrager med data til forskning, som politikerne orienterer sig i og bruger. Det gælder blandt andet følgeforskningen til folkeskolereformen.

På kommunalt niveau kan man også skelne mellem administrativ og politisk anvendelse. En survey blandt kommunale forvaltningschefer viser, at de ser de nationale test som et nyttigt redskab til at følge udviklingen i det faglige niveau i kommunen og på skoler. De vurderer også, at testsystemet har styrket både det administrative og det politiske fokus i kommunen på skolernes faglige resultater. En af cheferne udtaler: "*Nationale test er enormt godt. Hvis ikke vi havde dem, havde vi andre tests og evalueringstværværktøjer. Jeg ville være ked af, hvis man bare sagde, at alle former for tests skal væk*" (Bjørnholt m.fl., 2020, s. 82).

Ifølge cheferne følger forvaltningen også løbende og systematisk op på, om skolerne opnår de forventede resultater i de nationale test. Over 65 procent vurderer, at forvaltningerne i temmelig høj grad eller i meget høj grad følger op (Bjørnholt m.fl., 2020, s. 91). En sammenligning med data fra den tidligere evaluering af de nationale test (Rambøll, 2013) viser, at opfølgning på testene sker

i langt større omfang end tidligere (Bjørnholt m.fl., 2020, s. 93). De nationale test er tilsyneladende blevet konsolideret som et integreret led i kommunernes administrative ledelse på skoleområdet. Blandt de kommunale politikere finder man dog forskellige opfattelser af og holdninger til testsystemet.

På skoleniveauet tegner undersøgelsen et blandet billede af opfattelsen og anvendelsen af de nationale test som ledelsesværktøj. Over halvdelen af skolelederne mener, at testresultaterne giver ledelsen anvendelig information om skolens faglige niveau, mens en langt mindre andel mener, at testene giver ny viden om elevernes faglige niveau (Bjørnholt m.fl., 2020, s. 61). En skoleleder udtaler:

"Hele ideen med nationale test er god – at man får et billede af, hvordan tingene ser ud i forhold til nationalt plan. Hvis man bruger det rigtigt i sin hverdag, så kan jeg se det som noget gavnligt. Jeg bokser med validitet, som jeg konstant hører fra mine lærere, for det er man i tvivl om. Og så stiller man spørgsmålstegn ved værdien af nationale tests."

(Bjørnholt m.fl., 2019, s. 68).

Skolelederne bliver også spurgt, hvordan og hvor meget de bruger resultater fra de nationale test som grundlag for handlinger. Det viser sig, at de bruges i begrænset omfang. På spørgsmål om, hvor meget de bruger nationale test i prioritering af nye indsatser, planlægning af kompetenceudvikling eller formulering af nye mål, er det kun omkring halvdelen af skolelederne, som svarer positivt (kategorierne i nogen grad, i høj grad og i meget høj grad), og i forhold til udarbejdelse af budget er det kun en fjerdedel (Bjørnholt m.fl., 2020, s. 78). I det hele taget tyder undersøgelsen på, at de nationale test udgør en mindre del af skoleledernes styrings- og ledelsesgrundlag. Lederne bruger mest testresultater til at vurdere skolens samlede, faglige progression og i mindre grad til at vurdere den enkelte elev eller klasse.

Selv om billedet på skoleniveauet er blandet, dokumenterer undersøgelsen, at ledere på de forskellige niveauer inden for skoleområdet opfatter de nationale test som et værdifuldt bidrag til vidensgrundlaget at opretholde og udvikle kvaliteten af folkeskolens undervisning. Det er især i de større enheder såsom kommuner og skoler, at testresultater – ud fra et forvaltningsperspektiv – ses som værdifulde og i et vist omfang kan påvirke handlinger. I ledernes forhold til de mindre enheder, klasserne og de enkelte elever tillægges testresultater mindre betydning. Undersøgelsen bekræfter, at de nationale test har kunnet bidrage til performancemåling af folkeskolen på nationalt og lokalt plan.

Nationale test som formativ pædagogisk evaluering

De nationale test blev som nævnt indført med en Lov om ændring af Folkeskoleloven, L101 2005-06. Den har titlen Styrket evaluering og anvendelse af nationale test som pædagogisk redskab samt obligatoriske prøver. Lovens titel signalerer klart, at testene skal anvendes som et led i den formative evaluering, og denne beskrives nærmere i folkeskolelovens § 13, hvor det hedder:

Som led i undervisningen skal der løbende foretages evaluering af elevernes udbytte heraf, herunder af elevens tilegnelse af kundskaber og færdigheder i fag og emner set i forhold til kompetencemål og opmærksomhedspunkter. Evalueringen skal danne grundlag for vejledning af den enkelte elev og for den videre planlægning og tilrettelæggelse af undervisningen, og for underretning af forældrene om elevens udbytte af undervisningen. Børne- og undervisningsministeren udarbejder test i udvalgte fag og på bestemte klassetrin, som skal anvendes som led i den løbende evaluering (Folkeskoleloven, § 13, stk. 2 og 3).

Den løbende evaluering er grundlæggende tiltænkt et formativt formål, idet evaluering af elevens læringsudbytte skal lede til konkret vejledning af den enkelte med det formål at understøtte læring. Spørgsmålet er så, hvordan lovtekstens krav er blevet indfriet, og hvilken rolle de nationale test har spillet i denne sammenhæng. I det

følgende belyser vi, hvordan lærere, elever og skoleledere vurderer værdien af nationale test som værktøj i den formative evaluering.

”I overensstemmelse med tidligere undersøgelser (Puck og Bundsgaard, 2016) peger VIVE’s undersøgelse på, at lærere generelt er skeptiske over for nationale test som pædagogisk redskab, og at de i mindre grad bruger test som afsæt for deres pædagogiske praksis.”

(Bjørnholt m.fl., 2020, s. 39).

Der er dog forskel på begrundelserne for denne skepsis. Én begrundelse bygger på kritik af testenes indhold, validitet og kvalitet, som afføder manglende tillid til, at testene bidrager til et bedre indblik i enkeltelevers eller klassers faglige niveau på de områder, hvor eleverne bliver testet. Nogle lærere fortæller dog, at testresultater stemmer med deres forventninger til eleverne. En anden begrundelse er, at der blandt lærere hersker usikkerhed om, hvordan og hvorvidt de skal bruge nationale test i deres pædagogiske arbejde, og at det ikke giver mening at teste børnene, når de pædagogiske formål med testene er uklare. Generelt vurderer lærere, at nationale test i langt højere grad er et styringsredskab end et pædagogisk redskab. Lærere oplever ikke, at indholdet i testene afspejler det, de underviser i, og derfor finder de det vanskeligt at anvende resultaterne i det videre arbejde med læring og undervisning.

”Og omkring halvdelen af lærerne mener, at der kun i mindre grad eller slet ikke findes retningslinjer for den pædagogiske brug af testene.”

(Bjørnholt m.fl., 2020, s. 40).

Lærerne peger også på forskellige sideeffekter ved nationale test. Én af disse optræder i forholdet mellem kolleger, når en klasse ikke klarer sig godt testene. I sådanne situationer kan lærere opleve at stå meget alene med ansvaret.

En lærer siger: ”Pludselig bliver det en personlig ting, og man kan jo ikke gøre for, hvad man ramler ind i forhold til classesammensætning. Hvis det er en tung klasse, så bliver det en byrde, hver gang man får nationale testresultater.”

(Bjørnholt m.fl., 2020, s. 52).

Lærerne føler også, at deres undervisningskompetencer bliver vurderet ud fra, om eleverne klarer sig godt. Derfor ses ”teaching to the test” som en sideeffekt eller konstitutiv virkning, som især synes at forekomme i kernefagene dansk og matematik. Et nævnt eksempel er, at dansklærere underviser i ordsprog og fremmedord, fordi dette indgår i de nationale test, også selv om flere dansklærere ikke mener, at ordsprog og fremmedord er et godt mål for elevernes ordforråd. Samtidig fremgår det af analyserne, at eleverne mange steder trænes i at gennemføre de nationale test. Således angiver 28 procent af skolelederne, at de i nogen, temmelig høj eller meget høj grad opfordrer lærerne til at bruge tid på opgaver, der vil gavne eleverne i nationale test (Bjørnholt m.fl., 2020, s. 53-54). Skolelederne begrundet det primært med, at eleverne skal blive mere trygge ved opgaverne, og ikke ud fra et ønske om, at eleverne skal opnå bedre resultater. Også tidligere undersøgelser peger på, at en betydelig del af lærerne tilrettelægger meget af undervisningen i perioden før de nationale test som en forberedelse til testene (Puck og Bundsgaard, 2016, s. 13).

Når lærere i højere grad anvender testresultater i matematik og dansk, skyldes det, at skoleledere ofte følger op på disse fag, og at der på skolerne er ansat faglige vejledere i dansk og matematik, som kan være behjælpelige med at tolke og gøre testresultaterne operative.

Undersøgelsen peger på, at nationale test i et vist omfang understøtter en tillidsbaseret dialog mellem lærere og skoleledere, men at dette arbejde sjældent resulterer i konkrete indsatser eller nye beslutninger (Bjørnholt m.fl., 2020, s. 61). I de tilfælde, hvor nationale test indgår som grundlag for nye indsatser, sker det i samspil med andre kilder til viden, og her spiller de faglige vejlederes vurderinger en stor rolle. I kraft af et bedre indblik i lærernes praksis kan vejlederne understøtte en mere

konkret og faglig tolkning af resultater fra nationale test som ledelsesinformation, der kan bruges til at styre og udvikle skolen.

VIVE’s undersøgelse peger på, at eleverne oplever testsituationen som udramatisk, og at de synes, at lærerne gør et stort stykke arbejde for at gøre testsituationen så god som muligt og understøtte elever, der går i stå. De små elever har dog vanskeligt ved at holde koncentrationen. Nogle elever oplever testsituationen som ubehagelig og frustrerende, mens andre oplever den som en dejlig afveksling fra den daglige undervisning. De fleste elever er dog hverken særlig positive eller negative over for testen (Bjørnholt m.fl., 2020, s. 27).

En generel kritik fra elever er, at testen er for lang, og at der er for mange spørgsmål. Undersøgelsen viser da også, at koncentrationen falder i løbet af testen – mest for mindre elever, men altså også for de større. I tiden op til en national test oplever nogle elever nervøsitet eller en forventning om, at testen bliver kedelig, når de skal sidde stille og ikke må snakke med andre. Nogle elever henviser til deres erfaringer fra tidligere test:

”Jeg havde tænkt, at den ikke skulle blive så langtrukket som en, jeg prøvede nede i indskolingen, hvor jeg sad i, jeg ved ikke hvor lang tid, hvor den bare blev ved med at spytte spørgsmål ud. Den ville bare ikke stoppe.”

(elev på mellemtrinnet) (Bjørnholt m.fl., 2020, s. 27).

Elever er generelt optagede af at klare sig godt i testene, for de er klar over, at deres forældre får resultaterne. Beskrivelser af nervøsitet handler oftest om, at elever er usikre på resultaterne eller bange for at lave fejl. Nogle elever beskriver fysiske symptomer, for eksempel hjertebanken, ondt i maven, rysten eller gråd.

”Jeg synes, mit hjerte begyndte at banke i går. Der tænkte jeg: ’Åh nej, det er i morgen, vi skal have national test’. Så begyndte jeg at hive mig lidt op.” (indskolingselev) (Bjørnholt m.fl., 2020, s. 27).

En nyere undersøgelse af opfattelsen af nationale test blandt folkeskoleelever i 8. og 9. klasse (Erichsen, Lindholst og Knudsen, 2020) peger også på, at store dele af eleverne oplever at være nervøse i tiden op til en national test – og at være kede af det, efter at de har modtaget testresultater.

Alt i alt viser både VIVE's undersøgelse og anden forskning, at værdien af nationale test som et led i den pædagogiske evaluering er særdeles tvivlsom. Lærere og skoleledere er usikre på testenes validitet og finder resultaterne komplekse og svære at tolke. De fremhæver, at nationale test aldrig kan stå alene, og de peger på en række andre evalueringsredskaber, som understøtter læring og undervisning på et "oplyst grundlag". I samspillet mellem lærere, klasser og skoleledere kan de nationale test fremme konstitutive virkninger, hvilket også er vist i kvalitative undersøgelser (Andreasen og Kousholt, 2015; Kousholt, 2015). De fleste elever oplever testsituationen som udramatisk, men hos nogle medfører den usikkerhed og nervøsitet.

Konklusion

Den omfattende faglige diskussion, som har været ført op til og i sammenhæng med evalueringen af de nationale test, har især fokuseret på spørgsmålet om testenes validitet og graderne af usikkerhed i resultaterne. Det er væsentlige spørgsmål, men i denne artikel har vi ikke taget dem op, fordi vi har fokuseret på spørgsmålet om testsystemets formål og dets anvendelighed i forhold til disse formål.

De vedtagne formål med de danske nationale test blev i høj grad præget af de uddannelsespolitiske dagsordener og konstellationer frem til vedtagelsen i 2006. Perioden omkring årtusindeskiftet var både internationalt og i Danmark præget af et øget fokus på de enkelte landes institutionelle konkurrenceevne, signaleret gennem begrebet "konkurrencestat" (Pedersen, 2011). På skoleområdet kom OECD's PISA-undersøgelser til at udgøre en arena for landes vurdering af deres styrke eller svaghed, og i Danmark rejste især liberale og konservative kræfter krav om foranstaltninger, som kunne forbedre skoleelevers faglige præstationer. Én af foreslåede foranstaltninger var nationale test; men det stødte på modstand i reformpædagogiske kredse, som blandt andet stod stærkt i Socialdemokratiet og Radikale Venstre. De natio-

nale test blev vedtaget, men vedtagelsen måtte begrundes reformpædagogisk; som en styrkelse af formativ evaluering og potentiel forøgelse af den sociale lighed. Det prægede formuleringerne i den ændrede folkeskolelov.

De nationale test er imidlertid primært anvendelige som performancemåling. Det er de også designet til, men i lovgivning og offentlig debat er de som nævnt først og fremmest blevet tildelt pædagogiske funktioner. Det har været en hovedkilde til deres legitimitet. I forhold til formativ pædagogisk evaluering viser forskningen, at testenes fokus er for snævert, og at resultaterne er svære at tolke og håndtere for de lærere, som skal bruge resultaterne til styrke elevernes præstationer. Som performancemåling og bidrag til kvalitetssikring har de nationale test haft en værdi, dels på nationalt plan, men især på kommunalt niveau. I forhold til disse funktioner er et system, hvor alle elever fra 2. klasse i alle skoler skal testes i en række fag, dog en krævende og dyr foranstaltning, og meget taler for, at det nødvendige vidensgrundlag for kvalitetssikring ville kunne frembringes på andre måder. Danmark har jo grundige, individbaserede uddannelsesregistre, som kan nyttiggøres i sammenhæng med andre informationskilder og stikprøveundersøgelser.

Nytten af de nationale test skal også vurderes i sammenhæng med, at de i et vist omfang belaster skolernes arbejde og har u hensigtsmæssige konstitutive virkninger. Både VIVE's undersøgelse af testsystemets anvendelse og anden forskning viser, at systemet fremmer "teaching to the test" som en konstitutiv virkning, og at testresultater som konkurrenceparameter kan påvirke lærernes arbejdsmiljø. Desuden styrker systemet alt andet lige det præstationspres, som mange børn og unge oplever, og som går ud over deres trivsel.

Disse vanskeligheder og ulemper ved systemet med nationale test fremstod egentlig tydeligt allerede nogle år efter systemets indførelse, men en række embedsmænd, politikere og forskere har ikke villet opgive den store investering, som testsystemet har været. Derfor ser vi det som positivt, at regeringen har besluttet at indskrænke anvendelsen af de nationale test. Om den valgte model er fornuftig, er for tidligt at sige.



Litteratur

- Andersen, M., Wahlgren, B. & Wandel, J. (2017). *Evaluering – af læring, undervisning og uddannelse*. København: Hans Reitzels Forlag.
- Andreasen, K.E. (2011). *Summative test i formativ praksis*. I Andreasen, K.E., Friche, N. & Rasmussen, A. (red.). *Målt og vejlet*. Uddannelsesforskning om evaluering. Aalborg: Aalborg Universitetsforlag.
- Andreasen, K.E. & Kousholt, K. (2015). *De nationale test som ny praksis i den danske folkeskole. Betydninger i klassens sociale fællesskab*. I Andreasen, K.E., Buchardt, M., Rasmussen, A. & Ydesen, C. (red.). *Test og prøvelser. Oprindelse, udvikling, aktualitet*. Aalborg: Aalborg Universitetsforlag.
- Ball, S.J. (2017). *The Education Debate*. Bristol: Policy Press.
- Bjørnholt, B., Flarup, L.H., Hindsholm, M.F., Brændgaard, N.W., Andersen, C.M. ... & Jensen, S.V. (2020). *Anvendelsen af de nationale test. Delrapport 5: Evalueringen af de nationale test*. København: VIVE.
- Bundsgaard, J. (2018). *Pædagogisk brug af test*. *Sakprosa*, 10(2), 1-40.
- Bundsgaard, J. & Kreiner, S. (2019). *Undersøgelse af de nationale tests måleegenskaber*. Aarhus: DPU, Aarhus Universitet.
- BUVM (Børne- og Undervisningsministeriet) (2020): *Rådgivningsgruppen for evaluering af de nationale test. Anbefalinger, januar 2020*. København: Børne- og Undervisningsministeriet.
- Dahler-Larsen, P. (2006). *Evalueringskultur – et begreb bliver til*. Odense: Syddansk Universitetsforlag
- Eriksen, T.B., Lindholst, A.C. & Knudsen, S.V. (2020). *Resultater fra en spørgeskemaundersøgelse om folkeskoleelevers opfattelse/oplevelse af de nationale test*. Aalborg: Aalborg Universitet.
- Gustafsson, L.R. (2011). *What did You Learn in School Today? How Ideas Mattered for Policy Changes in Danish and Swedish Schools 1990-2011*. Ph.d.-afhandling. Aarhus: Politica.
- Hansen, L.R. (2015). *Hvordan blev de nationale test indført? Ideer som politiske våben*. *CEPRA-sriben* 18, 6-12.
- Holm, L. (2015). *Læreropfattelser og praksis i relation til det nationale test i dansk/læsning*. *CEPRA-Sriben*, 18, 58-63.
- Jensen, T.P. (2011). *PISA – en makroevaluering af uddannelse*. I Andreasen, K., Friche, N. & Rasmussen, A. (red.). *Målt og vejlet*. Uddannelsesforskning om evaluering. Aalborg: Aalborg Universitetsforlag.
- Kousholt, K. (2015). *Børns gætterier ved nationale test*. *CEPRA-Sriben*, 18, 46-57.
- Krogstrup, H.K. (2016). *Evalueringsmodeller*. København: Hans Reitzels Forlag.
- Miller, T. (2003): *Evaluering og legitimitet*. I Gaarsmand, B. & Jacobsen, B. (red.). *Folkeskolen – ekstern tilpasning og intern organisering*. Værløse: Billesø & Baltzer.
- Mitra, D.L. (2018). *Educational Change and the Political Process*. New York, NY: Routledge.
- Pedersen, O.K. (2011). *Konkurrencestaten*. København: Hans Reitzels Forlag.
- Puck, M.R. & Bundsgaard, J. (2016). *Nationale test: Danske lærere og skolelederes brug, holdninger og viden*. Aarhus: DPU, Aarhus Universitet & Center for Anvendt Skoleforskning, University College Lillebælt.
- Rambøll (2013). *Evaluering af de nationale test i folkeskolen. Rapport*. København: Rambøll.
- Rasmussen, P. (2019). *De politiske partiers uddannelsespolitiske programmer ved valget til Folketinget d. 5. juni 2019*. Arbejdsnotat. Aalborg: Aalborg Universitet: Institut for Kultur og Læring.
- Regeringen (2013). *Aftale mellem regeringen, Venstre og Dansk Folkeparti om fagligt løft af folkeskolen, 7. juni 2013*. København: Regeringen.
- Rigsrevisionen (2010). *Beretning til Statsrevisorerne om udviklingen af de nationale test*. København: Rigsrevisionen.
- Uddannelsesstyrelsen (2004). *OECD-rapport om grundskolen i Danmark – 2004*. København: Uddannelsesstyrelsen, Undervisningsministeriet.