

Til
**Undervisningsministeriet
(Kvalitets- og Tilsynsstyrelsen)**

Dokumenttype
Rapport

Dato
September 2013

EVALUERING AF DE NATIONALE TEST I FOLKESKOLEN **RAPPORT**



INDHOLD

1.	Resumé	1
2.	Indledning	3
2.1	Om de nationale test	3
2.2	Baggrund	4
2.3	Om evalueringen	4
2.3.1	Formål	5
2.3.2	Metode og analysestrategi	5
2.4	Læsevejledning	9
3.	Testenes effekt	10
3.1	Er elevernes faglige niveau styrket efter indførelse af de nationale test?	10
3.1.1	Hvordan kan de nationale test tænkes at have effekt?	11
3.1.2	Hvordan undersøges effekten af de nationale test?	12
3.1.3	Hvad viser analyserne af effekten af de nationale test?	14
3.2	Er skolernes evalueringskultur styrket efter indførelse af de nationale test?	18
3.2.1	Er skolernes evalueringskultur blevet styrket?	18
3.2.2	Hvad er årsagen til den styrkede evalueringskultur?	21
3.2.3	Har de nationale test medvirket til en styrkelse af skolernes evalueringskultur?	21
3.3	Er der sammenhæng mellem skolernes evalueringskultur og elevernes faglige resultater?	30
3.3.1	Fremgangsmåde for dannelse af mål for skolernes evalueringskultur	31
3.3.2	Sammenhæng mellem evalueringskultur og faglige resultater	33
3.4	Delkonklusion	36
4.	De nationale test som pædagogisk værktøj	38
4.1	Lærernes anvendelse af testresultaterne til vurdering af eleverne	39
4.2	Lærernes anvendelse af testresultaterne til planlægning af undervisning	42
4.3	De nationale tests indflydelse på lærernes prioritering af undervisningens indhold	51
4.4	Delkonklusion	55
5.	Testenes betydning for elever og skole-hjem-samarbejde	57
5.1	Eleverne	57
5.1.1	Lærernes tilbagemelding til eleverne og opfølgning på resultaterne	58
5.1.2	De nationale tests betydning for elevernes forståelse af egne styrker og svagheder	60
5.1.3	De nationale tests betydning for elevernes (motivation for) læring	60
5.1.4	Elevernes oplevelse af testsituationen	61

5.2	Skole-hjem-samarbejde	63
5.2.1	Inddragelse af hjemmet i dialog om læringsmål	67
5.2.2	Testresultaternes værdi for forældre i skole-hjem-samarbejdet	67
5.3	Delkonklusion	68
6.	De nationale test som dialog- og styringsredskab	70
6.1	Anvendelse af de nationale test som værktøj for skolelederens kvalitetsarbejde	70
6.1.1	Skolernes evalueringspraksis	70
6.1.2	Skoleledelsens anvendelse af de nationale test	71
6.1.3	Brug af de nationale test som dialog- og styringsredskab	77
6.2	De nationale test som værktøj for kommuners kvalitetsarbejde	78
6.2.1	Kommuners generelle kvalitetsarbejde på skoleområdet	79
6.2.2	Kommunernes anvendelse af de nationale test	80
6.2.3	Samarbejde og dialog mellem forvaltning og skoler	87
6.3	Delkonklusion	87
7.	Testenes indhold	90
7.1	Testenes indhold, udformning og kvalitet	90
7.1.1	Fag og profilområder	90
7.1.2	Trinmål	92
7.1.3	Opgaveemner	92
7.1.4	Antal opgaver	93
7.1.5	Opgavetyper	93
7.1.6	Testopgavernes kvalitet	94
7.1.7	Udvikling af opgaver	95
7.2	Fordele og ulemper ved det adaptive princip	95
7.2.1	Kort beskrivelse af det adaptive princip	96
7.2.2	Teknisk om det adaptive princip	96
7.2.3	Økonomisk om det adaptive princip	97
7.3	Delkonklusion	98
8.	Testenes udformning	100
8.1	Den praktiske tilrettelæggelse af testene på skolerne, herunder selve gennemførelsen	100
8.2	Det adaptive princip ud fra et pædagogisk perspektiv	104
8.3	Tilbage melding til lærerne	107
8.4	Delkonklusion	112
9.	Konklusioner	114
9.1	Tematiske hovedkonklusioner	114
9.1.1	Testenes effekt	114
9.1.2	Testenes anvendelse	115
9.1.3	Testenes indhold	117
9.1.4	Testenes udformning	118
9.2	Tværgående hovedkonklusion	118
9.3	Perspektivering	119

Bilag:

Bilag 1: Metodebilag

Bilag 2: Model for evalueringskultur

Bilag 3: Detaljeret redegørelse for registeranalyser

Bilag 4: Ekspertvurdering 2 (det adaptive princip)

Bilag 5: Fagspecifikke opmærksomhedspunkter fra ekspertvurdering 1

1. RESUMÉ

Rambøll Management Consulting (herefter Rambøll) har i perioden november 2012 til september 2013 gennemført en evaluering af de nationale test. Evalueringen er udarbejdet på opdrag af Kvalitets- og Tilsynsstyrelsen, Undervisningsministeriet. Evalueringen er gennemført i samarbejde med en række underleverandører. Professor Helena Skyt Nielsen og lektor Simon Calmar Andersen, Aarhus Universitet, har været ansvarlige for at gennemføre en række registeranalyser af effekten af de nationale test. Professor Peter Allerup, Aarhus Universitet, har været ansvarlig for at foretage en vurdering af fordele og ulemper ved det adaptive princip. Endelig har lektor Carsten Bendixen og chefkonsulent Mikael Axelsen, Professionshøjskolen UCC, været sparringspartnere for Rambøll i forhold til udvikling af en model for evalueringskultur.

De nationale test er introduceret som et værktøj til at styrke den løbende evaluering i folkeskolen. Eleverne skal i løbet af deres skoletid gennemføre 10 obligatoriske nationale test inden for forskellige fag. Testene er it-baserede, selvscorende og adaptive. Sidstnævnte betyder, at testene tilpasser sig den enkelte elevs niveau undervejs i testforløbet. De nationale test blev lanceret i fuldt omfang i 2010.

Formålet med evalueringen er at vurdere, om de nationale test som pædagogisk værktøj har styrket skolernes evalueringskultur og elevernes faglige niveau. Evalueringen skal belyse positive såvel som negative effekter af indførelsen af de nationale test i forhold til formålet med testene.

Rambøll har i nærværende evaluering kortlagt, analyseret og vurderet de nationale test i forhold til følgende temaer: Effekt, anvendelse, indhold og udformning. Data er indsamlet ultimo 2012 og i første halvdel af 2013, og evalueringen er således udtryk for status for testenes implementering og effekt på dette tidspunkt. Evalueringen har haft sit hovedfokus på at vurdere, hvordan testene anvendes i praksis, samt eventuelle resultater og effekter heraf.

Metodisk er der anvendt et evalueringsdesign, der kombinerer effektevaluering og virkningsevaluering. Herved opnås viden om resultater og effekter, og tilsvarende en afdækning af *hvordan* og *hvorfor* eventuelle effekter er skabt. Datagrundlaget for evalueringens observationer, vurderinger og konklusioner er omfattende. De anvendte datakilder omfatter desk research, registerdata, spørgeskemaundersøgelser til lærere, skoleledere og forvaltninger, casestudier, inddragelse af eksperter samt et internationalt litteraturstudie om det adaptive princip. Datakvaliteten vurderes, samlet set, at være særdeles god.

I forhold til evalueringens første tema, **testenes effekt**, indikerer analysen, at indførelsen af de nationale test har haft en positiv effekt på elevernes faglige resultater. Evaluator har ved hjælp af registerdata gennemført analyser på elev-, skole- og landsniveau, og data indikerer, at de nationale test har en positiv effekt på elevernes faglige niveau. Resultaterne tyder endvidere på, at denne effekt ikke alene skyldes *teaching to the test*, eller at eleverne blot er blevet bedre til at gennemføre testene, men at de faktisk er blevet dygtigere til det fag, de er blevet testet i. Det har dog alene været muligt at se på effekten af et midlertidigt teknisk nedbrud i dansk, læsning, og det vil – når data for testresultater fra 2013 bliver tilgængelige – være muligt at undersøge, om den samme effekt ses i matematik. Evalueringen konkluderer videre, at skolernes evalueringskultur ser ud til at være styrket i de senere år, samt at der er flere indikationer på, at de nationale test har medvirket til at styrke skolernes evalueringskultur. Det er dog samtidig evaluators vurdering (som det anføres i lovbemærkningerne), at det ikke er de nationale test alene, der skaber en stærkere evalueringskultur. Det er derimod helheden i anvendelsen af nationale test i samspil med en række evalueringsværktøjer, samt en stærk ledelse, der understøtter en stærk evalueringskultur.

Evalueringsstemaet **testenes anvendelse** omfatter en analyse af en række aktørers anvendelse, oplevelse og vurdering af de nationale test. Selv om der er flere forskelle mellem læreres og skolelederes vurdering af de nationale test (herunder betydning af testene), er der flere fællesnævner mellem lærernes og skoleledernes *anvendelse* af de nationale test. Den overordnede konklusion er, at størstedelen af lærerne og skolelederne anvender de nationale test i henholdsvis en pædagogisk og ledelsesmæssig sammenhæng. Det er endvidere en fællesnævner, at både lærere

og ledere i overvejende grad bruger testresultater *summativt*, dvs. til at skabe overblik og monitorere det faglige niveau blandt elever, klasser og skolen som helhed. Den fremadrettede (formative) anvendelse, som var tiltænkt i lovgivningen, sker (indtil videre) i langt mindre omfang. Derudover er det en hovedkonklusion, at testresultater sjældent står alene som redskab i en evalueringsskulturs kultur. Evalueringen viser videre, at kommunerne anser de nationale test som en væsentlig indikator til at vurdere skolernes faglige resultater, men de nationale test er blot ét ud af flere værktøjer, og mange kommuner delegerer typisk ansvaret for opfølgning på testresultater til skolelederne. Endelig konkluderer evalueringen, at de nationale test i nogen grad har styrket skole-hjem-samarbejdet, samt at eleverne overordnet set er positivt indstillede over testene. En væsentlig konklusion er dog, at individuel tilbagemelding til og dialog med eleven om testresultater kun sker i begrænset omfang.

Hvad angår **testenes indhold**, konkluderer evalueringen, at valg af testfag og profilområder samlet set synes hensigtsmæssige. Tilsvarende afspejler testene i hovedtræk testegnede trinmål. Det er endvidere en hovedkonklusion, at der inden for flere testfag er mangel på svære opgaver i opgavebanken. Evalueringen peger på flere opmærksomhedspunkter i forhold til testopgaverne, såvel fagfaglige som relateret til layout, grafik samt brug af illustrationer. Det konkluderes videre på baggrund af evalueringens litteraturstudie, at det adaptive princip sikrer, at de nationale test fremskaffer sikker viden om elevens faglige niveau med høj præcision og med færrest mulige opgaver. Samtidig konstateres det, at de adaptive test er betydeligt mere ressourcekrævende end lineære test, og systemet har en lav agilitet og fleksibilitet.

Testenes udformning omfatter en analyse af rammerne for testafvikling ud fra flere perspektiver. Evalueringen viser, at afviklingen af de nationale test (herunder den praktiske forberedelse) fungerer på tilfredsstillende vis. Det medgåede tidsforbrug udgør generelt set heller ikke en udfordring. Evalueringen viser, at der er en klar tidsmæssig gevinst ved at gennemføre it-baserede test, når man sammenholder med papirbaserede test. Det skyldes, at de nationale test automatisk genererer et resultat, hvorfor lærerne sparer rettetid. Hvad angår det adaptive princip vurderet ud fra et pædagogisk perspektiv, er konklusionen, at størstedelen af lærerne ser dette som en fordel for eleverne og deres testafvikling. Dette understøttes af litteraturstudiet. Der peges dog samtidig på en række ulemper relateret til elevernes oplevelse af de nationale test. Endelig viser evalueringen, at testsystemets muligheder for at vise resultater (resultatvisningen) umiddelbart er brugervenlig, men der er et ønske om, at resultaterne kan vises på måder, der i endnu højere grad kan understøtte lærernes arbejde med at implementere resultaterne/arbejde med disse pædagogisk.

Evalueringens **samlede konklusion** er, at de nationale test har en positiv effekt på elevernes faglige niveau, samt at de nationale test i nogen grad har været en medvirkende årsag til en styrket evalueringsskulturs kultur, der har udviklet sig i folkeskolen i de senere år. I sammenhæng hermed konkluderer evaluator, at det ikke er de nationale test alene, der har resulteret i en stærkere evalueringsskulturs praksis. De nationale test har været en del af en bredere skoleudvikling, der i højere grad end tidligere er kendetegnet ved systematik, dokumentation af resultater og (i nogen grad) fastsættelse af mål. Evalueringen viser samtidig, at der er en række af de opstillede hypoteser, relateret til de nationale test, som evaluator ikke har kunnet bekræfte baseret på den indsamlede empiri. Med andre ord er der flere forbedringspunkter, der kan styrkes fremadrettet. I relation til såvel den pædagogiske som styringsmæssige anvendelse er den væsentligste udfordring at anvende testresultater *fremadrettet* (formativt) til tilrettelæggelse af undervisning, øget dialog og målretning af indsatsen.

Evaluator har som afslutning på evalueringen udarbejdet en **perspektivering** med en række opmærksomhedspunkter af relevans for den fremtidige udvikling af de nationale test.

2. INDLEDNING

Evalueringen af de nationale test er gennemført af Rambøll i samarbejde med en række underleverandører. Professor Helena Skyt Nielsen og lektor Simon Calmar Andersen, Aarhus Universitet, har været ansvarlige for at gennemføre en række registeranalyser af effekten af de nationale test. Professor Peter Allerup, Aarhus Universitet, har været ansvarlig for at foretage en vurdering af fordele og ulemper ved det adaptive princip. Endelig har lektor Carsten Bendixen og chefkonsulent Mikael Axelsen, Professionshøjskolen UCC, været sparringspartnere for Rambøll i forhold til udvikling af en model for evalueringskultur¹.

Evalueringen er gennemført på opdrag for Kvalitets- og Tilsynsstyrelsen, Undervisningsministeriet, i perioden fra november 2012 til september 2013.

Målgruppen for evalueringsrapporten er Kvalitets- og Tilsynsstyrelsen, der har fungeret som opdragsgiver for evaluatoren. Derudover er det politiske niveau en relevant målgruppe, da rapporten skal danne grundlag for en redegørelse til Folketinget ultimo 2013. Også kommuner og skoler med interesse i effekten og anvendelsen af de nationale test kan have glæde af at læse rapporten.

Evaluatoren vil gerne benytte lejligheden til at takke alle de kommuner og skoler, der har deltaget i nærværende evaluering. Vi har gennemført omfattende dataindsamling med stor involvering af de kommunale forvaltninger såvel som skoleledere, lærere, elever og forældre, og dataindsamlingen skulle for skolernes vedkommende passes ind i en i forvejen travl hverdag. Det er evaluatorens oplevelse, at vi er blevet mødt med stor interesse, samarbejdsvilje og hjælpsomhed, hvilket har bidraget til at skabe en god proces samt et solidt datagrundlag for evalueringen.

2.1 Om de nationale test

De nationale test er introduceret som et værktøj til at styrke den løbende evaluering af elevernes udbytte af undervisningen i folkeskolen. De nationale test skal fungere som et pædagogisk redskab for lærerne, og testresultaterne skal bidrage til den videre tilrettelæggelse af undervisningen². Derudover skal de nationale test fungere som værktøj i skolelederes og kommuners kvalitetsarbejde.

De nationale test tester elevernes kundskaber og færdigheder med henblik på at give lærerne et indblik i den enkelte elevs niveau. Eleverne skal i løbet af deres skoletid gennemføre 10 obligatoriske nationale test inden for forskellige fag. Tabellen nedenfor giver et overblik over de fag og klassetrin, hvor der skal gennemføres obligatoriske test.

Tabel 2.1: Oversigt over nationale test

Fag og klassetrin	2.	3.	4.	5.	6.	7.	8.
Dansk, læsning	x		x		x		x
Matematik		x			x		
Engelsk						x	
Geografi							x
Biologi							x
Fysik/kemi							x

De nationale test er kendetegnede ved at være:

- It-baserede – eleverne besvarer testene på en computer
- Selvscorende – lærerne skal ikke selv rette testene, men får leveret resultaterne
- Adaptive – testene tilpasser sig den enkelte elevs niveau undervejs i testforløbet.

¹ I den resterende del af rapporten bruges "evaluator" som samlet betegnelse for disse parter.

² www.uvm.dk.

De nationale test tester udvalgte områder i de enkelte fag (profilområder), og testenes funktion er således først og fremmest at vurdere elevernes faglige niveau inden for hvert af disse områder³.

I skoleåret 2011/2012 blev der gennemført ca. 529.000 obligatoriske test og ca. 420.000 frivillige test over hele landet. Der gennemføres således et stort antal frivillige test som supplement til de obligatoriske test. En frivillig test kan gennemføres på det klassetrin, den er målrettet mod, men også på klassetrinnet over og under. Det skal i den sammenhæng bemærkes, at de frie grundskoler også har mulighed for at gennemføre frivillige test.

2.2 Baggrund

I 2006 besluttede Folketinget at indføre nationale test i den danske folkeskole. Det skete ikke mindst på baggrund af en OECD-rapport fra 2004, der pegede på, at evalueringskulturen i den danske folkeskole var utilstrækkelig⁴. Rapportens konklusion blev understøttet af to samtidige evalueringsrapporter fra Danmarks Evalueringsinstitut (EVA)⁵, der også pegede på, at der savnedes dokumentation af folkeskolens arbejde og af elevernes udbytte af undervisningen.

OECD-rapporten konkluderede, at etableringen af en ny evalueringskultur sikkert er den enkeltforandring, der er mest vigtig at opnå, hvis andre initiativer skal kunne indføres, så de får virkning, og standarderne hæves. Svaret på denne evalueringssudfordring blev bl.a. indførelse af 10 nationale test under overskriften "Fremme af evalueringskulturen", en indførelse af kommunale kvalitetsrapporter og elevplaner samt en præcisering af, at den løbende evaluering af elevens udbytte af undervisningen skal ses i forhold til trin- og slutmål for de enkelte fag. De to sidstnævnte initiativer understreger det øgede fokus på styrkelse af elevernes faglige niveau, som også indførelsen af de nationale test var udtryk for. Som en del af projektet "Fremme af evalueringskulturen" blev der lanceret en evalueringsportal med en lang række evalueringsværktøjer og inspirationsmateriale, der skulle understøtte lærernes brug af de nationale test samt lærernes, skolernes og kommunernes evalueringspraksis generelt⁶.

Det fremgår af bemærkningerne til Forslag til Lov om ændring af lov om folkeskolen (Styrket evaluering og anvendelse af nationale test som pædagogisk redskab samt obligatoriske prøver mv.⁷), at det overordnede mål med indførelse af testene er en forbedring af det faglige niveau med henblik på at give alle elever mulighed for at opnå sikre kundskaber og færdigheder, som sætter dem i stand til at gennemføre en ungdomsuddannelse, når de har afsluttet folkeskolen.

Det fremgår samtidig af lovbemærkningerne, at det er en forventning, at man ved at indføre nationale test vil styrke skolernes evalueringskultur og dermed på sigt elevernes faglige niveau. Det er denne sammenhæng, der belyses i nærværende evaluering.

Det var et omfattende udviklingsarbejde og lovforberedende arbejde, der lå forud for lanceringen af de nationale test. Det er værd at bemærke, at lancering og implementering af de nationale test i årene 2006-2010 var genstand for stor debat og bevågenhed i medier, politisk og ikke mindst i skolen bredt set. En ambitiøs tidsplan, tekniske udfordringer og en grundlæggende debat om nødvendigheden af nationale test medførte, at projektets implementering oplevede betydelig turbulens indtil den endelige lancering i foråret 2010⁸. Det skal understreges, at driften af testsystemet i dag er velfungerende.

2.3 Om evalueringen

Dette afsnit redegør for evalueringens formål og på overordnet niveau for evalueringens metode og datakilder. For en mere detaljeret beskrivelse af metode og datakilder henvises til evalueringens bilag.

³ www.uvm.dk

⁴ OECD-rapport om grundskolen i Danmark, OECD, 2004.

⁵ Undervisningsdifferentiering i folkeskolen, Danmarks Evalueringsinstitut, 2004 og Løbende evaluering af elevernes udbytte af undervisningen i folkeskolen, Danmarks Evalueringsinstitut, 2004.

⁶ Evalueringsportalen er i dag nedlagt, og artikler og værktøjer er lagt ind på Undervisningsministeriets hjemmeside.

⁷ Herefter anvendes betegnelsen "Lovforslag om de nationale test".

⁸ Rigsrevisionens beretning om de nationale test fra 2010 indeholder en oversigt over de milepæle, som projektet har gennemgået, se [http://www.rigsrevisionen.dk/media\(1438,1030\)/05-2009.pdf](http://www.rigsrevisionen.dk/media(1438,1030)/05-2009.pdf).

2.3.1 Formål

Formålet med evalueringen er at vurdere, om de nationale test som pædagogisk værktøj har styrket skolernes evalueringskultur og elevernes faglige niveau.

Med udgangspunkt i dette overordnede formål definerer kommissoriet for evalueringen 16 evalueringsspørgsmål, der falder inden for følgende fire temaer:

- Tema 1: Testenes effekt
- Tema 2: Testenes anvendelse
- Tema 3: Testenes indhold
- Tema 4: Testenes udformning.

For oversigt over samtlige evalueringsspørgsmål henvises til bilag 1.

2.3.2 Metode og analysestrategi

Analyse af effekt og årsagssammenhænge

Nærværende evaluering undersøger sammenhængen mellem nationale test og skolernes evalueringskultur samt elevernes faglige resultater, jf. formålet ovenfor. Disse analyser skal belyse, hvorvidt anvendelsen af de nationale test som et pædagogisk værktøj har styrket skolernes evalueringskultur og herigennem elevernes faglige niveau.

Analysen af årsagssammenhænge er i evalueringen omsat til følgende tre selvstændige analyser:

1. I den første analyse udføres der en række regressionsanalyser, der sammenligner faglige resultater mellem elever, der har deltaget i nationale test, og elever der ikke har deltaget i nationale test. Variationen skabes ved at analysere testresultater for elever, der går på forskellige årgange og derfor adskiller sig mht. deltagelse i test. Ligeledes analyserer vi, om der er forskelle på resultater hos elever, der har været udsat for testnedbrud⁹ og altså ikke fik respons på deres resultat, og elever, der deltog i testen og således fik mulighed for at få feedback på deres præstation.
2. I en anden analyse søger vi via spørgeskemaundersøgelser at vurdere, om skolernes evalueringskultur er styrket efter indførelse af nationale test. Dette sker bl.a. ved at spørge skolerne om deres evalueringskultur har ændret sig over tid, samt ved at inkludere et mindre antal spørgsmål, der er anvendt i en spørgeskemaundersøgelse, inden de nationale test blev indført¹⁰. Skolernes svar uddybes ved dybdegående casestudier, og der sammenlignes med tidligere studier af danske skolers evalueringskultur for at vurdere, om der er sket en eventuel forbedring.
3. I den tredje form for analyse kombineres spørgeskemadata om skolernes evalueringskultur med testresultater. Metodisk set foretages en korrelationsanalyse for at undersøge, hvorvidt der er en sammenhæng mellem styrken af skolernes evalueringskultur og elevernes faglige resultater.

Der henvises til bilag 1 og 3 for en detaljeret beskrivelse af de anvendte tilgange til at afdække effekten af de nationale test.

Kombination af effektevaluering og virkningsevaluering

For at kunne undersøge ovenstående årsagssammenhænge har evaluator til nærværende evaluering udviklet et evalueringsdesign, der kombinerer virkningsevaluering og effektevaluering, mens tilgangen **contribution analysis** danner rammen for analysen¹¹. Evalueringsdesignet er udviklet

⁹ Begrebet "nedbrud" (eller lignende begreber) bruges som betegnelse for et midlertidigt teknisk nedbrud i testsystemet, der fandt sted i en kortere periode primo marts måned 2010. Det skal understreges, at driften af testsystemet efter evaluators oplysninger generelt er velfungerende.

¹⁰ The Impact of Public Management Reforms on Student Performance in Danish Schools, Public Administration, Vol. 86, No. 2, pp. 541-558, Simon Calmar Andersen, 2008.

¹¹ Addressing Attribution Through Contribution Analysis: Using Performance Measures Sensibly, The Canadian Journal of Program Evaluation, Vol. 16 (1), pp. 1-24., John Mayne, 2001.

med henblik på at kunne etablere viden om effekter af nationale test samt systematiske sammenhænge til forklaring af, hvordan effekter er skabt. Dette uddybes nedenfor.

Evalueringsstilgangen contribution analysis er udviklet med henblik på at sandsynliggøre sammenhængen mellem indsats og effekt i situationer, hvor det ikke er muligt at etablere en evalueringstilgang med brug af eksperimentelt design. I forhold til de nationale test er det ikke muligt at lave et randomiseret kontrolleret forsøg, da alle folkeskoler i henhold til loven skal gennemføre nationale test. En indsatsgruppe og kontrolgruppe kan derfor ikke etableres. Contribution analysis-perspektivet kan imidlertid bidrage til at sandsynliggøre, i hvilken grad effekt kan tilskrives den pågældende indsats, og via udvikling af en forandringsteori identificeres endvidere de faktorer, som forventes at lede frem til de endelige mål.

For at kunne sandsynliggøre årsagssammenhænge benytter contribution analysis teoribaseret evalueringsmetode med definition af en forandringsteori for årsag-virkningskæder samt opstilling af hypoteser om drivkræfter og barrierer for at realisere antagelserne i forandringsteorien. Endvidere kræver contribution analysis opstilling af alternative forklaringer på en positiv eller negativ effekt, der kan testes og dermed bidrage til at be- eller afkræfte forandringsteoriens vs. alternative forklaringskraft som drivende for de skabte effekter. Evaluatoren har på denne baggrund udviklet en forandringsteori med to spor: et spor med fokus på lærerne og den pædagogiske praksis og et spor med fokus på styring og ledelse. Med afsæt heri har evaluatoren endvidere udviklet en række hypoteser, der løbende testes i evalueringens analyser. Forandringsteori(er) og hypoteser er præsenteret i bilag 1.

Model for evalueringskultur

Som en del af evalueringsdesignet har evaluatoren operationaliseret begrebet evalueringskultur i relation til evalueringen af de nationale test og udviklet en model herfor. Modellen er præsenteret nedenfor og beskrives nærmere i bilag 2.

Operationaliseringen er sket med henblik på at kunne anskueliggøre og sandsynliggøre de nævnte årsagssammenhænge, herunder både for at kunne undersøge hvorvidt evalueringskulturen er styrket ved brug af de nationale test, og for at kunne undersøge testene som bidragende årsag til styrket fagligt niveau hos eleverne. Modellen bliver således brugt til at vurdere, om der er sammenhæng mellem graden af skolernes evalueringskultur og skolernes faglige resultater, og den giver samtidig et billede af, i hvor høj grad lærere, skoler og kommuner arbejder systematisk med de nationale test. Jo højere grad af systematik, jo stærkere evalueringskultur.

Modellen kan læses både vertikalt og horisontalt og afspejler dermed to forskellige dimensioner af evalueringskultur. En vertikal læsning af modellen viser en progression i aktørernes brug af resultatet af de nationale test gående fra monitorering alene til aktiv fremadrettet anvendelse på baggrund af samarbejde og dialog. En horisontal læsning af modellen afspejler en udvikling i sammenhængen mellem de tre aktørers arbejde med evaluering. Jo højere placeret en skole er i modellen, jo stærkere sammenhæng (eller overensstemmelse) er der mellem aktørernes arbejde med evaluering.

Af hensyn til overskuelighed indgår faglige vejledere, forældre og elever ikke direkte som aktører, men de er implicit tænkt ind både i relation til lærer- og skoleledersøjle.

Figur 2.1: Model for evalueringskultur

Aktørfokus	Lærer	Leder	Kommune
	Pædagogisk udvikling (Undervisning)	Forandringsledelse (Ledelse af pædagogisk udvikling)	Styring
Niveau 4	Kollegialt samarbejde om formativ anvendelse af NT som ét blandt flere evalueringsredskaber mhp. undervisningsdifferentiering.	Organisatorisk reflekteret anvendelse af NT som ét blandt flere evalueringsredskaber mhp. på udvikling af skolens undervisning.	Formativ anvendelse af og dialog om NT og øvrige evalueringsresultater mhp. at målrette skolernes faglige og pædagogiske arbejder.
Niveau 3	Delvist kollegialt samarbejde og delvis formativ anvendelse af NT som ét blandt flere evalueringsredskaber mhp. På undervisningsdifferentiering.	Organiseret kollegialt samarbejde mhp. formativ anvendelse af NT som ét blandt flere evalueringsredskaber.	Delvis formativ anvendelse af og begrænset dialog om NT mhp. at målrette skolernes faglige og pædagogiske arbejder.
Niveau 2	Klasserumsbaseret monitorering. Summativ anvendelse af nationale test som enkeltredskab.	Skolebaseret monitorering. Summativ anvendelse af NT som enkeltredskab.	Kommunal monitorering. Summativ anvendelse af nationale test som enkeltredskab.
Niveau 1	Elevbaseret monitorering af NT som enkeltredskab.	Monitorering af NT som enkeltredskab.	Monitorering af NT som enkeltredskab.

Modellen har i evalueringen endvidere fungeret som udgangspunkt for udarbejdelse af spørgeskemaer, interviewguides til casestudier og effektanalyser.

Dokumentationsgrundlag

For at kunne belyse evalueringens fire temaer omfatter evalueringsdesignet brug af såvel kvantitative som kvalitative samfundsvidenskabelige metoder. Evalueringen baserer sig på et omfattende datamateriale, der er indsamlet gennem de dataindsamlingsaktiviteter, der fremgår af tabellen nedenfor.

Tabel 2.2: Dataindsamlingsaktiviteter

Datakilde	Beskrivelse
Desk research	Gennemgang af relevant lovgivning, baggrundsmateriale, undersøgelser og forskning.
Registeranalyser	Regressionsanalyser af elevers resultater i nationale test, afgangsprøver mv. med henblik på at vurdere effekten af de nationale test.
Spørgeskemaundersøgelser	Spørgeskemaundersøgelser blandt forvaltninger, skoleledere og lærere med henblik på at identificere og forklare eventuelle sammenhænge mellem indførelsen af de nationale test og hhv. evalueringskultur og elevernes faglige niveau. Dertil kommer afdækning af en lang række øvrige dimensioner af de nationale test.
Casestudier	Casestudier af to dages varighed på 10 skoler med interview med skoleledere, lærere, faglige vejledere, forældre, elever og forvaltningsrepræsentanter samt observation af afvikling af de nationale test.
Ekspertvurdering 1 og 2 (testenes indhold)	Evaluator har i samarbejde med en række faglige eksperter vurderet indholdet af de nationale test med henblik på at identificere og beskrive indholdsmæssige styrker og svagheder.

Sammenfattende hviler evalueringen på et omfattende datamateriale, der har sikret, at evalueringsspørgsmålene er belyst ved hjælp af forskellige perspektiver og datakilder. Generelt har datatriangulering således været et grundlæggende princip i vores analyse af data og efterfølgende vurdering og konklusion i relation til evalueringsspørgsmålene.

Datakvalitet

Evaluatoren vurderer, at datakvaliteten i nærværende evaluering samlet set er særdeles god. Evalueringen bygger på et omfattende datagrundlag og gør brug af såvel kvantitative som kvalitative data.

De *kvantitative data* (registeranalyser og spørgeskemaundersøgelser) bidrager overordnet set til at indsamle viden i bredden og til at etablere den kausale sammenhæng mellem indsats og effekt (effektevaluering). Testresultater og karakterer i afgangsprøver er anvendt som datakilder til at vurdere effekt og årsagssammenhænge.

Registeranalyserne er udført ved at anvende registerdata i form af anonymiserede baggrundsoplysninger fra Danmarks Statistik samt oplysninger om faglige resultater i nationale test samt folkeskolens afgangsprøver fra UNI-C. Fordelene ved registerdata er, at de er indsamlet på konsistent vis over tid, de kan kobles til andre datakilder via cpr-nummer og giver mulighed for at følge individer over tid. Der er benyttet anonymiserede registerdata fra Danmarks Statistik til at sammenligne den sociale baggrund for de elever, der gennemførte testen, med de elever, der ikke gennemførte på grund af det midlertidige nedbrud. Denne analyse bekræfter antagelsen, at det var tilfældigt, hvilke elever der blev ramt af nedbrud. Således har vi et grundlag for at lave en robust analyse af forskelle i faglige resultater mellem elevernes faglige resultater og kan tillægge eventuelle forskelle deres deltagelse i de nationale test.

Spørgeskemaundersøgelserne er gennemført med høje svarprocenter for forvaltning, skoleledere og lærere. Mellem 59 og 80 pct. for hver af de tre respondentgrupper har besvaret spørgeskemaet. Dertil kommer, at spørgeskemaundersøgelsen blandt forvaltningen dækker næsten alle kommuner i Danmark og således næsten udgør en totalundersøgelse. Samlet set medfører de høje besvarelsesprocenter et repræsentativt kvantitativt datagrundlag, der bidrager til at sikre en høj datakvalitet.

De kvalitative data (casestudier, internationalt litteraturstudie og ekspertseminarer) bidrager til at indsamle dybdegående viden om anvendelse, indhold og udformning af de nationale test. Mere specifikt anvendes de kvalitative data til at belyse, forstå og forklare de kontekstuelle faktorer, der kan sandsynliggøre, hvorvidt effekter kan tilskrives indsatsen (virkningsevaluering).

Der er gennemført 10 casestudier af to dages varighed. De 10 cases er udvalgt, så der sikres dybdegående kvalitativ viden fra forskellige typer skoler. Der er sikret spredning i relation til de af evaluatoren udvalgte kriterier. Casestudierne er gennemført som en kombination af interview og observation. Interviewene sikrer, at der er indsamlet viden om de nationale test fra alle de involverede parter, mens observationerne bidrager med viden om praksis i relation til gennemførelsen af de nationale test.

Vurderingen af indholdet af de nationale test omfatter en vurdering af fordele og ulemper ved det adaptive princip samt en vurdering af styrker og svagheder ved udformning og indhold af testopgaver. Disse analyser forudsætter indgående teknisk og faglig indsigt, hvorfor evaluatoren har gennemført disse i samarbejde med førende eksperter på de pågældende områder. Dette bidrager til at sikre kvaliteten af disse analyser.

Kombination af kvantitative og kvalitative data giver et solidt grundlag for metodetriangulering, hvilket er et grundlæggende princip i tilgangen til evalueringen. Metodetriangulering giver mulighed for at belyse evalueringens temaer og evalueringsspørgsmål gennem såvel datakilder (fx lærere, skoleledere og forvaltningsrepræsentanter) som dataindsamlingsmetoder (fx spørgeskemaundersøgelser, casestudier mv.). Denne grundlæggende tilgang øger evalueringens validitet.

2.4 Læsevejledning

Rapporten indeholder foruden resumé og nærværende indledning seks analysekapitler og et konkluderende kapitel. Kapitlerne ser ud som følger:

- Kapitel 3 består af en analyse af **de nationale tests effekt** på elevernes faglighed og skolernes evalueringskultur samt sammenhængen herimellem (evalueringens tema 1)
- Kapitel 4, 5 og 6 indeholder delanalyser af **de nationale tests anvendelse**. Kapitel 4 omfatter den pædagogiske anvendelse af de nationale test, kapitel 5 undersøger forældre og elevs oplevelse af de nationale test, og i kapitel 6 undersøges de nationale test som dialog- og styringsredskab (evalueringens tema 2)
- Kapitel 7 består af en analyse af **de nationale tests indhold**, herunder udformning og indhold af testopgaver samt fordele og ulemper ved det adaptive princip (evalueringens tema 3)
- Kapitel 8 indeholder en analyse af **de nationale tests udformning**, herunder den praktiske tilrettelæggelse af testene på skolerne (evalueringens tema 4)
- Kapitel 9 består af evalueringens **konklusioner** på tværs af de seks analysekapitler og indeholder dermed en samlet besvarelse af evalueringsspørgsmålene. Dertil kommer en perspektivering af evalueringens resultater.

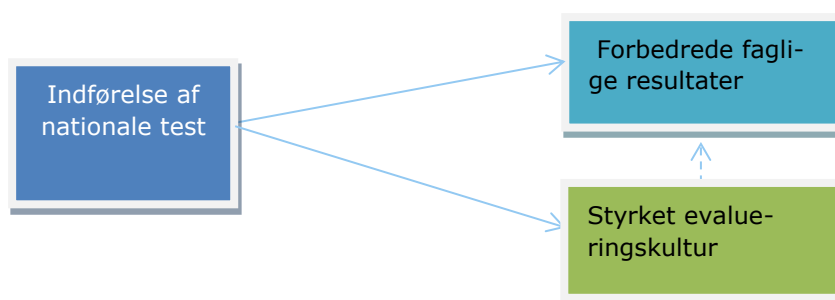
Rapportens bilagsmateriale omfatter en uddybende metodebeskrivelse (bilag 1), en model for evalueringskultur (bilag 2), en detaljeret redegørelse for registeranalyser (bilag 3), et selvstændigt notat om ekspertvurdering 2 (bilag 4) og fagspecifikke opmærksomhedspunkter fra ekspertvurdering 1 (bilag 5).

3. TESTENES EFFEKT

I dette kapitel besvares en række evalueringsspørgsmål, der er rubriceret under det følgende overordnede evalueringstema: *Testenes effekt*. Kapitlet indeholder således en samlet analyse af de effekter, som indførelsen af de nationale test har haft, ved at undersøge sammenhængen mellem nationale test og skolernes evalueringskultur samt elevernes faglige resultater.

De forskellige årsagssammenhænge, der analyseres i kapitlet, fremgår af figur 3.1 nedenfor. For en grundig beskrivelse af analysestrategien henvises til kapitel 2 samt metodebilaget (bilag 1).

Figur 3.1: Analyse af årsagssammenhænge



Kapitlet indledes med en analyse af, hvorvidt elevernes faglige niveau er styrket efter indførelsen af de nationale test. Dernæst undersøges det, hvorvidt skolernes evalueringskultur er styrket efter indførelsen af de nationale test. Disse to analyser er de centrale i evalueringens effektanalyse, idet de relaterer sig direkte til besvarelsen af de to evalueringsspørgsmål omhandlende effekt fra evalueringens opdrag.

Derudover har evaluatoren foretaget en supplerende analyse af sammenhængen mellem skolernes evalueringskultur og elevernes faglige resultater (markeret med stiplede linje i figuren ovenfor). Denne analyse bidrager til at belyse, hvorvidt anvendelsen af de nationale test som et pædagogisk værktøj har styrket skolernes evalueringskultur og herigennem elevernes faglige niveau.

Analyserne i nærværende kapitel er samlet set baseret på et datagrundlag bestående af følgende:

- Testresultater fra de obligatoriske nationale test i 2010, 2011 og 2012
- Resultater fra folkeskolens afgangsprøver
- Data om elevernes socioøkonomiske baggrund
- Spørgeskemaundersøgelser til kommunale forvaltninger, skoleledere samt lærere med erfaring inden for testfag
- Omfattende casestudier på udvalgte skoler
- Desk research af relevant forskning, undersøgelser mv.

Kapitlet er bygget op om hver af de nævnte sammenhænge, således at kapitlet besvarer hvert af de evalueringsspørgsmål, der relaterer sig til testenes effekt. I kapitlets delkonklusion foretages en samlet vurdering af de overordnede effekter ved indførelse af de nationale test.

3.1 Er elevernes faglige niveau styrket efter indførelse af de nationale test?

Som det fremgår af lovbemærkningerne til lovforslaget om de nationale test, er det overordnede formål med indførelsen af testene en forbedring af elevernes faglige niveau. I sidste ende er målet at give alle elever mulighed for at opnå sikre kundskaber og færdigheder, som sætter dem i stand til at gennemføre en ungdomsuddannelse, når de har afsluttet folkeskolen. Således er én af evalueringens nøglehypoteser, at anvendelsen af nationale test styrker elevernes faglige niveau. I det følgende vil det blive efterprøvet, om denne sammenhæng kan bekræftes¹².

¹² Der henvises til evalueringens forandringsteori og tilhørende hypoteser i bilag 1.

3.1.1 Hvordan kan de nationale test tænkes at have effekt?

Nationale test kan i princippet have effekt på både den enkelte elev, på skolen og på landsplan. Den enkelte elev kan blive dygtigere af at skulle gennemføre en test. Laboratorieforsøg har dokumenteret, hvordan elever bedre kan huske en tekst, hvis de er blevet testet i teksten (uden at få at vide, om de svarede rigtigt), end hvis de har fået tid til at repetere teksten. Så selve det at gennemføre en test kan hjælpe til at huske det, man er blevet testet i¹³. Den enkelte elev kan også blive bedre til at gennemføre nationale test ved at have prøvet det før. Således kan der dels være en øget fortrolighed med den it-baserede adaptive testform eller – for nogens vedkommende – at blive testet i det hele taget, ligesom der kan være en effekt af at have stiftet bekendtskab med typen af opgaver. Endvidere kan der opstå en effekt for den enkelte elev, fordi læreren (og eleven) ved hjælp af resultaterne af testen kan se, hvor eleven har behov for at blive dygtigere og efterfølgende indrette undervisningen, så den svarer til den enkelte elevs behov og forudsætninger. Den sidste antagelse om, at lærerens pædagogiske anvendelse af testresultaterne skaber en forbedret faglig indlæring hos eleverne, står centralt i evalueringens forandringsteori. Derudover er der i forandringsteorien formuleret en forventning om, at lærerens tilbagemelding om testresultaterne gør eleven mere bevidst om egne faglige styrker og svagheder og medvirker til at motivere eleven til at lære.

På skoleniveau kan der opstå en effekt af de nationale test i den forstand, at skolens lærere samlet set bliver bedre til at få deres elever til at præstere godt. Det kan skyldes, at lærerne for hver test får øget kendskab til, hvilke typer spørgsmål testen stiller, og på den baggrund kan indrette undervisningen på, at eleverne bliver gode til at besvare netop den type spørgsmål. Dette øgede kendskab vil spredes på skolen. For det første vil informationen blive delt blandt lærerne på lærerværelset og således potentielt påvirke undervisningen hos lærere, som endnu ikke selv har set en national test. For det andet vil informationen om testenes indhold hurtigt kunne påvirke stort set alle elever på en given skole, fordi vi i Danmark anvender faglærere, der som oftest underviser flere klasser på flere klassetrin samtidig. Der kan således opstå, hvad man ofte betegner som *teaching to the test*. Ikke blot hos en enkelt lærer, men i store dele af lærerkollegiet på en skole. Hvis testen dækker (en del af) de discipliner, som man ønsker, at eleverne skal blive dygtige til for at opfylde Fælles Mål, behøver der ikke være noget negativt i *teaching to the test*. Nogle forskere peger endda på, at det er idealet, at læringsmål, undervisningsformer og test er afstemte i forhold til hinanden (alignment).¹⁴ Men hvis testene kun ligger på kanten af det, som man ønsker, at eleverne skal blive gode til, eller hvis *teaching to the test* tager for meget tid og opmærksomhed væk fra andre dele af faget, som er vigtige(re), kan det være problematisk.¹⁵ Der henvises til kapitel 4 for en yderligere analyse af dette emne i relation til testene.

Et centralt aspekt af forandringsteorien er tilsvarende, at der på skoleniveau kan opstå en effekt ved, at testene bidrager til at forstærke en evalueringskultur på hele skolen. Det kan ske ved, at resultater af test benyttes af lærere og skoleledelse til en dialog om, hvor skolen samlet set har behov for at styrke undervisningen, og hvad der kan gøres for at bidrage til dette. En sådan praksis beskrives i modellen for evalueringskultur som en organisatorisk reflekteret anvendelse af testene og er udtryk for det stærkeste niveau for skolers evalueringskultur.

Endelig kan der på landsplan være en tilsvarende effekt af de nationale test, som består i, at skoler landet over benytter resultater og erfaringer fra de gennemførte test til at indrette undervisningen, så eleverne generelt præsterer bedre. Igen kan det ske, dels ved at hele skolesystemet begynder at indsnævre undervisningen til de områder, der bliver testet, dels ved at man generelt lærer af og reflekterer over resultaterne, og på den baggrund udvikler bedre undervisningsformer og -materialer mv., der styrker elevernes læring inden for de faglige områder, hvor testene viser, der er et behov.

¹³ The Power of Testing Memory: Basic Research and Implications for Educational Practice, *Perspectives on Psychological Science* 1(3): 181-210, Roediger, H. L., III, and J. D. Karpicke, 2006.

¹⁴ Instructional Alignment: Searching for a Magic Bullet. *Educational Researcher* vol. 16 no. 8 16-20, Cohen, 1987. Teaching for Quality Learning at University. What the student does, Buckingham and Philadelphia: Open University Press, John Biggs, 2003, samt Alignment - how to bridge between qualifications and the learning process. Paper prepared for the conference on Quality Assurance, Accreditation and European Legal Education, Utrecht, The Netherlands, Torben K. Jensen, 2004.

¹⁵ School Accountability, Ch. 8 in E. A. Hanushek, S. Machin and L. Woessmann (eds.), *Handbook of the Economics of Education*, vol. 3., 383-421, Figlio, D. and S. Loeb, 2011.

3.1.2 Hvordan undersøges effekten af de nationale test?

Som udgangspunkt er det meget vanskeligt statistisk at undersøge effekten af de nationale test på faglige færdigheder. Data fra det nationale testsystem kan vise, hvordan eleverne har klaret testene over tid, men som udgangspunkt er det ikke til at sige, om en eventuel fremgang er en effekt af testene eller en effekt af andre faktorer. Udfordringen er, at der ikke er nogen at sammenligne med. De nationale test er obligatoriske for alle elever (dog med mulighed for fritagelse), og alle ti test blev udrullet til alle landets folkeskoler i foråret 2010. Det er således svært at finde en sammenligningsgruppe, så man kan vurdere, hvordan de faglige færdigheder hos de testede elever ville have udviklet sig, hvis de ikke havde taget en test. Man kan desværre heller ikke anvende de frivillige test til formålet, eftersom de skoler, lærere og elever, som har gennemført frivillige test, ikke udgør et tilfældigt udsnit.

Der findes meget få studier internationalt, der forsøger at undersøge effekten af nationale, standardiserede testsystemer – til trods for, at sådanne testsystemer, der skal holde skolerne ansvarlige for deres resultater, har bredt sig med stor hast over hele verden.¹⁶ Nogle steder har man søgt at sammenligne elevpræstationer i stater, der har indført sådanne systemer, med elevpræstationer i lignende stater uden samme systemer.¹⁷ Generelt tyder disse sammenligninger på, at indførelsen af disse systemer har positive effekter,¹⁸ men for det første er det en udfordring for den type undersøgelser, at indførelsen af sådanne testsystemer ofte er ledsaget af en række andre reformelementer, hvilket gør det vanskeligt at adskille effekten af testsystemet fra de andre aspekter af sådanne reformer. Dette er også tilfældet i Danmark, hvor de nationale test blev indført som en del af initiativer til at fremme af evalueringskultur i folkeskolen. For det andet er disse undersøgelser gennemført i en amerikansk kontekst med såkaldte "high stakes test", hvor der generelt har været knyttet stærke sanktioner til dårlige testresultater (fx lukning af skoler). I den sammenligning er de nationale test i Danmark "low stakes" og har først og fremmest været fremført som et pædagogisk værktøj. Det gør det vanskeligt at overføre resultaterne – specielt fordi undersøgelserne også tyder på, at truslen om sanktioner får skoler til at udnytte svagheder i systemet eller ligefrem til at snyde, hvilket modvirker de ellers positive effekter.¹⁹

For at imødegå disse vanskeligheder med at undersøge effekten af de nationale test i sig selv, har evaluator til denne evaluering anvendt to metoder til at opnå eksogen variation i deltagelse i nationale test. Til effekter på elev- og skoleniveau benytter vi os af et midlertidigt nedbrud i it-systemet i 2010, der gjorde, at ikke alle elever blev testet det år. Til effekter på landsplan undersøger evaluator forskelle mellem årgange lige før og lige efter testenes indførelse.

Til førstnævnte analyser af elev- og skoleniveau udnyttes det, at it-systemet brød sammen i en periode på ni dage fra den 2. til 10. marts 2010. Det betød, at halvdelen af de elever, der skulle have gennemført en test på det tidspunkt, ikke fik nogen test.²⁰ Ved at sammenligne disse elevers resultater ved efterfølgende nationale test, med resultaterne for tilsvarende elever, der gennemførte testen i 2010, kan effekten af de nationale test undersøges. Den anvendte metode kaldes "two stage least squares" og foregår i to trin (deraf navnet): I det første trin beregnes sandsynligheden for, at en elev er blevet testet. Her udnyttes det, at den gruppe der blev ramt af nedbrud, har en lavere sandsynlighed for at blive testet, fordi kun nogle af dem blev genbooket til en ny test. I andet trin estimeres sammenhængen mellem denne sandsynlighed for at blive testet og testresultater to år senere. Denne analyse hviler på den antagelse, at det var tilfældigt, hvilke elever der blev ramt af det midlertidige nedbrud. Analyser, der sammenligner de "nedbrudsramte" elever med de øvrige elever på alle de baggrundsfaktorer, som evaluator har registeroplysninger om (deres køn, alder, familieforhold, forældres uddannelse, indkomst, indvan-

¹⁶ School Accountability, Ch. 8 in E. A. Hanushek, S. Machin and L. Woessmann (eds.), Handbook of the Economics of Education, vol. 3., 383-421, Figlio, D. and S. Loeb, 2011.

¹⁷ Bl.a. The Impact of No Child Left Behind on Student Achievement. Journal of Policy Analysis and Management 30, 418-446, Dee, T. D. and B. Jacob, 2011.

¹⁸ School Accountability, Ch. 8 in E. A. Hanushek, S. Machin and L. Woessmann (eds.), Handbook of the Economics of Education, vol. 3., 383-421, Figlio, D. and S. Loeb, 2011.

¹⁹ Left Behind by Design: Proficiency Counts and Test-Based Accountability. Review of Economics and Statistics 92, 263-283, Neal, D. and D. W. Schanzenbach, 2010, samt Teaching to the rating: School accountability and the distribution of student achievement Journal of Public Economics 92: 1394:1415, Reback, R., 2008.

²⁰ Skolerne blev tilbudt mulighed for at booke en ny test, hvilket skete for halvdelen af de 21.697, der blev udsat for et nedbrud i dansk, læsning.

drerstatus mv.), bestyrker, at den antagelse er korrekt.²¹ Der er kun få statistisk signifikante forskelle mellem de to grupper.

Med en såkaldt skole "fixed-effects" (FE)-analyse, er det endvidere muligt at isolere den effekt af de nationale test, som gælder på elevniveau, dvs. for de enkelte elever på en skole, fra den effekt af de nationale test, der påvirker hele skolen (skoleniveau). Med en FE-analyse undersøges det, om præstationerne er bedre eller dårligere for elever, der ikke blev testet, i forhold til elever på samme skole, der blev testet. Det vil sige, at det her undersøges, om der er en effekt for den enkelte elev udover, hvad der måtte være af effekter, som løfter hele skolen og dermed alle elever (fx effekter af *teaching to the test*, som kan opstå på skoleniveau, hvis lærernes viden om testenes præcise indhold spredes på skolen og får dem til at træne eleverne specifikt i at blive gode til at gennemføre næste test).

Endvidere undersøges det, om effekten på elevniveau skyldes, at eleverne generelt bliver bedre til at gennemføre en *hvilken som helst* af de adaptive, it-baserede, selvscorende test, eller om de blot præsterer bedre i det fag, de er blevet testet i før. Dette gøres ved at sammenligne om effekten af det midlertidige nedbrud – fx i dansk, læsning – er den samme for alle efterfølgende fag – fx i dansk, læsning, fysik/kemi, geografi og biologi – eller om den er størst i det fag, eleven oplevede nedbrud i (her dansk, læsning).

Til disse analyser benyttes registerdata fra UNI-C og Danmarks Statistik for de elever, der gik i 2. til 8. klasse på en folkeskole i skoleåret 2009/10. Det vil sige de elever, der var til national test i foråret 2010, da der var midlertidigt nedbrud i it-systemet. Evaluator ser på effekten af nedbrud i dansk, læsning på resultaterne i dansk, læsning to år efter. Evaluator benytter det midlertidige nedbrud i dansk, læsning, fordi det er det eneste fag, som eleverne bliver testet i igen to år senere (2012). Der ses endvidere på, om der er effekt af nedbrud i dansk, læsning på andre fag end dansk, læsning. Når data senere bliver tilgængelige for 2013, vil det være muligt tilsvarende at undersøge, om det midlertidige nedbrud i matematik i 3. klasse i 2010 har effekt på resultaterne i matematik i 6. klasse i 2013.

Omtrent 14 pct. af eleverne var udsat for det midlertidige nedbrud i den korte periode i marts 2010. Der benyttes baggrundsdata om eleverne som en ekstra kontrol for, at resultaterne ikke er drevet af, om gruppen af elever, der blev udsat for test, adskilte sig systematisk fra de øvrige elever. Resultaterne påvirkes ikke substantielt af at inddrage disse oplysninger, hvilket bestyrker resultaternes robusthed.

Som nævnt undersøges det også, om der på *landsplan* kan ses en effekt af indførelsen af de nationale test. Denne analyse kan indfange, at der udover effekterne på elev- og skoleniveau også kan ske at være en effekt på hele skolesystemet. Det er blandt andet interessant, fordi den anvendte metode betyder, at analyserne på elev- og skoleniveau ikke kan vise, om de elever, der som følge af nedbruddet ikke blev testet, også har oplevet en positiv effekt af indførelsen af de nationale test. Derfor har evaluator også undersøgt effekter på landsplan.

Det gøres ved at sammenligne karaktererne for de årgange af elever, der var til folkeskolens afgangsprøver i 2009 og 2010 – det vil sige, lige inden de nationale test var indført for dem (i 8. klasse) – med karaktererne for de efterfølgende to årgange i 2011 og 2012 (der havde nået at blive testet i 8. klasse). Endvidere sammenlignes resultaterne i de fag, eleverne blev testet i i 8. klasse, med de fag, der ikke blev testet i 8. klasse. Hvis der skulle være en generel ændring i elevernes resultater fra 2009 og 2010 til 2011 og 2012, som ikke har at gøre med indførelsen af nationale test i bestemte fag, må det forventes, at dette afspejles i de forskellige fag generelt. Hvis man udover en sådan generel udvikling kan se en særlig positiv udvikling i de fag, der blev indført test i, tages det som et udtryk for en samlet effekt af de nationale test på landsplan. Denne analyse hviler på relativt strenge antagelser, idet der ikke må være systematiske forskelle i karaktererne som følge af opgaveudformningen fra år til år og heller ikke i de faglige evner hos eleverne på tværs af årgange, og at det i den forstand var tilfældigt, at de nationale test blev

²¹ Beskrivende analyser fra UNI-C af bortfaldet ved den første testning bestyrker ligeledes denne antagelse. Se Bortfaldsanalyse, Nationale Test, UNI-C, 2010.

indført lige i foråret 2010 og ikke fx et år før eller et år senere.²² Denne antagelse kan ikke yderligere testes, og evaluator må derfor tage disse analyser med et vist forbehold. Omvendt kan man sige, at hvis resultaterne peger i samme retning som analyserne af elever, der blev ramt af nedbrud, styrker det yderligere vores tiltro til, at resultaterne faktisk påviser den reelle effekt af de nationale test.

Til analyserne benyttes registerdata fra UNI-C og Danmarks Statistik for børn født i perioden 1994-2002. Der henvises i øvrigt til rapportens bilag 3 for yderligere specifikationer.

3.1.3 Hvad viser analyserne af effekten af de nationale test?

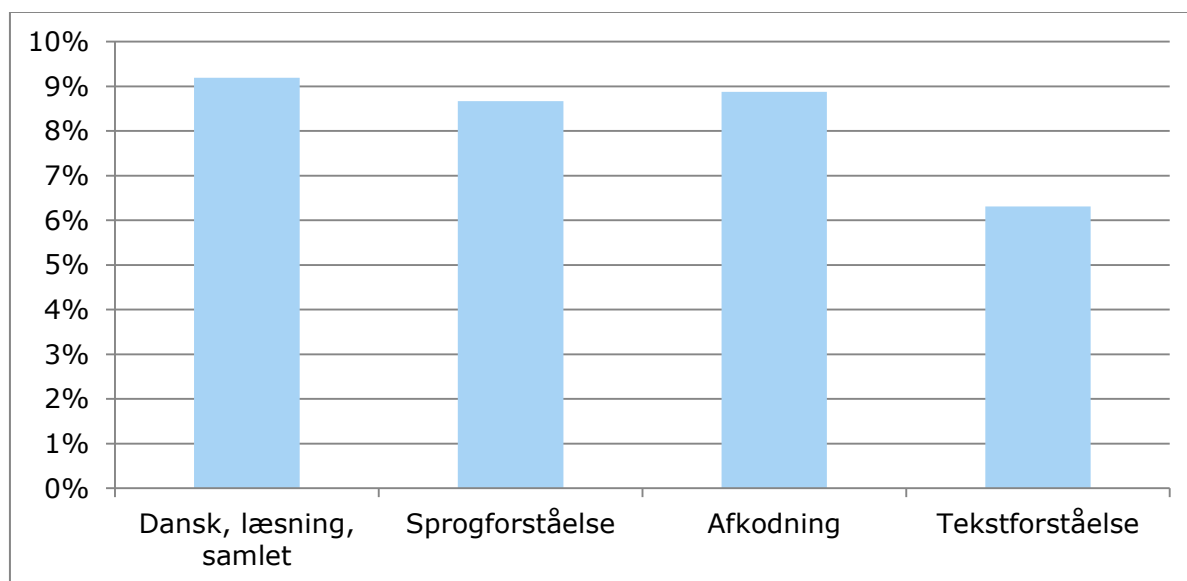
Effekter på elev- og skoleniveau

I det følgende præsenteres først resultaterne af, at der var et midlertidigt nedbrud i it-systemet i 2010. Vi undersøger effekten af at opleve et nedbrud i faget dansk, læsning, fordi dette fag bliver testet hvert andet år (2., 4., 6. og 8. klasse), og det derfor på nuværende tidspunkt er muligt at undersøge, om der er en effekt på resultaterne i samme fag to år efter.²³

Hovedresultaterne ses i figur 3.2. Figuren viser, hvor meget bedre, de elever, der gennemførte en test i 2010, klarede sig i 2012 i dansk, læsning (i 4., 6. og 8. klasse). Det fremgår, at eleverne scorede omkring 9 pct. (af en standardafvigelse²⁴) bedre end elever, der ikke blev testet. Det er vanskeligt at give en substantiel fortolkning af, hvor stor denne effekt er, men det svarer cirka til gabet mellem drenge og piger. Til sammenligning er det en effekt, der er mindst lige så stor som estimerne af effekten af de amerikanske testsystemer.

Figuren viser endvidere, at effekten af de nationale test i dansk, læsning er mindst på profilområdet tekstforståelse. Effekten er dog statistisk signifikant på alle tre profilområder.

Figur 3.2: Effekten af at gennemføre en test i dansk, læsning på præstation to år efter



Note: Samlet testscore og opdelt på de tre profilområder målt på standardiseret skala. Effekten er statistisk signifikant i alle fire tilfælde.

I nedenstående figur 3.3 har vi med den såkaldte skole "fixed-effects"-model isoleret den del af effekten, der kan tilskrives forskelle i resultater for elever fra samme skole (altså effekter på

²² Der kan godt være andre ikke-tilfældige årsager til, at det blev indført i 2010 – for eksempel den politiske udvikling, den tid det tog at udvikle it-systemet mv. Det afgørende for vores analyse er antagelsen om, at eleverne i disse årgange ikke i udgangspunktet havde systematisk forskellige faglige forudsætninger. Evaluator slår årgangene hhv. 2009 og 2010 samt 2011 og 2012 sammen for at gøre denne antagelse mere robust.

²³ Det eneste andet fag, der bliver testet to gange i løbet af skoletiden, er matematik. Først senere vil data blive tilgængelige for 2013, så det kan undersøges, om nedbrud i matematik i 3. klasse i 2010 har effekt på resultaterne i matematik i 6. klasse i 2013.

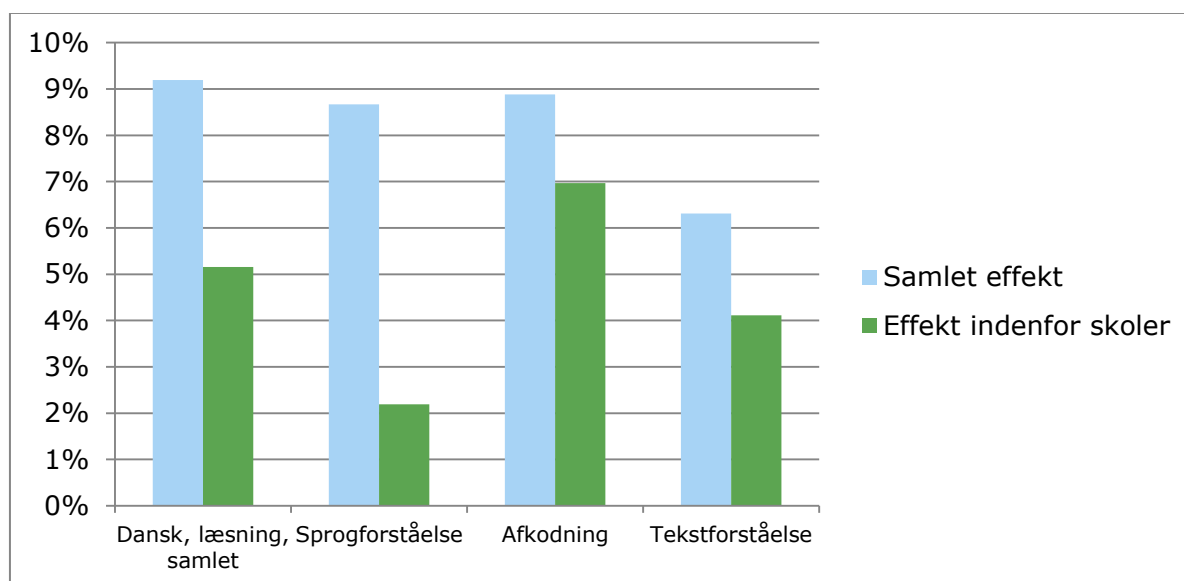
²⁴ De nationale test måles som udgangspunkt på en såkaldt Rasch-skala. Evaluator har standardiseret skalaen, så den har gennemsnit på 0 og standardafvigelse på 1. Effekten måles derfor i, hvor meget af en standardafvigelse (der er mål for den gennemsnitlige spredning i resultaterne) de nationale test flytter eleverne.

elevniveau), og som derfor ikke kan skyldes, at hele skolen er blevet bedre, fx som følge af forbedret evalueringskultur eller øget *teaching to the test*. De blå søjler i figur 3.3 er de samme som i figur 3.2.

De nye grønne søjler viser, at den del af effekten, der skyldes forskelle inden for elever på samme skole, samlet set udgør knap halvdelen af den samlede effekt. På denne måde udelukker den effekt af de nationale test, der skyldes, at nogle skoler samlet set opnår bedre resultater, og man fokuserer alene på den forbedring, enkelte elever inden for en skole opnår. Alligevel er der altså stadig en statistisk signifikant og betydelig positiv effekt af de nationale test. Disse resultater viser med andre ord, at den positive effekt af de nationale test ikke kun skyldes, at skoler samlet set opnår bedre resultater af at have gennemført de nationale test (hvilket blandt andet kunne hænge sammen med, at lærerne på skolen i højere grad indretter undervisningen efter testens indhold). Noget af den positive effekt skyldes også, at de enkelte elever på en skole, der har gennemført en test, klarer sig bedre end deres skolekammerater, der ikke blev testet (på grund af det midlertidige nedbrud). Det indikerer, at eleverne lærer noget af at blive testet, eller at lærerne benytter de enkelte elevers resultater fra de gennemførte test som et pædagogisk værktøj til at målrette undervisningen til de enkelte elevers behov (jf. diskussionen i kapitel 4).

Effekten inden for skoler stammer hovedsageligt fra bedre resultater inden for profilområdet afkodning. Der er ikke nogen statistisk signifikant effekt på de to andre profilområder. Forklaringen kan være, at afkodning er en forholdsvis konkret færdighed (kobling af bogstaver og lyde til ord), og at lærerne på området afkodning har de mest konkrete værktøjer til at sætte ind over for elever, hvor testresultaterne viser, at de har problemer på det område.

Figur 3.3: Effekten af at gennemføre en test i dansk, læsning på præstation to år efter



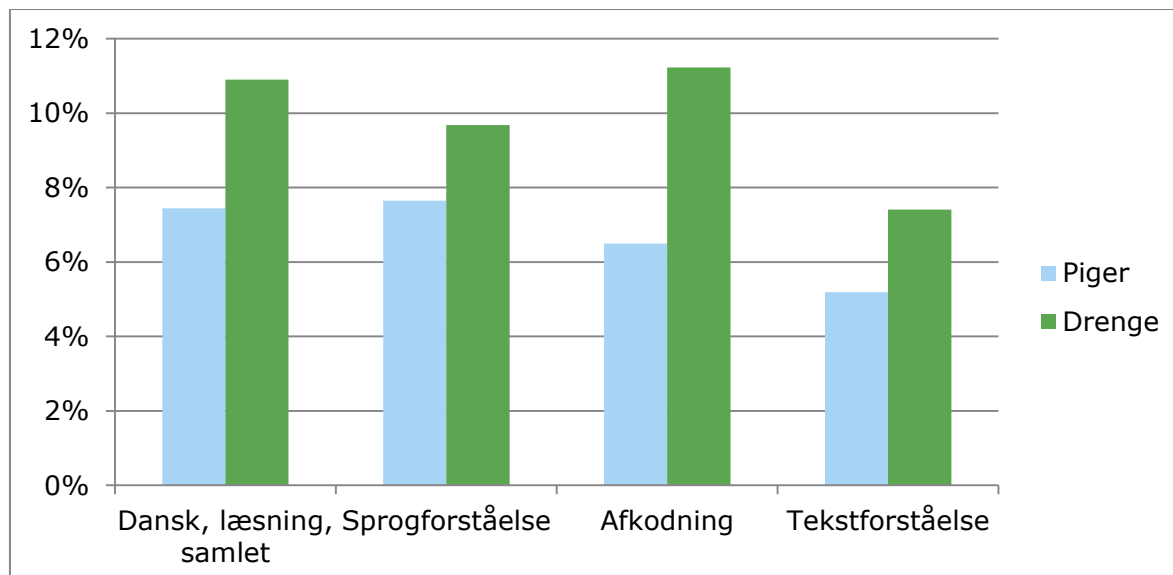
Note: Den samlede effekt (identisk med figur 3.2) og effekt inden for skoler. Samlet testscore og opdelt på de tre profilområder målt på standardiseret skala. Effekten er statistisk signifikant i alle tilfælde på nær sprogforståelse inden for skoler.

Evaluator har endvidere undersøgt, om effekten af de nationale test er forskellig for forskellige grupper af elever. Det er der nogle tendenser til, men generelt ikke større, end at det kan skyldes tilfældigheder i målingerne. Som eksempel er effekten af de nationale test i figur 3.4 opdelt for piger og drenge. Som det fremgår, er der en tendens til, at drengene har større effekt af at blive testet, men forskellene er ikke store og stabile nok til at være statistisk signifikante.²⁵ Der ses også en tendens til, at effekten af at blive testet er større for elever med indvandrerbaggrund (figur udeladt, se bilag 3), men igen er effekten på den samlede testscore ikke statistisk signifikant. Det er også undersøgt, om der er forskelle i effekten af at have gennemført en national test på tværs af forældres uddannelsesbaggrund (figur udeladt, se bilag 3). Der er ingen tegn på det-

²⁵ Vi anvender et konventionelt signifikansniveau på 5 pct., hvilket betyder, at vi kun regner det for et signifikant resultat, hvis forskellen er så markant, at der er mindre end 5 pct. sandsynlighed for at finde en sådan forskel, hvis det faktisk bare var tilfældigheder, der var på spil.

te. Generelt findes en positiv effekt af at gennemføre nationale test for både drenge og piger, på tværs af forældres uddannelsesbaggrund, og for både elever med og uden indvandrerbaggrund.²⁶

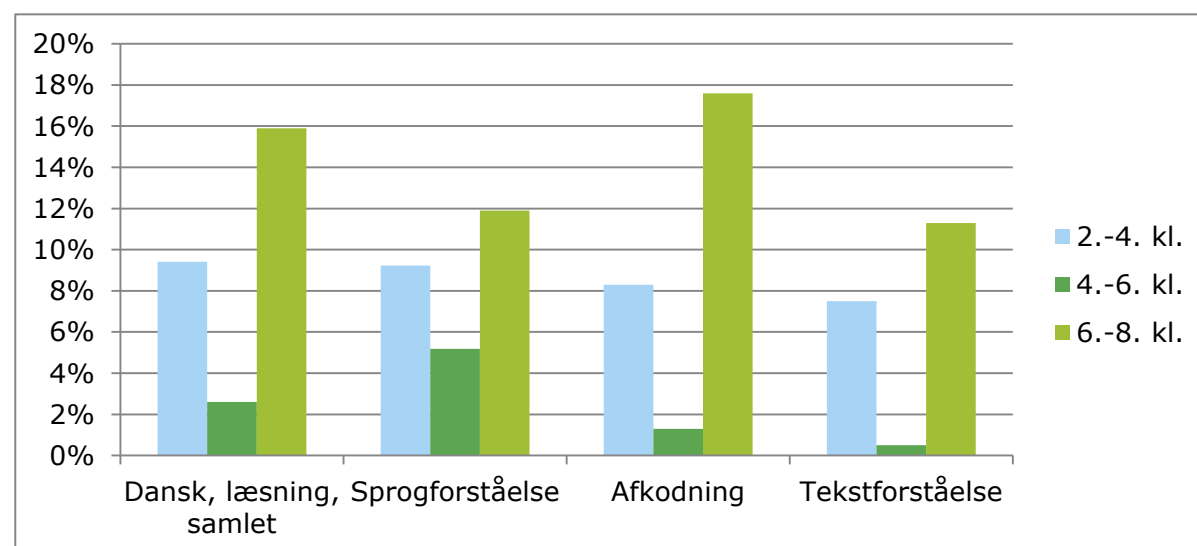
Figur 3.4: Effekten for piger og drenge af at gennemføre en test i dansk, læsning på præstation to år efter



Note: Samlet testscore og opdelt på de tre profilområder målt på standardiseret skala. Forskellene mellem piger og drenge er ikke statistisk signifikante.

I nedenstående figur 3.5 er effekten af de nationale test opdelt på klassetrin. Her ses, at effekten af at blive testet er mindre for elever, der gik i 4. klasse i 2010, end for elever der gik i 6. klasse. Der er ikke nogen statistisk signifikant effekt af at blive testet i 4. klasse.

Figur 3.5: Effekten af at gennemføre en test i dansk, læsning på præstation to år efter, opdelt på klassetrin



Note: Samlet testscore og opdelt på de tre profilområder målt på standardiseret skala. Effekten på 6. klassetrin er signifikant forskellig fra 4. klassetrin (på nær profilområdet Sprogforståelse) og forskellig fra 2. klassetrin på profilområdet Afkodning.

²⁶ Vi har også set på, om gennemførelse af nationale test i 8. klasse har en sammenhæng med afgangsprøvekaraktererne i 9. klasse. Resultaterne indikerer en positiv effekt og støtter således de øvrige resultater. Men da noget i data kunne tyde på, at det ikke er helt tilfældigt, hvilke 8. klasseelever der blev ramt af nedbrud, forfølger vi ikke dette spor yderligere.

Endelig har evaluator i forhold til analyserne af effekter på elevniveau undersøgt, om nedbrud i ét fag har effekt på resultaterne i andre fag. Hvis det er tilfældet, vil det indikere, at effekten af de nationale test til dels skyldes, at eleverne bliver bedre til det at gennemføre en test generelt og ikke alene skyldes, at de bliver dygtigere af den feedback, som testresultaterne giver.

Resultaterne viser, at en test i dansk, læsning i 2. klasse alene påvirker resultatet i dansk, læsning i 4. klasse og ikke påvirker resultatet i matematik i 3. klasse. Til gengæld har det at blive testet i dansk, læsning på 6. klassetrin både en effekt på resultatet i dansk, læsning og i fysik/kemi, biologi og geografi i 8. klasse. Disse resultater indikerer, at noget af effekten af de nationale test skyldes, at eleverne bliver dygtigere til det fag, de bliver testet i, og at noget af effekten af de nationale test skyldes, at eleverne bliver dygtigere til at gennemføre test i det hele taget. Det havde dog generelt styrket analysen, såfremt man havde haft en længere tidsperiode end den eksisterende, samt tilsvarende havde kunnet se på matematik over en treårig periode.

Effekt på landsplan

I forhold til effekten på landsplan, som sammenfatter både den effekt, der måtte være for den enkelte elev, for hele skoler og for hele folkeskolen (herunder også eventuelle effekter på kommuneniveau), har evaluator som nævnt sammenlignet afgangsprøvekaraktererne for årgangene lige før og lige efter indførelsen af de nationale test. For at gøre resultaterne mere robuste over for tilfældige forskelle mellem årgangene, sammenlignes årgangene, der gik til afgangsprøve i 2009 og 2010 (og som derfor ikke havde nået at gennemføre en test i 8. klasse), med dem, der gjorde det i 2011 og 2012. For at tage højde for en generel forskel mellem disse to grupper af årgange sammenlignes udviklingen i afgangskaraktererne i de fag, hvor eleverne var blevet testet ved nationale test i 8. klasse (dansk, læsning, fysik/kemi, biologi og geografi), med udviklingen i afgangskarakterer i de fag, hvor eleverne endnu ikke var blevet testet (matematik for årgang 2011 og 2012 og engelsk for årgang 2011). Analyserne viser et generelt billede af, at karaktererne er gået lidt mere frem i fag, hvor eleverne er blevet testet, sammenlignet med fag, hvor eleverne ikke er blevet testet. Dette bestyrker de foregående analysers resultater om, at de nationale test har en positiv effekt for eleverne. Denne analyse på landsplan skal dog tages med forbehold, fordi der er mange andre forhold, som påvirker ændringer i karakterer i fag over tid, herunder udformning af opgaver.

En samlet vurdering af, om de nationale test har styrket elevernes faglige niveau

Samlet set kan det konkluderes, at både analyserne på elev-, skole- og landsniveau samstemmende indikerer, at de nationale test har en positiv effekt på elevernes faglige præstationer. Resultaterne tyder endvidere på, at denne effekt ikke alene skyldes, at hele skoler er blevet bedre (hvilket kunne skyldes, at lærerne på skolerne har indrettet undervisningen efter *teaching to the test*). Noget af effekten ses også for de enkelte elever på skolerne, der gennemførte testen, i sammenligning med tilsvarende skolekammerater, der på grund af nedbruddet ikke fik gennemført en test. Det tyder på, at lærere og elever blandt andet har brugt resultaterne af testene til at arbejde på at blive dygtigere i faget. For elever, der blev testet i dansk, læsning på 6. klassetrin ses samtidig en tendens til, at de også opnår bedre resultater i andre fag i 8. klasse. Det indikerer, at noget af effekten skyldes, at eleverne bliver bedre til at gennemføre test i det hele taget.

Disse analyser på elev- og skoleniveau ser alene på effekten af nedbrud i dansk, læsning. Når data for 2013 bliver tilgængelige, vil det være muligt at undersøge, om den samme effekt ses i matematik.

Analyserne på elev- og skoleniveau kan ikke vise, om de, der som følge af det midlertidige nedbrud ikke blev testet, også har oplevet en positiv effekt af indførelsen af de nationale test. Derfor er der også set på effekter på landsplan. Denne analyse af årgangene før og efter, testene blev indført, viste, at der er en betydelig positiv effekt af de nationale test også på landsplan. Noget af denne effekt må tilskrives de effekter på elev- og skoleniveau, som de første analyser viste. Men effekten er så stor, at det tyder på, at der også er en yderligere effekt på landsplan, som kan skyldes, at lærerne i kraft af de nationale test har haft et større fokus på undervisningen i disse fag. Denne effekt kan være kommet alle elever til gode – også de, der ikke blev testet på grund af nedbruddet i 2010.

Størrelsen på disse effekter er betydelige. Disse effekter er lige så store som i de amerikanske testsystemer, hvor der kan være meget stærke sanktioner knyttet til dårlige testresultater og omvendt store økonomiske incitamenter til gode resultater. Det kunne tyde på, at nationale test har en effekt som et pædagogisk værktøj uden, at det er fordi det er drevet af trussel om sanktioner eller løfte om økonomiske belønninger.

Sammenfattende kan **hypotesen** om, at anvendelsen af nationale test styrker elevernes faglige niveau, således bekræftes. Analyserne indikerer, at de nationale test har haft en positiv effekt på elevernes faglige resultater. Det har dog alene været muligt at se på effekten af at eleven har gennemført én enkelt national test i dansk, læsning, hvorfor resultaterne skal læses med dette for øje. Derudover har det kun været muligt at undersøge de generelle effekter af de nationale test, dvs. om testene har en effekt på elevernes faglige niveau, hvorimod evaluatoren i kapitel 4 og 6 mere dybdegående vil undersøge, *hvordan* lærere og skoleledere anvender de nationale test. Således vil evaluatoren her forsøge at relatere effektanalysernes resultater til forandringsteoriens forventninger om, hvordan de nationale test opnår deres effekter.

3.2 Er skolernes evalueringsskultur styrket efter indførelse af de nationale test?

En væsentlig bevæggrund for at indføre nationale test i den danske folkeskole var opfattelsen af, at evalueringsskulturen i den danske folkeskole var utilstrækkelig, jf. beskrivelsen i kapitel 2. Udover dokumentation fra OECD og Danmarks Evalueringssinstitut var der samtidig bred enighed blandt folkeskolens parter om, at der var en udfordring med skolernes evalueringsskultur²⁷.

Et af svarene på denne evalueringssudfordring blev indførelsen af ti nationale test. Det fremgik af lovbemærkningerne, at man forventede, at man ved at indføre nationale test vil styrke skolernes evalueringsskultur og dermed på sigt elevernes faglige niveau. Dette er tilsvarende kernen i nærværende evalueringss forandringsteori, jf. bilag 1. Således er én af evalueringssens hypoteser, at anvendelsen af nationale test styrker skolernes evalueringsskultur.

I det følgende vurderes det, hvorvidt der i de senere år, **overordnet set**, er sket en styrkelse af skolernes evalueringsskultur. Her vil evaluatoren anvende skolernes egne kvalitative vurderinger samt tidligere undersøgelser til at foretage en vurdering af, om der samlet set er sket en positiv udvikling i evalueringsskulturen i den danske folkeskole (som forventet i lovbemærkningerne og i forandringsteorien). Dernæst vil det blive undersøgt, hvorvidt en eventuel styrkelse kan tilskrives de nationale test for at kunne afdække det relevante evalueringssspørgsmål.

3.2.1 Er skolernes evalueringsskultur blevet styrket?

For at kunne undersøge hvorvidt der, **overordnet set**, er sket en styrkelse af den danske folkeskoles evalueringsskultur, er vi nødt til at definere evalueringsskultur som begreb. I evalueringssens udbudsmateriale præsenterer Kvalitets- og Tilsynsstyrelsen nedenstående definition af evalueringsskultur.

²⁷ Folkeskolens svar på OECD's anbefalinger. Tilbage melding fra KL, Danmarks Læreforening, Lederforeningen, Danmarks skolelederforening, Skole og Samfund, Børne- og kulturchefforeningen, 2004.

Boks 3.1: Definition af evalueringskultur

Evalueringskultur handler om en lærers, en skoles eller en kommunes evne til at arbejde med og anvende evalueringer på forskellige niveauer, så evalueringer adapteres og gøres til rutiner. Arbejdet med evaluering skal foregå på både kommune-, skole-, team-, klasse- og elevniveau og skal foretages systematisk og løbende med henblik på at styrke elevernes udbytte af undervisningen.

Evalueringer skal være baseret på fastsættelse af mål samt efterfølgende refleksioner og vurderinger. Resultater af evalueringerne skal formidles og dokumenteres skriftligt, være suppleret af mundtlig formidling og have et fremadrettet perspektiv.

Der skal løbende følges op på evalueringsarbejdet, så der skabes grobund for, at arbejdet med evalueringskulturen hele tiden udvikler sig og tilpasses inden for konteksten.

Kilde: Kvalitets- og Tilsynsstyrelsen, baggrundsnotat om nationale test i folkeskolen, 2012.

Kernen i denne definition af evalueringskultur handler om skolers evne til *systematisk* og *løbende* at arbejde med og *anvende* evalueringer på forskellige niveauer med henblik på at styrke elevernes udbytte af undervisningen. Det er denne definition, som evaluator har taget udgangspunkt i ved udviklingen af en model for evalueringskultur, og det er denne definition, som vi i det følgende vil anvende til at vurdere, om skolernes evalueringskultur som helhed er styrket.

Til at understøtte denne vurdering har evaluator inddraget andre undersøgelser, som har undersøgt udviklingen af den evalueringskultur, som man ønskede at styrke med ændringerne af folkeskoleloven i 2006-2007. Der ses først på, hvordan evalueringskulturen blev vurderet, *inden* de nationale test blev indført. Derefter vurderer evaluator data om samme emne *efter* indførelsen af de nationale test. I denne forbindelse suppleres der med data, indhentet via evaluators casestudier. Samlet skal dette bidrage til at vurdere, om der reelt er sket en styrkelse af skolernes evalueringskultur, men altså uden at tillægge en eventuel udvikling af de nationale test. Denne sammenhæng vil først blive undersøgt i de efterfølgende afsnit.

Evalueringskulturen *inden* de nationale test

Som allerede nævnt udgav OECD i 2004 et meget omtalt review, hvis overordnede budskab var fraværet af en evalueringskultur i den danske folkeskole²⁸. Samme år kritiserede EVA folkeskolen for det samme i deres evaluering af undervisningsdifferentiering²⁹ og i deres evaluering af løbende evaluering i grundskolen³⁰. EVA konkluderede, at evaluering ikke fylder ret meget og er en løstrevet del af den daglige læringspraksis, samt at valget af evalueringsredskaber er mere eller mindre tilfældig. I forlængelse heraf opfordrede bl.a. KL i en pjece fra 2005 skolerne til at begynde at arbejde med evalueringskultur, idet dette er med til at skabe en bedre faglig kvalitet i folkeskolen.³¹

I 2006 undersøgte EVA læringsmiljøer i folkeskolen³², hvor de igen anførte, at evalueringskulturen endnu ikke er etableret i skolevæsenet. Året efter vurderede EVA dog, at skolerne var i gang med at opbygge en evalueringskultur, selv om man naturligvis ikke var nået lige langt på alle skoler.³³

I OECDs TALIS-undersøgelse fra 2008 undersøgte man danske samt en række andre landes læreres anvendelse af evaluering i skolen. Undersøgelsen viste, at danske lærere i langt mindre grad end andre landes lærere benytter evaluering. Der er relativt få lærere, der mener, at evalu-

²⁸ OECD-rapport om grundskolen i Danmark, OECD, 2004.

²⁹ Undervisningsdifferentiering i folkeskolen, Danmarks Evalueringsinstitut, 2004.

³⁰ Løbende evaluering af elevernes udbytte af undervisningen i folkeskolen, Danmarks Evalueringsinstitut, 2004.

³¹ Evalueringskultur – en ny dansk skoletradition, Kommunernes Landsforening, 2005.

³² Læringsmiljøer i folkeskolen – Samspillet mellem læringssynet, de fysiske rammer, undervisningens tilrettelæggelse og evalueringskulturen, Danmarks Evalueringsinstitut, 2006

³³ Klog på egen praksis. Danmarks Evalueringsinstitut 2007.

ering har medført ændringer i deres undervisningspraksis³⁴, og ca. en fjerdedel af lærerne angiver, at de arbejder på en skole, hvor de inden for de seneste fem år aldrig har evalueret hverken internt eller eksternt på skolen³⁵.

Endvidere har Skolerådet siden deres etablering i 2006 vurderet udviklingen i den danske folkeskoles evalueringskultur. I Skolerådets årsberetninger fra 2007-2009 – årene efter de nationale test blev vedtaget – pegede Skolerådet på en række redskaber, hvis anvendelse kan være med til at styrke evalueringskulturen og elevernes faglige resultater, heriblandt de nationale test. Det fremgår dog hverken, om evalueringskulturen i folkeskolen er blevet styrket eller svækket.

Evaluators konstaterer, at dokumentationen for begrænset/manglende evalueringskultur er ganske omfattende, inden indførelsen af de nationale test samt en række andre initiativer til styrkelse af evalueringskulturen i folkeskolen.

Evalueringskulturen efter de nationale test

I 2010 fremhæver Skolerådet i deres beretning, at evalueringskulturen er blevet styrket de seneste år, hvor blandt andet elevplaner og kvalitetsrapporter har haft en positiv indvirkning. De påpeger dog, at der stadig er plads til forbedring. I 2011 fremhæver de igen, at evalueringskulturen er blevet styrket, og nu indgår de nationale test som en del af forklaringen. I 2012 og 2013 konkluderer Skolerådet igen, at evalueringskulturen er styrket, men at de nye redskaber³⁶ har potentiale til at styrke evalueringskulturen endnu mere. Man vurderede bl.a., at én af udfordringerne er den pædagogiske anvendelse af de nationale test. Det er altså gennemgående en vurdering i Skolerådets beretninger, som baserer sig på den løbende dokumentation, der udmøntes i deres observationer og anbefalinger, at evalueringskulturen i den danske folkeskole er forbedret over årene.

I 2011 udgav OECD et omfattende review af evalueringspraksis i den danske folkeskole. OECD konkluderer i deres rapport, at der siden 2004 er iværksat mange initiativer, som har medført en styrkelse af folkeskolens evalueringskultur³⁷, og at Danmark *"over en kort periode har indført en vifte af obligatoriske foranstaltninger for elevernes læring, et system for kvalitetsrapportering, der involverer kommuner og skoler, og en national struktur til at overvåge resultaterne og vurdere prioriteringer i folkeskolen."*³⁸Således anfører forfatterne, at de første skridt i opbygningen af en evalueringskultur er taget, samt at skolerne har adgang til mere information om og flere redskaber til selvevaluering og forbedring.

I 2011 udgav EVA en evaluering af sammenhænge mellem evalueringsfaglighed og differentieret undervisning³⁹, hvor det konkluderes, at evalueringsarbejdet er styrket, og at lærerne i stigende grad påtager sig evalueringsopgaven. EVA finder dog ingen tydelig kobling mellem evalueringsindsatsen og undervisningsdifferentiering. De forklarer det blandt andet med, at evaluering stadig er frakoblet undervisningen, da lærerne primært anvender evalueringsresultaterne bagudrettet. De konkluderer bl.a., at *"i teorien burde lærerne anvende evalueringsresultaterne i tilrettelæggelsen af undervisningen, men enten indsamler de ikke tilstrækkelige data, eller også anvender de ikke den indsamlede viden til at fokusere undervisningen i forhold til læringsmålene"*.

Gennemgangen af eksisterende viden om udviklingen i den danske folkeskole viser, at der er igangsat et omfattende arbejde med fokus på at evaluere elevernes udbytte i folkeskolen, herunder at denne indsats tilsyneladende *overordnet set* har bidraget til en styrkelse af evalueringskulturen i den danske folkeskole.

³⁴ TALIS – Lærere og skoleledere om undervisning, kompetenceudvikling og evaluering – i et internationalt perspektiv, Skolestyrelsen, 2009.

³⁵ Creating Effective Teaching and Learning Environments – First Results from TALIS, Teaching And Learning International Survey – OECD, 2009.

³⁶ Elevplaner, kvalitetsrapporter og de nationale test.

³⁷ OECD Reviews of Evaluation and Assessment in Education: Denmark, OECD, 2011.

³⁸ Oversat fra "Over a short period of time, Denmark has introduced a suite of compulsory measures of student learning, a system of quality reporting involving municipalities and schools, and a national structure to monitor outcomes and evaluate priorities in compulsory education."

³⁹ Undervisningsdifferentiering som bærende pædagogisk princip – En evaluering af sammenhænge mellem evalueringsfaglighed og differentieret undervisning, Danmarks Evalueringsinstitut, 2011.

Denne vurdering, foretaget på baggrund af sekundære datakilder, understøttes af data fra de omfattende casestudier, som evaluator har gennemført på et bredt udsnit af skoler i forbindelse med evalueringen. I den forbindelse er det værd at fremhæve, at vi i vores casestudier finder opbakning til, at skolernes evalueringsskulptur er styrket i løbet af de senere år. Det afspejler sig først og fremmest i et større fokus på evaluering overordnet set, og et større fokus på faglige resultater. De interviewede aktører på skolerne fortæller, at det især er inden for de senere par år, at man er begyndt at få en systematik for, hvordan evalueringsskulpter anvendes i en samlet evalueringsskulptis. Som det påpeges i flere af de ovennævnte rapporter, viser casestudierne også, at den formative del af evalueringen – dvs. en fremadrettet pædagogisk brug af resultaterne med henblik på tilrettelæggelse af undervisningen, så den tilpasses den enkelte elevs behov. – er en udfordring.

3.2.2 Hvad er årsagen til den styrkede evalueringsskulptur?

Nedenfor undersøges det, hvorvidt indførelsen af de nationale test har medvirket til den umiddelbare styrkelse i evalueringsskulpturen.

Fremgangsmåde

En (i statistisk forstand) kausal undersøgelse af, hvorvidt skolernes evalueringsskulptur er blevet styrket efter indførelse af de nationale test, forudsætter, at man enten a) har valide data for skolernes evalueringsskulptur *før* og *efter* indførelsen af de nationale test og/eller b) at man kan sammenligne skoler, der har indført nationale test, med skoler, der ikke har, for at se om der er forskel i skolernes evalueringsskulptur som følge af brugen af nationale test.

Disse tilgange er ikke umiddelbart mulige at gennemføre, dels fordi der ikke foreligger solide, nationalt dækkende førmålinger i relation til evalueringsskulptur, dels fordi de nationale test ved lov er obligatoriske for alle folkeskoler i Danmark. Evaluator har i stedet benyttet andre data til at undersøge, hvorvidt skolernes evalueringsskulptur er blevet styrket efter indførelse af de nationale test. Dette er bl.a. sket gennem landsdækkende spørgeskemaundersøgelser med skoleledere, lærere og forvaltninger samt dybdegående casestudier på skoler. Evaluators model for evalueringsskulptur har fungeret som analytisk ramme for udarbejdelse af såvel spørgeskemaer som interviewguides.

Evaluator har anvendt følgende fremgangsmåde til at foretage analysen med:

1. For det første undersøges det, hvilke evalueringsskulpter skolelederne angiver, at de anvender, og hvor længe de har anvendt dem. Dette giver et indblik i omfanget af brugen af evalueringsskulpter, herunder om en given ændring i anvendelse er sket inden, samtidig med eller efter de nationale test blev indført. Ligeledes benyttes data fra samme spørgeundersøgelse til at vurdere, hvorvidt en række elementer er en del af skolernes evalueringsskulptis. I lighed med ovenstående har skolelederne skulle angive, hvornår de givne elementer blev en del af skolens evalueringsskulptis. Denne analyse giver et indblik i skolernes indstilling til samt anvendelse af evaluering
2. Dernæst vil vi benytte data fra de tre spørgeskemaundersøgelser med skoleledere, lærere og forvaltninger, hvor de nævnte målgrupper er blevet bedt om deres direkte vurdering af, hvorvidt de nationale test har medført en stærkere evalueringsskulptur i folkeskolen, og på hvilken måde dette kommer til udtryk. I dette tilfælde baserer analysen sig således på den *oplevede effekt af årsagssammenhængen*.

Evaluators kvalitative erfaringer fra en række casestudier på danske folkeskoler inddrages gennemgående i de tre analyser.

3.2.3 Har de nationale test medvirket til en styrkelse af skolernes evalueringsskulptur?

Analyse 1: Skolernes anvendelse af evalueringsskulpter samt deres evalueringsskulptis – hvornår ændrede det sig?

Jf. definitionen af evalueringsskulptur er skolernes evne til at anvende evalueringsskulpter og arbejde med dem på forskellige niveauer en central del af evalueringsskulpturen. Nedenstående tabel viser en fordeling af skoleledernes svar på spørgsmålet om, hvorvidt de anvender en række evalueringsskulpter i ledelsen af skolen, og i så fald, hvor længe de har gjort det. I spørgeskemaun-

dersøgelsen er spørgsmål fra Simon Calmar Andersens (2004)⁴⁰ undersøgelse af styringsmetoder i de danske folkeskoler inkluderet (disse spørgsmål er markeret med gråt). Dette muliggør i en vis udstrækning en form for sammenligning mellem delaspekter af skolernes evalueringskultur inden og efter indførelse af nationale test.

Tablet 3.1: Hvis din skole anvender nogle af følgende redskaber i ledelsen af skolen, hvor længe har de da været i anvendelse⁴¹? Angivet i pct.

Udsagn	Anvender ikke	Mindre end 1 år	1-2 år	3-5 år	6-10 år	Flere end 10 år	Ved ikke	I alt (N)
Undervisningsmiljøvurdering blandt elever	2	2	7	36	40	9	5	100 (728)
Opgørelse af fravær blandt elever	3	1	9	37	21	24	5	100 (728)
Opgørelse af fravær blandt lærere	4	2	8	35	22	24	6	100 (728)
Standardiserede test (fx fra Dansk Psykologisk Forlag/Hogrefe eller lignende)	5	2	5	18	28	35	8	100 (728)
Elevernes resultater i de nationale test	6	4	23	55	6	1	5	100 (728)
Skriftlige målsætninger for netop din skole (2004)	7	3	13	32	44*		1	100 (892)
Skriftlige målsætninger for netop din skole (2013)	15	3	11	28	23	13	7	100 (728)
					36			
Skriftlige evalueringer eller tilbagemeldinger om opnåede resultater (2004)	26	6	18	27	20*		2	100 (895)
Skriftlige evalueringer eller tilbagemeldinger om opnåede resultater (2013)	21	4	12	30	17	8	8	100 (728)
					25			
Skriftlig plan for indfrielse af skolens målsætninger	25	4	10	26	19	8	8	100 (728)
Andre evalueringsværktøjer som fx lærerlogbog og værdsættende samtale	25	3	12	26	17	6	12	100 (728)
Afgangselevernes karaktergennemsnit	28	2	6	23	17	17	9	100 (728)
Elevernes efterfølgende deltagelse i ungdomsuddannelse	33	2	9	22	13	6	15	100 (728)

*:I 2004-undersøgelsen hed svarkategorien blot "Flere end 5 år". I 2012/13-spørgeskemaundersøgelsen sondres mellem 6-10 og flere end 10 år.

Kilde: Spørgsmål 3 og 18, spørgeskemaundersøgelse til skoleledere, 2013

De første nationale test blev afviklet i 2010⁴². Det indebærer, at svarene '6-10 år' eller 'flere end 10 år' vil sandsynliggøre, at evalueringsredskabet blev taget i brug, *inden* de nationale test var en realitet. Således er anvendelsen af redskabet ikke påvirket af indførelsen af de nationale test. Svarene 'Mindre end 1 år' og '1-2 år' vil i evaluators analyse betyde, at redskabet først blev taget i brug efter de nationale test. Svaret '3-5 år' er taget som udtryk for, at skolernes benyttelse af

⁴⁰ The Impact of Public Management Reforms on Student Performance in Danish Schools, Public Administration, Vol. 86, No. 2, 2008 (s. 541-558), Simon Calmar Andersen, 2004.

⁴¹ Svarene er kontrolleret for, hvor længe skolelederen har været ansat. Der er en signifikant sammenhæng mellem ansættelsestid og svar. Skolelederne svarer primært ud fra den tid, de har været ansat. Skoleledere, der har været ansat i kort tid, svarer oftere ved ikke, men ikke oftere, at de ikke anvender redskabet.

⁴² Nogle test blev dog lanceret i 2007, om end den første fulde obligatoriske runde var i 2010.

evalueringsredskabet fandt sted omtrent samtidig med, at de nationale test blev indført. Evaluator skal gentage, at der ikke er tale om en kausal test i statistisk forstand, men svarene kan give en indikation af, om der er et tidssammenfald med ændringer i anvendelse af evalueringsredskaber og fremkomsten af de nationale test.

Tabellen ovenfor viser interessante forskelle i anvendelsen af redskaber blandt skoleledelsen.

Der er en gruppe redskaber, som i meget vid udstrækning anvendes af langt de fleste skoleledere. Således er der meget få skoler, der *ikke* anvender undervisningsmiljøvurderinger blandt elever (2 pct.), elev- (3 pct.) og lærerfravær (4 pct.) eller standardiserede test (5 pct.). En stor del af skolelederne angiver, at disse redskaber blev taget i brug, inden de nationale test blev indført, men der er samtidig over en tredjedel, der svarer, at det skete omtrent samtidig med indførelsen af de nationale test (3-5 år siden). Mange skoleledere angiver også, at de anvender de nationale test som redskab, og ikke så overraskende svarer hovedparten, at de begyndte med det samtidig med indførelsen af de nationale test. Over en fjerdedel er dog først begyndt at bruge de nationale test som redskab inden for de seneste 2 år. Da både undervisningsmiljøvurderinger samt nationale test er lovpligtige at benytte, siger anvendelsen af disse to instrumenter ikke voldsomt meget om skolernes evalueringskultur.

Der er en anden gruppe redskaber, som benyttes i langt mindre udstrækning end de ovennævnte. Disse redskaber vedrører skoleledernes anvendelse af skriftlige planer for indfrielse af skolens målsætninger, skriftlige evalueringer eller tilbagemeldinger om opnåede resultater eller skriftlige målsætninger for skolen. Disse redskaber anvendes *ikke* af henholdsvis 25, 21 og 15 pct. af skolelederne. Den største gruppe af skolelederne svarer, at de begyndte at bruge redskaberne for 3-5 år siden, men en del angiver også, at det skete *inden*, de nationale test blev indført. I relation til skoleledernes anvendelse af de to sidstnævnte redskaber har vi her mulighed for at sammenligne skoleledernes svar med den tidligere nævnte undersøgelse i 2004, foretaget af Simon Calmar Andersen, hvor der indgår identiske spørgsmålformuleringer.

Sammenligner man spørgsmålene fra den aktuelle undersøgelse og dem fra 2004, må det konstateres, at der ikke er sket markante ændringer. Der er dog sket en forøgelse af skoler, der *ikke* anvender skriftlige målsætninger for deres skole. Andelen af skoler, der ikke anvender skriftlige målsætninger for deres skole, er fordoblet siden undersøgelsen i 2004 fra 7 til 15 pct. På den anden side er der sket en reduktion i antallet af skoler, der *ikke* anvender skriftlige evalueringer på fem procentpoint fra 2004 til 2013. Man kan således forsigtigt konkludere, at der er sket en styrkelse i form af forøget anvendelse af skriftlige evalueringer, mens der er en udvikling væk fra benyttelsen af skriftlige målsætninger. Styrkelsen af anvendelsen af de skriftlige evalueringer er i en vis udstrækning sket, efter de nationale test blev indført.

Endelig er der en tredje gruppe evalueringsinstrumenter, der slet ikke benyttes af mellem en fjerdedel og tredjedel af skolerne. Således angiver skolelederne, at følgende redskaber *ikke* anvendes i ledelsen af skolen: elevernes overgangsfrekvens (33 pct.), elevernes afgangskarakterer (28 pct.) samt andre evalueringsværktøjer som fx lærerlogbog og værdsættende samtale (25 pct.).

Evalueringen viser, at skolelederne benytter en bred vifte af evalueringsværktøjer i deres ledelse af skolerne. Især de nationale test benyttes i vid udstrækning sammen med andre standardiserede test samt undervisningsmiljøvurderinger og fraværsopgørelser blandt både lærere og elever. I casestudierne fremhæves det tilsvarende af både skoleledere og lærere, at de nationale test blot er ét af mange evalueringsredskaber, og at testene derfor ikke kan stå alene (se afsnit 4.1).

Nedenstående tabel indeholder skoleledernes svar på en række udsagn, hvor de har skullet angive, hvor længe udsagnene kan siges at have afspejlet skolens generelle evalueringspraksis. Spørgsmålene er formuleret på baggrund af modellen for evalueringskultur og reflekterer, hvad der kan betegnes som stærk evalueringskultur. Nogle af spørgsmålene vedrører ledelsens vurdering af deres egen samt lærernes indstilling til evaluering, andre handler mere om ledelsens evne til at arbejde med evaluering, og endelig er der spørgsmål, der afdækker om evaluering bruges fremadrettet, fx til at formulere mål med.

Tabel 3.2: Hvis følgende udsagn kan siges at afspejle skolens generelle evalueringspraksis, hvor længe har det så været praksis? Angivet i pct.⁴³

Udsagn	Ikke en del af praksis	Mindre end 1 år	1-2 år	3-5 år	6-10 år	Flere end 10 år	Ved ikke	I alt (N)
Skoleledelsen ser evaluering som et vigtigt redskab til at understøtte den faglige kvalitet og udvikling	1	2	11	38	29	13	6	100 (728)
Lærergruppen ser evaluering som et vigtigt redskab til at understøtte den faglige kvalitet og udvikling	1	3	15	40	22	11	7	100 (728)
Skoleledelsen har et godt indblik i forskellige evalueringsværktøjer og deres styrker og svagheder	6	6	18	36	17	6	11	100 (728)
Skoleledelsen understøtter, at der er ressourceperson(er), der har et stort kendskab til evalueringsmetoder og deres anvendelighed	12	4	18	38	14	7	6	100 (728)
Skolen har klare målsætninger for arbejdet med evaluering	14	6	17	34	17	5	8	100 (728)
Skoleledelsen arbejder systematisk og løbende med at inddrage forskellige test- og evalueringresultater i en refleksiv dialog med lærergruppen med henblik på at opstille klare mål for skolens udvikling	15	8	22	30	13	6	6	100 (728)
Skoleledelsen benytter evalueringresultater til at sætte årlige mål	35	8	17	24	7	2	8	100 (728)

Kilde: Spørgeskemaundersøgelse til skoleledere, 2013.

Overordnet viser tabellen, at de fleste elementer af evalueringspraksis er udbredt på skolerne. Tabellen viser også, at det hyppigst forekommende svar er, at de pågældende praksisser blev etableret for 3-5 år siden, og derefter er det mest hyppige svar 1-2 år. Hvad angår skolernes *indstilling* til evaluering, angiver alle skolelederne (på nær 1⁴⁴pct.), at skoleledelsen såvel som lærergruppen ser evaluering som et vigtigt redskab til at understøtte den faglige kvalitet og udvikling med. Hovedparten af skolerne angiver, at denne positive indstilling til evaluering blandt skoleledere og lærergruppen har været til stede samtidig med eller efter, de nationale test kom til (hhv. 51 pct. og 55 pct.). En vis andel angiver dog, at denne indstilling har været praksis, inden de nationale test blev indført (henholdsvis 42 pct. og 33 pct. svarer 6 år eller flere).

Hvad angår skolernes *evne* til at arbejde med evaluering, svarer mange, at skoleledelsen har et godt indblik i forskellige evalueringsværktøjer (83 pct.). Ligeledes svarer forholdsvis mange, at skoleledelsen understøtter, at der er ressourcepersoner, der har et stort kendskab til evalueringsmetoder og deres anvendelighed (82 pct.). Begge dele er en forholdsvis nylig praksis (36-38 pct. svarer 3-5 år og 18 pct. svarer 1-2 år).

⁴³ Svarene er kontrolleret for, hvor længe skolelederen har været ansat. Der er en signifikant sammenhæng mellem ansættelsestid og svar. Skolelederne svarer primært ud fra den tid, de har været ansat, Skoleledere, der har været ansat i kort tid, svarer oftere ved ikke, men ikke oftere, at de ikke anvender redskabet.

⁴⁴ 7-8 pct., hvis man inkluderer 'Ved ikke' i svaret om, at de ikke anvender redskabet.

Der er især ét aspekt af evalueringspraksis, som ud fra ovenstående tabel ser ud til at være mangelfuld. Evalueringer bør, jf. definitionen af evalueringskultur, være baseret på *fastsættelse af mål* samt efterfølgende refleksioner og vurderinger og have et *fremadrettet* perspektiv. Der er dog over en tredjedel af skolelederne, der angiver, at det *ikke* er en del af praksis, at skoleledelsen benytter evalueringsresultater til at sætte årlige mål med (35 pct.)⁴⁵. 15 pct. angiver ligeledes, at det ikke er en del af skolens praksis, at skoleledelsen arbejder systematisk og løbende med at inddrage forskellige test- og evalueringsresultater i en refleksiv dialog med lærergruppen med henblik på at opstille klare mål for skolens udvikling. Endelig angiver 14 pct., at det ikke er en del af praksis, at skolen har klare målsætninger for arbejdet med evaluering. Disse resultater stemmer overens med resultaterne i tabel 3.1, der viste, at en del skoleledere ikke anvendte skriftlige målsætninger for netop deres skole, skriftlige evalueringer eller tilbagemeldinger om opnåede resultater eller en skriftlig plan for indfrielse af skolens målsætninger. Der henvises også til analyser heraf i et styringsperspektiv i kapitel 6.

Ovenstående resultater udfordrer til dels en central del af evalueringens forandringsteori. I forandringsteorien anføres, at ledelsen skal opstille konkrete faglige mål på baggrund af testresultater og øvrig evaluering, formulere pædagogiske indsatsområder på baggrund af resultaterne og løbende revidere de pædagogiske mål for skolens udvikling. Sker dette ikke, vil der være en risiko for, at den tiltænkte styrkelse af skolens evalueringskultur og evalueringskapacitet ikke finder sted. Det ser dog ud til, at der er en positiv udvikling i gang, hvor flere skoler i de senere år benytter evalueringsresultater til at sætte mål med. Kun 9 pct. svarer, at skoleledelsen for 6 år eller flere år siden benyttede evalueringsresultater til at sætte årlige mål med. En fjerdedel angiver, at det har været en del af praksis for 3-5 år siden, og knap en femtedel angiver, at det påbegyndtes for 1-2 år siden. Det samme mønster ses ved de andre spørgsmål.

Evalueringsens kvalitative data giver et indblik i nogle af de barrierer, der er i forhold til at anvende evalueringsresultater fra de nationale test til, på skoleniveau, at sætte eller justere mål og indsatser. Ud fra casestudierne og svar fra ledere og lærere i de åbne svarkategorier i spørgeskemaerne er det tydeligt, at de nationale test på nogle skoler rummer en potentiel interessekonflikt mellem ledere og lærere. Sammenlignet med mange andre evalueringsredskaber giver de nationale test mulighed for at sammenligne på tværs af klasser og over tid. En skoleleder forklarer det på følgende måde: "Det store dokumentationsmateriale giver mulighed for at evaluere både elever, men også lærernes undervisning på en måde, som det ikke tidligere har været tilfældet" (spørgeskemaundersøgelsen til skoleledere, 2013). Dét, at de nationale test indeholder sammenlignelige resultater, er med til at give en stor gruppe lærere et indtryk af, at testene er en kontrolforanstaltning, jf. kapitel 6.

Flere lærere, som er interviewet i forbindelse med casestudierne, fortæller, at de efter de nationale tests indførelse ansvarliggøres for deres elevers resultater af forældre, skoleledelse, forvaltning og offentligheden generelt. Spørger man til, hvordan de holdes til ansvar for resultaterne, kan de fleste dog ikke komme med konkrete eksempler på, hvordan dette sker. Casestudierne viser eksempelvis, at det ikke er alle skoleledere, der direkte har en dialog med lærere om resultaterne af de nationale test. Det er heller ikke alle skoleledere, som bruger evalueringsresultaterne til vurdering og eventuelt justering af målene for skolen som helhed såvel som for den enkelte lærer, årgang eller klasse. Således er der i nogle tilfælde modstrid mellem lærernes fornemmelse af kontrol og den praksis, der foregår på skolen.

Ovenstående resultater stemmer overens med resultaterne af et forskningsprojekt, der fandt frem til, at en relativ stor andel af danske lærere mener, at evaluering kan inspirere til udvikling af en bedre undervisning samt kan anvendes til at adressere seriøse problemer på skolen, samtidig med at de mener, at vedvarende evaluering udviser mangel på tillid til lærernes arbejde⁴⁶. Dette resultat kan forekomme paradoksalt, idet lærerne på én og samme tid har en positiv såvel som negativ attitude over for evaluering. Svarene kan tages som udtryk for, at lærerne støtter, at man som lærer kan anvende evaluering pædagogisk, men at de ser med bekymring på, at

⁴⁵ Denne tendens findes ligeledes i andre publikationer som fx Ledelse af folkeskolerne, SFI, 2011.

⁴⁶ Chp.10: Teachers' perceptions of quality assurance and evaluation" in Fabricating Quality in Education. Data and governance in Europe, Ozga, Jenny, Peter Dahler-Larsen, Christina Segerholm og Hannu Simola, 2011.

skoleledelse, forvaltning eller andre bruger evalueringresultater til at vurdere og kontrollere læreres indsats med.

Resultaterne viser, at de fleste af de evalueringsskisser, som evaluator har defineret som en del af en stærk evalueringsskulturer – med få undtagelser – er udbredt på de deltagende skoler. Der er dog ikke tale om en meget markant udvikling, hvor samtlige skoler fra den ene dag til den anden anfører, at de efter indførelsen af nationale test omlagde deres evalueringsskisser. Men for flertallet af skolerne er ændringerne sket inden for de sidste 5 år. Dette resultat bekræftes ligeledes i casestudierne, hvor det samstemmende anføres, at ændringerne i evalueringsskulturer har været undervejs i nogle år, men særligt har fundet sted de sidste 3-5 år. På denne baggrund er det evaluators samlede vurdering (på tværs af datakilder), at skolernes evalueringsskulturer er i en stadig bevægelse mod at blive styrket, og at denne styrkelse falder tidsmæssigt sammen med indførelsen af de nationale test (og projektet til fremme af evalueringsskulturer i folkeskolen). Der er dog tilsyneladende en udfordring i forhold til at bruge testresultater fremadrettet til at sætte eller justere skolens faglige eller pædagogiske mål og indsatser med.

Analyse 2: Den oplevede effekt af de nationale tests bidrag til at styrke skolernes evalueringsskulturer

Nedenfor ser evaluator nærmere på, i hvilket omfang skolerne selv – skoleledere og lærere – samt forvaltninger vurderer, at de nationale test har styrket evalueringsskulturer i den danske folkeskole.

Tabel 3.3: Styrker de nationale test skolens evalueringsskulturer? Angivet i pct.

	Slet ikke	I mindre grad	I nogen grad	I temmelig høj grad	I meget høj grad	Total
Skoleledere	5	19	47	23	6	100 (728)
Lærere	21	35	36	8	1	100 (2162)

Kilde: Spørgeskemaundersøgelse til skoleledere og lærere, 2013.

Tabellen viser, hvordan skoleledere og lærernes svar fordeler sig i forhold til deres vurdering af, hvorvidt de nationale test styrker skolernes evalueringsskulturer. Det hyppigste svar for både skoleledere og lærerne er, at de nationale test i nogen grad har styrket skolens evalueringsskulturer (47 pct. af skoleledere og 36 pct. af lærerne). Ser man nærmere på de to gruppers svarfordelinger, er det dog tydeligt, at skoleledere vurderer effekten af nationale test for skolernes evalueringsskulturer markant mere positivt end lærerne. 29 pct. af skoleledere angiver, at de nationale test i temmelig høj eller i meget høj grad har styrket skolens evalueringsskulturer. Den tilsvarende andel af lærere er kun 9 pct. 55 pct. af lærerne angiver, at de nationale test slet ikke eller i mindre grad har styrket skolens evalueringsskulturer, mens det tilsvarende tal for skoleledere kun er 24 pct.

Ovenstående billede stemmer godt overens med data, indhentet i forbindelse med casestudier på en række skoler. Størstedelen af skoleledere på de besøgte skoler, mener, at deres skolers evalueringsskulturer er styrket efter indførelsen af de nationale test. Der er dog en del skoler, der mener, at evalueringsskulturer også ville være blevet styrket i fraværet af test. Flere skoler nævner, at det er en tendens i tiden, hvor der fokuseres på at måle og vurdere med PISA-undersøgelserne som afsæt for dette fokus. Således er evaluering i langt højere grad blevet italesat end tidligere, hvilket har afstedkommet både statslige og lokale initiativer.

Nedenstående tabel viser, hvordan forvaltningerne har svaret på et enslydende spørgsmål om sammenhæng mellem nationale test og evalueringskultur.

Tabel 3.4: De nationale test har generelt medvirket til at styrke evalueringskulturen på kommunens skoler. Angivet i pct.

Helt uenig	Overvejende uenig	Hverken enig eller uenig	Overvejende enig	Helt enig	Ved ikke	Total (N)
1	6	22	50	14	6	100 (78)

Kilde: Spørgeskemaundersøgelse til kommunale forvaltninger, 2013.

Sammenligner man skolernes svar med forvaltningernes, fremgår det, at forvaltningerne er endnu mere positive end skolelederne. Således svarer hele 64 pct. af de adspurgte repræsentanter fra forvaltningerne, at de er overvejende eller helt enige i, at de nationale test har styrket evalueringskulturen på kommunens skoler. 22 pct. er hverken enige eller uenige, og kun 7 pct. er overvejende eller helt uenige i, at de nationale test har medvirket til at styrke evalueringskulturen på kommunens skoler. Dette billede er i overensstemmelse med det billede, som forvaltningerne gav i casestudierne.

Baseret på ovenstående er der altså en klar gruppering af de forskellige respondenters svar i forhold til vurderingen af, om de nationale test har styrket skolernes evalueringskultur eller ej. Jo længere væk fra undervisningen, man kommer, jo mere positivt vurderes de nationale tests betydning at være for evalueringskulturen. Således er forvaltningerne meget positive i deres vurdering, skolelederne er positive og lærerne mere skeptiske. Samlet set er der således tale om et blandet billede fra de aktører, der anvender de nationale test i dagligdagen. Særligt kredsen af lærere med erfaringer inden for testfag har en blandet vurdering af testenes betydning for skolernes evalueringskultur – endog med en klar tendens til, at størstedelen finder testenes effekt begrænset.

Dette resultat kan læses på flere måder: For det første kan det være udtryk for et øjebliksbillede og samtidig et udtryk for, at testene af mange lærere endnu ikke bruges på et niveau, der betegner en stærk evalueringskultur (den fremadrettede og formative anvendelse), jf. kapitel 4. For det andet kan svarene være udtryk for, at formålet med de nationale test aldrig har været, at de nationale test *alene* skulle styrke skolernes evalueringskultur, men at de skulle gøre det i sammenhæng med andre redskaber. Således vurderer evaluator samlet, at data fra spørgeskemaundersøgelserne *i nogen grad* peger på en positiv sammenhæng mellem nationale test og udvikling af evalueringskultur.

Det bemærkes endvidere, at skolelederne og lærerne i spørgeskemaundersøgelsen samstemmende fremhæver, at de nationale test først og fremmest har medført, at skolen er blevet bedre til at dokumentere og formidle resultaterne af evalueringerne. På den anden side er både skoleledere og lærere mindst tilbøjelige til at erklære sig enige i, at de nationale test har fremmet, at skolen foretager systematiske og løbende evalueringer af elevernes udbytte af undervisningen.

Tabel 3.5: Hvordan har de nationale test været med til at styrke jeres skoles evalueringskultur? Vi er blevet bedre til ... Angivet i pct.

	Slet ikke	I mindre grad	I nogen grad	I temmelig høj grad	I meget høj grad	Total (N)
... at foretage systematiske og løbende evalueringer af elevernes udbytte af undervisningen	1	13	57	24	4	100 (550)
... at dokumentere og formidle resultaterne af evalueringerne	0 (0,4)	10	54	32	3	100 (550)
... at anvende evaluering fremadrettet med henblik på at tilrettelægge en undervisning, der tager udgangspunkt i den enkelte elev	1	14	56	26	3	100 (550)

Kilde: Spørgeskemaundersøgelse til skoleledere, 2013.

Bemærk: Spørgsmålet er kun stillet til de skoleledere, der har svaret bekræftende på, om de nationale test har medvirket til at styrke skolens evalueringskultur.

For de tre spørgsmål angiver flest skoleledere, at de nationale test har medført, at skolen er blevet bedre til at dokumentere og formidle resultaterne af evalueringerne (36 pct. svarede i temmelig eller i meget høj grad). Det hyppigst angivne svar er dog 'I nogen grad' til alle tre spørgsmål (54-57 pct.).

Tabel 3.6: Hvordan har de nationale test været med til at styrke evalueringskulturen. Angivet i pct.

De nationale test har...	Slet ikke	I mindre grad	I nogen grad	I temmelig høj grad	I meget høj grad	Total
... medvirket til at fremme, at vi på skolen foretager systematiske og løbende evalueringer af elevernes udbytte af undervisningen	26	32	34	8	1	100 (2162)
... medvirket til at fremme, at vi på skolen dokumenterer og formidler resultaterne af evalueringer	20	30	37	12	2	100 (2162)
... styrket mit fokus på elevernes faglige resultater	21	29	36	12	2	100 (2162)

Kilde: Spørgeskemaundersøgelse til lærere, 2013.

Bemærk: Alle de deltagende lærere er blevet stillet spørgsmålet.

Lærernes svar er noget mindre positive end skoleledernes. En relativ stor andel af lærerne angiver således for hvert af de tre spørgsmål, at de nationale test ikke har gjort en positiv forskel.

Casestudierne nuancerer skolernes forklaringer på, hvordan de nationale test har styrket skolerne evalueringskultur. I den forbindelse angiver flere skoler, at de nationale test har understøttet og forstærket det fokus, der allerede var på evaluering. Skolelederne understreger, at selvom evaluering også var på dagsordenen tidligere, har de nationale test medført en bevægelse fra en evalueringskultur, der baserer sig på mavefornemmelser og "synsninger", til i dag at bygge på resultater. Endvidere nævnes det, at man med testene har flyttet fokus fra en evaluering af, om undervisningsforløbet er godt, til et fokus på om læringsudbyttet er godt. Casestudierne viser, at de nationale test har medført et større fokus på faglige resultater. På denne vis har testene bidraget til en styrket evalueringskultur.

En samlet vurdering af de nationale tests betydning for skolernes evalueringskultur

Evaluators vurderer overordnet, at der i de senere år er sket en styrkelse af skolernes evalueringskultur. Denne vurdering er baseret på sekundære kilder (undersøgelser mv.) samt evaluators egne data.

Evaluators vurderer samlet set, at ovenstående analyser i et vist omfang understøtter **hypotesen** om, at de nationale test har medvirket til at styrke skolernes evalueringskultur. Indførelsen af de nationale test har – sammen med en række andre evalueringsredskaber – været med til at cementere det ønske, der politisk har været om at styrke arbejdet med evaluering i skolen. De nationale tests effekt skal ses i sammenhæng med et generelt fokus på resultater og evaluering. Modsat andre evalueringsredskaber såsom standardiserede test, er de nationale test obligatoriske, og det er med til at fastholde et kontinuerligt fokus på den løbende evaluering. Der er dog interessante forskelle i de oplevede effekter af de nationale test. Forvaltninger og skoleledere er meget positive, mens lærere er mere skeptiske over for, om de nationale har styrket skolernes evalueringskultur. Denne skepsis kan hænge sammen med nogle læreres oplevelse af, at de nationale test er et kontrolinstrument, som kan bruges til at vurdere lærerens og ikke kun elevens indsats med.

Evaluators vurderer dog, at evalueringskulturen på enkelte punkter stadig er mangelfuld. Dette relaterer sig især til at anvende evaluering som udgangspunkt for målfastsættelse samt anvendelse af evalueringsresultater fremadrettet. Der henvises til kapitel 4 og 6 for yderligere analyser.

Som supplement hertil vil evaluators analysere, om den oplevede sammenhæng kan tilskrives andre faktorer end de nationale test. En metodisk udfordring i denne sammenhæng er, at der i den danske folkeskole er indført en række evalueringsredskaber (nationalt og lokalt), der alle har som overordnet formål at styrke evalueringskulturen. Således er det ikke muligt at isolere, hvilken effekt hvert redskab hver især har bidraget med i forhold til styrkelsen af evalueringskulturen. Det er dog evaluators vurdering, at de nationale test har haft en selvstændig positiv effekt på skolernes evalueringskultur. Casestudierne gav ikke anledning til at vurdere, at andre instrumenter som elevplaner, obligatoriske afgangseksamener eller skolernes egne kvalitetsrapporter, i sig selv har påvirket skolernes evalueringskultur i nogen nævneværdig grad. Ligeledes undersøgte EVA i 2011 kvalitetsrapportens betydning for praksis på skoler og i kommuner. De konkluderer, at kvalitetsrapporten kan styrke evalueringskulturen, men at kvalitetsrapporterne stadig kun anvendes i begrænset omfang⁴⁷.

Et andet forhold, som kan have haft betydning for styrkelsen af skolernes evalueringskultur, er det generelle fokus på faglige resultater og elevernes udbytte af undervisningen. Således kunne man argumentere for, at skolerne i fraværet af nationale test alligevel ville have øget andre former for evaluering for at kunne følge og forbedre de faglige resultater. Dette kan ikke udelukkes, men det er evaluators vurdering, at de nationale test har bidraget til at forstærke det fokus, der allerede var på faglige resultater. Med andre ord har de nationale test haft en understøttende (og dagsordenssættende) funktion, der på én og samme tid har understreget vigtigheden af at fokusere på elevernes udbytte af undervisningen samt nødvendigheden af at evaluere for at opnå, såvel som at fastholde, gode resultater.

Det er endvidere veldokumenteret, at skoleledelsen har enten direkte eller indirekte betydning for elevernes faglige resultater (se fx Skolerådets beretning 2012). Mange undersøgelser understreger ligeledes ledelsens betydning i forhold til at skabe et stærkt fagligt miljø⁴⁸. Det er evaluators vurdering, at en stærk evalueringskultur er en central bestanddel af et sådant fagligt miljø. Således kan en velbegrunnet hypotese være, at en stærk skoleledelse har en positiv betydning for skolers evalueringskultur. For at få en indikation af denne sammenhæng er det undersøgt, hvorvidt de skoleledere, der angiver, at de prioriterer, at eleverne får gode faglige resultater, er mere tilbøjelige til at mene, at de nationale test har styrket skolens evalueringskultur.

⁴⁷ Delrapport 3: Kvalitetsrapporten – undersøgelse af kvalitetsrapportens betydning for praksis på skoler og i kommuner. Danmarks Evalueringsinstitut, 2011.

⁴⁸ Ledelse, læring og trivsel i folkeskolerne, SFI, 2011.

Tabel 3.7: Sammenhæng mellem vægtning af elevernes faglige resultater og vurdering af nationale tests betydning for skolernes evalueringskultur. Angivet i pct.

Vurdering af i hvilken grad nationale test har styrket evalueringskultur						
Vægter elevernes faglige resultater	Slet ikke	I mindre grad	I nogen grad	I temmelig høj grad	I meget høj grad	Total
Slet ikke	0	0	0	0	0	0
I mindre grad	0	33	33	33	0	100 (3)
I nogen grad	7	29	38	26	0	100 (73)
I temmelig høj grad	4	20	50	22	4	100 (405)
I meget høj grad	6	16	45	24	10	100 (247)

Kilde: Spørgeskemaundersøgelse til skoleledere, 2013.

Der er en svag, men signifikant sammenhæng mellem, hvordan skoleledelsen har angivet, at de prioriterer, at eleverne får gode faglige resultater, og hvorvidt de mener, de nationale test har gavnet evalueringskulturen. Jo stærkere, man prioriterer elevernes faglige resultater, jo større tendens er der til, at man har en positiv vurdering af, om de nationale test har styrket skolens evalueringskultur.

Casestudierne viser tilsvarende, at de skoler, hvor man vægter elevernes faglige resultater højt, også er mest positivt stemt over for evaluering generelt, herunder især nationale test. Således kan man betragte et stærkt fagligt fokus mere som en fremmede faktor end som en konkurrerende forklaring, idet anvendelse af nationale test både kan påvirke det faglige fokus og blive påvirket af det. I kapitel 6 vil vi ligeledes fokusere på skoleledelsen.

Afsluttende vil evaluatoren pointere et væsentligt forhold omkring årsager til forskelle i skolers evalueringskultur. Lovbemærkningerne er som nævnt meget tydelige omkring, at det ikke er de nationale test *alene*, der skal skabe styrket evalueringskultur og stærkere faglighed. Der er dels andre virkemidler i spil nationalt (elevplaner, kvalitetsrapporter, obligatoriske afgangsprøver), dels skulle evaluering i skolerne også trække på andre værktøjer end nationale test (standardiserede test, logbøger mv.). I forlængelse heraf er det evaluators vurdering, at de ovenstående alternative forklaringer sammen med de nationale test skal ses som værende medvirkende til/understøttende for en stærkere evalueringskultur. Det er således ikke de nationale test alene, der skaber en stærkere evalueringskultur. Det er helheden i anvendelsen af nationale test i samspil med en række evalueringsværktøjer samt en stærk ledelse, der understøtter en stærk evalueringskultur. Etablering og videreudvikling af en evalueringskultur er en kompleks størrelse, og der er ikke nogen entydig sammenhæng mellem enkeltfaktorer – som nationale test – og skolers evalueringskultur.

Det bemærkes, at dette afsnit indeholder en mere overordnet vurdering af skolernes evalueringskultur. I kapitel 4, 5 og 6 undersøger evaluatoren skolernes evalueringspraksis nærmere, herunder hvordan de nationale test anvendes i folkeskolen. Herved tilvejebringes en mere nuanceret forståelse af status for, hvordan de nationale test anvendes i den danske folkeskole, og hvordan anvendelsen fremadrettet kan styrkes med henblik på en endnu stærkere evalueringskultur.

3.3 Er der sammenhæng mellem skolernes evalueringskultur og elevernes faglige resultater?

Som supplement til de to foregående effektanalyser har evaluatoren som led i vores test af evalueringens forandringsteori undersøgt, om elevernes styrkede faglige niveau understøttes af en stærkere evalueringskultur i folkeskolen. Hermed undersøges, om graden af evalueringskultur på en skole har betydning for elevernes faglige niveau. Evaluatoren har på den baggrund formuleret en hypotese om, at en stærk evalueringskultur i folkeskolen styrker elevernes faglige niveau.

I de følgende afsnit vil denne hypotese blive afprøvet. Denne analyse skal dog ses som et supplement til de to foregående analyser. For at undersøge denne sammenhæng vil der blive foretaget en statistisk analyse af den forventede sammenhæng, jf. nedenfor. Disse statistiske analyser vil blive suppleret af erfaringer fra casestudier samt evaluators vurdering på baggrund af de tilgængelige datakilder.

3.3.1 Fremgangsmåde for dannelse af mål for skolernes evalueringskultur

For at undersøge, hvorvidt der er en statistisk sammenhæng mellem skolernes evalueringskultur og deres faglige resultater, er det nødvendigt at finde et velegnet kvantificerbart mål for evalueringskultur. For at skabe et sådant mål har vi dannet indeks. I det følgende præsenteres kort fremgangsmåden.

Som beskrevet i kapitel 2 har evaluator operationaliseret begrebet evalueringskultur ved at udvikle en model for evalueringskultur. Modellen er anvendt i arbejdet med at formulere items til spørgeskemaundersøgelsen målrettet alle tre målgrupper (lærere, skoleledere og forvaltning) for at sikre, at alle relevante aspekter af evalueringskultur afdækkes. I dannelsen af et statistisk mål for evalueringskultur har vi taget udgangspunkt i skoleledernes besvarelse⁴⁹. Statistiske analyser af skoleledernes svar viser, at der er tale om flere selvstændige dimensioner i svarene, og at man derfor ikke kan slå alle relevante spørgsmål sammen til ét indeks for evalueringskultur. På baggrund af disse analyser har vi valgt at etablere tre indeks⁵⁰.

- Det første indeks repræsenterer **skoleledelsens anvendelse af de nationale test**: Hvilke resultater anvender de, hvordan anvender de dem osv.
- Det andet indeks illustrerer **skolens generelle evalueringspraksis**: Hvad er indstillingen til evaluering på skolen, bruges evalueringresultater systematisk til vurdering og målfastsættelse etc.
- Det tredje indeks opsummerer skolelederens vurdering af **lærernes anvendelse af de nationale test**: Hvordan anvender lærerne de nationale test, hvordan samarbejder de omkring testresultaterne, og inddrager de dem i skole-hjem-samarbejdet?

Skolernes placering på indeks for evalueringskultur

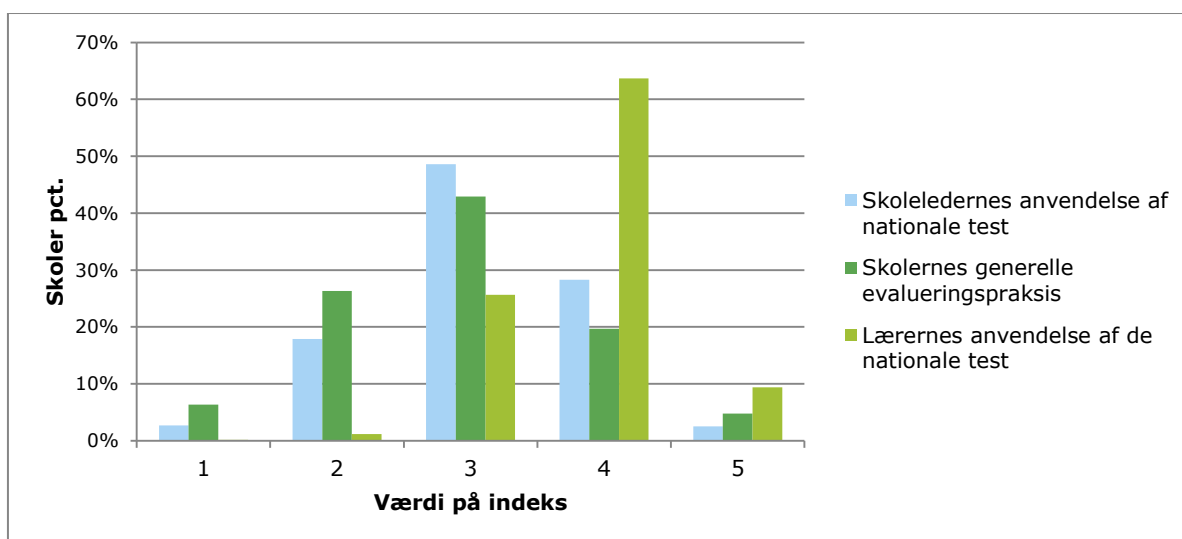
Nedenfor præsenteres, hvordan skolerne fordeler sig i relation til de tre indeks. Skoleledernes svar er summeret, således at de kan indplaceres på værdien 1-5, hvor 5 repræsenterer fx den stærkeste generelle evalueringspraksis (eller fremadrettede anvendelse af nationale testresultater) og 1 repræsenterer den svageste.

For at score højt på indekset omkring skolernes generelle evalueringspraksis, skal skolelederen i flere tilfælde have svaret 'I meget høj grad' eller 'I høj grad' til en række spørgsmål, fx hvorvidt skoleledelsen ser evaluering som et vigtigt redskab til at understøtte den faglige kvalitet og udvikling eller til, at skolen har klare målsætninger for arbejdet med evaluering. Svarer skolelederen derimod 'Slet ikke' eller 'I mindre grad' til de nævnte spørgsmål, vil skolen blive vurderet til at have en svagere generel evalueringspraksis.

⁴⁹ Fordelen ved at benytte skolelederens svar er, at vi spørger alle landets folkeskoleledere og således kan forvente en højere grad af repræsentativitet end hos lærerne, hvor vi spørger et udvalg af lærerne på skolerne.

⁵⁰ For en nærmere beskrivelse af fremgangsmåde, se kapitel 2 og metodebilaget.

Figur 3.6: Skolernes fordeling på tre indeks om evalueringskultur



Overordnet viser figuren, at skolernes fordeling på de tre indeks er relativ ens, men med enkelte undtagelser.

For så vidt angår det første indeks – skoleledernes fremadrettede anvendelse af de nationale test – så viser figuren, at halvdelen af skolerne fordeler sig i en midtergruppe, mens godt en fjerdedel kan siges at repræsentere den relativt stærkere gruppe, og knap en femtedel den relativt svage gruppe. Der er få skoler, der repræsenterer den stærkeste og svageste anvendelse (3 pct.). Der er således en tendens til, at der er flere skoler med en relativ stærk end med en relativ svag anvendelse af resultaterne af de nationale test.

Hvad angår det andet indeks – skolernes generelle evalueringspraksis – så fremgår det, at lidt færre skoler i forhold til det første indeks er i midtergruppen. Knap en fjerdedel er i den relativt svagere gruppe, mens en femtedel er i den relativt stærkere gruppe. Som ved indeks 1 er der få skoler i den meget svage (6 pct.) og meget stærke gruppe (5 pct.). Der er altså relativt færre skoler, der placerer sig i gruppen med den stærke evalueringskultur, end der er skoler, der placerer sig i gruppen med den stærke anvendelse af de nationale tests resultater.

Fordelingen på det tredje indeks – lærernes anvendelse af de nationale test – skiller sig ud fra de to andre ved at have en højere modalværdi. Hermed menes, at der er en stor gruppe, der ikke placerer sig i midten, men til højre for midten. Således er kun en fjerdedel af skolerne i midtergruppen, mens næsten to tredjedele af skolerne er i den relativt stærkere gruppe. Det illustrerer, at skolelederne har en forholdsvis positiv vurdering af lærernes anvendelse af de nationale test (i kapitel 4, 5 og 6 vil analysen dog blive nuanceret i forhold til, hvorvidt lærernes anvendelse af de nationale test er så stærk som vurderet af skolelederne).

Ovenstående analyse viser, at flertallet af danske skoler placerer sig i midten af et kontinuum gående fra meget svag til meget stærk evalueringskultur. Selvom der er tale om tre selvstændige dimensioner af skolernes samlede evalueringskultur, er der alligevel en positiv statistisk signifikant sammenhæng mellem indeksene, men den er ikke særlig stærk. Der er en tendens til, at jo stærkere generel evalueringspraksis en skole har, jo stærkere er skolelederens anvendelse af resultaterne af de nationale test. Ligeledes, jo stærkere en skoleleders anvendelse af resultaterne af de nationale test er, jo stærkere angiver de, at lærernes brug af de nationale test er. Den svageste sammenhæng finder man mellem skolens generelle evalueringspraksis og lærernes anvendelse af de nationale test.

Indeksene giver dog kun mulighed for at vurdere fordelingen af skolerne og deres relative styrke, og som nævnt er det ikke muligt at analysere resultaterne *over tid*, da der ikke tidligere er lavet lignende repræsentative undersøgelser af evalueringskulturen på de danske folkeskoler. Derudover har vi udelukkende benyttet data fra spørgeskemaundersøgelsen til skolelederne til at belyse skolernes evalueringskultur med, hvorfor lærernes perspektiv ikke er direkte repræsenteret. I

kapitel 4 vil vi mere indgående belyse, hvordan lærere og skoleledere anvender de nationale test som en del af deres samlede arbejde med evaluering. Dette vil nuancere det overordnede billede samt give indblik i, om der er områder, hvor anvendelsen af testene er særlig stærk eller omvendt mindre stærk.

3.3.2 Sammenhæng mellem evalueringskultur og faglige resultater

Evaluator har statistisk undersøgt sammenhængen mellem på den ene side de tre mål for evalueringskultur, der er beskrevet ovenfor, og på den anden side resultaterne af de fire nationale test, der gennemføres i 8. klasse (i 2012): Dansk, læsning, fysik/kemi, biologi og geografi.

Undersøgelsen mellem de tre mål for evalueringskultur og hver af de fire nationale test resulterer i 12 analyser. I de 11 af analyserne findes ikke nogen statistisk signifikant sammenhæng. I ét tilfælde findes en sammenhæng, nemlig mellem skolernes generelle evalueringspraksis og resultaterne i dansk, læsning.

Boks 3.2: Alternative forklaringer til årsagssammenhængen mellem evalueringskultur og faglige resultater

Evaluator har undersøgt, om denne ene sammenhæng kunne skyldes en række andre faktorer, der potentielt også kunne kendetegne skoler med en høj evalueringskultur. Således har vi kontrolleret sammenhængen for, om skolerne anvender lektiecafe (el.lign.), hvor stor andel af eleverne, der i givet fald anvender lektiecafeen, om skolerne anvender bestemte lærebogssystemer i dansk, læsning eller matematik, om skoleledelsen aktivt understøtter lærernes professionelle udvikling, og om skoleledelsen mere generelt prioriterer pædagogisk, strategisk eller administrativ ledelse. Resultaterne viser, at sammenhængen mellem den anden dimension af evalueringskultur og de nationale testresultaterne i dansk, læsning i 8. klasse forbliver signifikant. Det er altså ikke ovenstående konkurrerende forhold, der forklarer sammenhængen mellem en stærk evalueringskultur og høje faglige resultater i dansk, læsning.

Resultatet kan ikke siges at være særlig robust, idet der kun findes sammenhæng i 1 ud af 12 analyser. Der var heller ikke på forhånd nogle teoretisk baserede forventninger om at finde en sammenhæng mellem netop den dimension af evalueringskultur (generel evalueringspraksis) og lige præcis testen i dansk, læsning. Den på forhånd formulerede hypotese var, at en stærk evalueringskultur i folkeskolen styrker elevernes generelle faglige niveau. Såfremt der var en reel sammenhæng, burde der således findes en sammenhæng med flere af fagene. Ligeledes var forventningen, at de andre indeks omkring skoleledernes og lærernes anvendelse af nationale test også burde have en sammenhæng med faglige resultater. Således er det evaluators vurdering, at den ene fundne sammenhæng mest sandsynligt er udtryk for en tilfældighed.⁵¹

De statistiske analyser giver altså ikke umiddelbart støtte til, at der er en sammenhæng mellem skolers evalueringskultur og de faglige resultater blandt skolens elever. Resultaterne kan dog ikke tages som et endegyldigt bevis på et fravær af sammenhæng mellem skolernes evalueringskultur, og hvordan skolens elever klarer sig fagligt. De gennemførte analyser kan blot ikke bekræfte, at der er en statistisk sammenhæng.

Der kan være flere årsager til dette. For det første kan det handle om de muligheder, der er for at måle evalueringskultur. For at få et samlet mål for en skoles evalueringskultur har vi benyttet skoleledernes svar i en spørgeskemaundersøgelse og derefter dannet flere indeks, der belyser forskellige dimensioner af dette fænomen. Det kan muligvis handle om, at evaluering er en for individuel praksis til, at man kan danne et retvisende mål på skoleniveau, eller at skolelederen måske er for langt fra praksis til at kunne repræsentere skolens evalueringskultur. Således har SFI konkluderet, at evaluering gennem test har en positiv betydning for elevernes faglige præstationer. Jo flere test eleverne udsættes for, jo bedre klarer de sig. Vores analyse er dog foretaget på samlet skoleniveau i modsætning til fornævnte analyse. Således kan individuelle forskelle

⁵¹ Som nævnt ovenfor anvender vi et konventionelt signifikansniveau på 5 pct.. Det betyder, at man accepterer, at man i hver 20. analyse (5 pct.) kan ske at finde en signifikant sammenhæng, som i virkeligheden er en tilfældighed. Det skal man være særligt opmærksom på, når man som her gennemfører mange analyser på én gang og kun finder en sammenhæng i et enkelt tilfælde.

mellem lærere på en skole blive udvisket.⁵² Derudover er skolernes evalueringskultur selvfølgelig andet og mere end blot anvendelsen af test.

En anden forklaring kan handle om, at de bedst præsterende skoler i forvejen havde en stærk evalueringskultur og således ikke har haft en effekt af at anvende de nationale test. I supplerende analyser findes således en tendens til, at effekten af de nationale test er signifikant mindre for de skoler, der allerede inden testene blev indført, var blandt de 10 pct. bedst præsterende skoler (målt på afgangsprøvekaraktererne i 2009). Dette kan tolkes som, at disse skoler allerede på forhånd har tilstrækkelig god evalueringspraksis, så de nationale test ikke tilfører dem yderligere forbedring. Dette er dog en tolkning, som ikke direkte kan ses ud af de statistiske data. En anden mulighed kan være, at der er en vis loft-effekt i de nationale test i den forstand, at testen har vanskeligt ved at måle forskelle inden for gruppen af meget dygtige elever.

Den ovenstående analyse udgør et øjebliksbillede af sammenhængen mellem skolernes evalueringskultur og elevernes faglige resultater. Men kan forestille sig, at resultaterne havde været anderledes, såfremt man havde set på *udviklingen* både i evalueringskultur og faglige resultater. Hypotetisk er det muligt, at de skoler, der har oplevet den største styrkelse af evalueringskultur, er dem, der har opnået de største forbedringer i elevernes faglige resultater. Derudover er det en svaghed ved analysen, at vi måler evalueringskultur *efter* (december-januar 2012/2013), at vi måler de faglige resultater (forår 2012). Dette kan have haft en negativ betydning for at finde en sammenhæng, idet et centralt kriterium for kausalitet – tidsrækkefølge – er brudt. Det har dog som tidligere nævnt ikke været muligt at benytte nyere data for faglige resultater.

Den komplekse sammenhæng mellem evalueringskultur og faglige resultater har også været tydelig i de kvalitative casestudier. I denne forbindelse besøgte vi både skoler med en stærk evalueringskultur og gode faglige resultater, samt skoler med en svagere evalueringskultur og gode faglige resultater. Ligeledes besøgte vi skoler med en stærk evalueringskultur og dårlige faglige resultater, og skoler med en svag evalueringskultur og dårlige resultater. Det kan på den baggrund konstateres, at der ikke er en entydig sammenhæng mellem styrken af evalueringskultur og skolers faglige resultater. En stærk evalueringskultur er under ingen omstændigheder en tilstrækkelig betingelse for, at en skoles elever klarer sig godt. Dertil er sammenhængen for kompleks, og der er andre faktorer i spil. De skoler, der havde oplevet en positiv udvikling i elevernes faglige resultater, tillagde de nationale test såvel som en styrket evalueringskultur en vis betydning. Derudover fremhævede de vigtigheden af en tydelig ledelse, der klart prioriterer en høj faglighed som et centralt mål for skolen. Ligeledes oplevede de en stærk faglig sparring i form af velfungerende fagteammøder og klasseteammøder som befordrende for gode resultater. Endelig er det et tilbagevendende argument fra både godt og mindre godt præsterende skoler, at de faglige resultater i høj grad hænger sammen med skolens elevers socioøkonomiske grundlag (elever med en stærk socioøkonomisk baggrund får alt andet lige bedre resultater end elever med en svagere socioøkonomisk baggrund).

⁵² Lærere, undervisning og elevpræstationer i folkeskolen, SFI, 2013.

Boks 3.3: Et eksempel fra evalueringen: En stærk evalueringskultur understøtter gode faglige resultater

Evaluator har besøgt en folkeskole, der viser, at en høj evalueringskultur og et højt fagligt niveau er hinandens forudsætninger. Skolens overordnede mål er at lave verdens bedste folkeskole. Der er enighed om, at faglighed har førsteprioritet og trivsel et middel til at nå faglige resultater. Skolelederen påpeger, at skolens målsætning om et højt fagligt niveau er meget eksplicit, og at hverken lærere eller forældre er i tvivl om denne: "Det er fagligheden, der er i fokus her. Det betyder ikke, at alt det andet ikke er vigtigt, men her går vi op i, at eleverne lærer noget og trives med det. Ikke omvendt."

Skolens stærke evalueringskultur afspejles i følgende forhold:

- Skolen har en klar og strategisk evalueringskultur, herunder en årlig testoversigt som både kommune, lærere og forældre kan orientere sig i. Den beskriver, hvad og på hvilke klassetrin, der testes i indeværende skoleår.
- Skoleledelsen har sat den ramme, at alle de frivillige test er obligatoriske.
- Skolen har en målsætning for faglige resultater i nationale test såvel som folkeskolens afgangsprøver og læsetest på 1., 3. og 5. årgang.
- Internt i ledelsen er arbejdsdelingen den, at skolelederen har det strategiske overblik over testresultaterne, mens den pædagogiske afdelingsleder arbejder med at udmønte resultatet på elevniveau og er i tæt dialog med lærerne om den pædagogiske opfølgning. Det strategiske overblik indebærer bl.a., at skolelederen tager testresultaterne med i sine overvejelser om fagfordeling, ressourcefordeling m.v.
- De nationale test er et centralt delelement i en bred evalueringspraksis:
 - I dansk benyttes standardiserede test (OS64, Ordlæs, SL60, ST mv.) og nationale test som kvantitative metoder. Skolen benytter LUS (læseudviklingsskema) som kvalitativt supplement, da det efter deres vurdering kommer mere i dybden med elevernes læseudvikling. I matematik benyttes MAT- testen udover de nationale test.
 - Den kvalitative evaluering dækkes ofte i hverdagen, bl.a. ved at flere undervisningsmaterialer, herunder skolens dansksystem "Vild med dansk", har inkorporeret, at hvert emne følges af en mindre evaluering.
- Skolen har årlige klassekonferencer, hvor alle klassens fag tages op, og omdrejningspunktet er resultaterne fra de nationale test såvel som de andre test. Man kigger på tværs af fag og sammenligner testresultater. Målet er, at lærerne skal være gode til at bruge hinanden på tværs af fag, således at det fx opdages såfremt et problem i matematik eventuelt skyldes elevens ringe læsekundskaber. På klassekonferencen besluttet en indsats over for de svage elever.

De nationale test indgår altså som et delelement i en evalueringspraksis, der er testorienteret og systematisk, men som samtidig lægger vægt på løbende dialog om analyse af og opfølgning på resultatet. Alle kender og arbejder hen imod de faglige mål og anvender evaluering til løbende at vurdere, om man når derhen. Skolens evalueringskultur understøtter således det overgribende mål om at opnå gode faglige resultater.

Det er evaluators samlede vurdering, at de skoler, der har en stærk evalueringskultur, har en række fordele i deres arbejde for at fremme gode resultater end skoler, der ikke har en stærk evalueringskultur. Disse skoler har redskaber, der gør dem i stand til mere systematisk at vurdere årsager til en given præstation (for en elev, klasse, årgang eller skole som helhed) samt om en præstation skal give anledning til at justere indsatser og/eller mål. Ligeledes er det evaluators vurdering, at skoler med en stærk evalueringskultur er mindre følsomme for udsving i personale eller elevgrundlag, idet der er en samlet organisatorisk tilgang til kvalitetsarbejdet, der er mindre personbåren end på skoler med en svagere evalueringspraksis.

Derudover viser casestudierne, at effekten af de nationale test på mange skoler endnu ikke er slået fuldstændig igennem i forhold til evalueringskulturen, men mere i forhold til at cementere gode faglige resultater som et vigtigt mål. De fleste skoler har kun arbejdet med de nationale test i to til tre år. Således kan man forestille sig, at sammenhængen mellem evalueringskultur og faglige resultater bliver stærkere, når skolerne gennem længere tids anvendelse får styrket deres samlede evalueringskultur.

3.4 Delkonklusion

I dette kapitel er der foretaget en række analyser af effekterne af indførelsen af de nationale test. Effektanalysen er foretaget ved at undersøge eventuelle sammenhænge mellem nationale test og skolernes evalueringskultur samt elevernes faglige resultater. Nedenstående tabel indeholder en oversigt over de forskellige sammenhænge samt analysens resultater i oversigtsform.

Table 3.8: Oversigt over resultater af analyser omkring effekten af de nationale test

Sammenhæng	Hypotese	Data	Status
Nationale test → faglige resultater	Anvendelsen af nationale test styrker elevernes faglige niveau	Registerdata	Bekræftet +
Nationale test → skolens evalueringskultur	Anvendelsen af nationale test styrker folkeskolens evalueringskultur	Spørgeskemaundersøgelse, casestudier og øvrig dokumentation	Delvis bekræftet +
Skolens evalueringskultur → faglige resultater	En stærk evalueringskultur i folkeskolen styrker elevernes faglige niveau	Spørgeskemaundersøgelse, registerdata og casestudier	Ikke bekræftet -

Analysen af **sammenhængen mellem de nationale test og elevernes faglige resultater** indikerer en positiv effekt af de nationale test i relation til elevernes faglige resultater og bekræfter således hypotesen. Både analyserne på elev-, skole- og landsniveau indikerer, at de nationale test har en positiv effekt på elevernes faglige niveau. Resultaterne tyder endvidere på, at denne effekt ikke alene skyldes *teaching to the test*, eller at eleverne blot er blevet bedre til at gennemføre testene, men at de faktisk er blevet dygtigere til det fag, de er blevet testet i. Det har dog alene været muligt at se på effekten af nedbrud i dansk, læsning. Derudover har det kun været muligt at undersøge de generelle effekter af de nationale test – dvs. om testene har en effekt på elevernes faglige niveau og ikke, *hvordan* de har en effekt.

Analysen af **sammenhængen mellem de nationale test og skolernes evalueringskultur** tyder for det første på, at skolerne overordnet har styrket deres evalueringskultur i de senere år. Styrkelsen af evalueringskulturen ses ved, at skolerne som gruppe (som udtrykt af skolelederne) betragter evaluering som et vigtigt kvalitetsunderstøttende redskab, samt at de anvender en række evalueringsredskaber i deres arbejde. For det andet er der indikationer på, at de nationale test har medvirket til at styrke skolernes evalueringskultur. Styrkelsen af evalueringskulturen falder til en vis grad tidsmæssigt sammen med indførelsen af de nationale test. Undersøger man de oplevede effekter af de nationale test, er skolelederne samt forvaltningerne forholdsvis positive i deres vurdering af, hvorvidt de nationale test har bidraget til at styrke skolernes evalueringskultur. Dog er lærerne mere skeptiske. Evaluator vurderer samlet, at de nationale test har haft en understøttende og dagsordenssættende effekt i relation til at understrege et fokus på faglige resultater såvel som at bidrage til en større systematik i skolernes evaluering og har derved medvirket til en stærkere evalueringskultur. Det er dog samtidig evaluators vurdering (som det anføres i lovbemærkningerne), at det ikke er de nationale test alene, der skaber en stærkere evalueringskultur. Det er derimod helheden i anvendelsen af nationale test i samspil med en række evalueringsværktøjer samt en stærk ledelse, der understøtter en stærk evalueringskultur. Evaluator vurderer sammenfattende, at evalueringskulturen i den danske folkeskole er styrket, og at indførelsen af de nationale test i et vist omfang har bidraget positivt hertil.

De statistiske analyser af **sammenhængen mellem skolernes evalueringskultur og deres elevers faglige resultater** viser ikke en overbevisende sammenhæng, og hypotesen kan således ikke umiddelbart bekræftes. Det kan forekomme som et paradoks, at de nationale test tilsyneladende har en forholdsvis positiv effekt på elevernes faglige resultater og har styrket skoler-

nes evalueringskultur, hvorimod der ikke er en signifikant sammenhæng mellem evalueringskultur og faglige resultater. Som allerede nævnt, kan resultaterne dog ikke tages som et endegyldigt bevis på et fravær af sammenhæng mellem skolernes evalueringskultur, og hvordan skolens elever klarer sig. Resultaterne kan blot ikke bekræfte, at der er en statistisk sammenhæng ud fra de tilgængelige data. På baggrund af casestudierne vurderer evaluatoren dog, at en stærk evalueringskultur har en klar fremmende effekt på at opnå gode faglige resultater. Men en stærk evalueringskultur er hverken en tilstrækkelig eller en nødvendig betingelse for at opnå gode resultater.

På baggrund af ovenstående vurderer evaluatoren, at der har været positive effekter af de nationale test i relation til at styrke elevernes faglige resultater. Det er ligeledes vurderingen, at de nationale test har haft en positiv indvirkning på skolernes evalueringskultur. Der er dog stadig potentiale for en væsentlig styrkelse af evalueringskulturen, hvilket vil blive uddybet i kapitel 4.

4. DE NATIONALE TEST SOM PÆDAGOGISK VÆRKTØJ

I dette kapitel belyses ét af en række evalueringsspørgsmål, der relaterer sig til evalueringstemaet: *Testenes anvendelse*. Kapitlet indeholder en analyse af, hvordan **lærerne** anvender de nationale test som led i deres pædagogiske praksis. Dette omfatter lærernes anvendelse af testresultaterne i relation til vurderingen af elevernes faglige niveau og planlægning af undervisningen. Afslutningsvis vil det blive analyseret, hvorvidt testen har indflydelse på lærernes prioritering af undervisningens indhold.

Analyserne i nærværende kapitel er tilvejebragt af et datagrundlag bestående af følgende:

- Desk research af lovgrundlag, relevant litteratur samt viden fra undersøgelser, evalueringer og forskning
- Spørgeskemaundersøgelser til kommunale forvaltninger, skoleledere samt lærere med erfaring inden for testfag
- Omfattende casestudier på udvalgte skoler.

Supplerende inddrages tilbagemeldinger givet af de fageksperter, der har givet input til evaluatoren i forbindelse med evalueringsspørgsmål, der er afrapporteret under *testenes indhold* (jf. kapitel 7).

Baggrund

Det fremgår af folkeskolelovens § 13, stk. 2, at der løbende skal foretages en evaluering af elevernes udbytte af undervisningen. Et af de overordnede mål med at indføre nationale test er at styrke denne løbende evaluering, jf. bemærkningerne til lovforslaget om indførelse af nationale test.

De nationale test er tænkt som et pædagogisk værktøj, der bl.a. kan *bidrage* til en nuanceret vurdering af den enkelte elevs udbytte af udvalgte testegnede undervisningsemner samt medvirke til skabe overblik over den enkelte elevs faglige niveau⁵³. Det er ligeledes forventningen, at testresultatet anvendes fremadrettet *sammen med* resultatet af den øvrige løbende evaluering til brug for den videre planlægning af undervisningen, vejledningen af den enkelte elev samt underretning af forældre, jf. førnævnte lovgivning.

I forlængelse af bemærkningerne om en fremadrettet anvendelse af resultaterne fremgår det endvidere, at der forventes en målretning af undervisningen til den enkelte elevs behov og forudsætninger. En formulering, der afspejler folkeskolelovens § 18, stk. 2, der fastslår undervisningsdifferentiering som det bærende pædagogiske princip i den danske folkeskole. Der er altså en klar forventning om, at nationale test ikke kun anvendes til at skabe overblik over de enkelte elevs faglige niveau, men også understøtter lærernes arbejde med at differentiere undervisningen.

Ovenstående baggrund for og mål med at indføre de nationale test i folkeskolen er afspejlet i evaluators forandringsteori og hypoteser, jf. bilag 1. Forandringsteorien tydeliggør en forventet sammenhæng mellem gennemførelsen af test, lærerens analyse af testresultatet, lærerens planlægning af undervisningen på baggrund heraf samt gennemførelsen af en undervisning, der tager udgangspunkt i både klassen og den enkelte elevs faglige behov. Den forventede virkning er på sigt en styrket kvalitet af undervisningen.

For at understøtte de politiske målsætninger, der er beskrevet ovenfor, har Skolestyrelsen (nu Kvalitets- og Tilsynsstyrelsen) udsendt en vejledning, der giver råd om, hvordan læreren kan analysere og anvende testresultaterne⁵⁴. Også her fremgår det, at resultaterne af de nationale test kan bruges til at målrette undervisningen mod elevernes niveau, justere klassens eller elevernes mål eller iværksætte yderligere evalueringer. Det understreges, at eleverne ikke bliver dygtigere af selve testen, men de bliver dygtigere, hvis læreren kan bruge testen til at målrette undervis-

⁵³ L 101 Lovbekendtgørelse om nationale test s. 8.

⁵⁴ Brug testresultaterne – inspiration til pædagogisk brug af resultater fra de nationale test, Skolestyrelsen, 2011.

ningen mod den enkeltes behov. Det er opfølgningen, ifølge styrelsen, der gør forskellen⁵⁵. En pointe, der også spiller en væsentlig rolle i denne evaluerings forandringsteori.

Afslutningsvis skal der peges på, at det samtidig med indførelsen af de 10 nationale test, blev gjort obligatorisk at udarbejde minimum én elevplan om året for hver elev. Denne elevplan skal bl.a. indeholde oplysning om den besluttede opfølgning på resultaterne af den løbende evaluering, herunder opfølgningen på de nationale test⁵⁶. Altså et krav om, at den fremadrettede anvendelse af testresultaterne beskrives skriftligt i hver plan.

Evaluators vil i nærværende kapitel undersøge, hvorvidt disse bagvedliggende intentioner er indfriet.

4.1 Lærernes anvendelse af testresultaterne til vurdering af eleverne

I en analyse af lærernes brug af testresultater kan man, som det fremgår af indledningen, skelne mellem lærernes brug af resultatet til en **vurdering af elevernes faglige niveau** og som udgangspunkt for den videre **planlægning af undervisningen**. I dette afsnit analyseres førstnævnte, altså hvorvidt lærerne anvender testene til at monitorere elevernes faglige styrker og svagheder, jf. model for evalueringskultur.

For at undersøge hvorvidt lærerne anvender resultatet til en vurdering af eleven, er de deltagende lærere i spørgeskemaundersøgelsen blevet bedt om at tage stilling til, i hvilken grad de bruger testresultatet til at få et samlet overblik over henholdsvis klassen og de enkelte elevers faglige niveau.

Tabel 4.1: Tænk tilbage på, da du fik resultaterne af de nationale test i foråret 2012. Hvad har du brugt resultaterne af de nationale test til? Angivet i pct.

	Slet ikke	I mindre grad	I nogen grad	I temmelig høj grad	I meget høj grad	I alt (N)
Til at få et samlet overblik over klassens faglige niveau	2	8	32	41	17	100 (2162)
Til at få et overblik over de enkelte elevers faglige niveau	2	11	34	38	16	100 (2162)

Kilde: Spørgeskemaundersøgelse til lærere, 2013.

Tabellen ovenfor viser, at 58 pct. af lærerne svarer, at de i temmelig høj eller i meget høj grad benytter testresultatet til at få et overblik over klassens samlede niveau. Ser man på, hvorvidt lærerne bruger resultatet til at få et overblik over de enkelte elevers niveau, er tallet marginalt mindre. 54 pct. svarer, at de i temmelig eller i meget høj grad bruger resultaterne til at få overblik. For begge spørgsmål gælder det, at ca. en tredjedel af lærerne angiver, at de i nogen grad bruger resultaterne. Kun en meget lille andel af de adspurgte lærere giver udtryk for, at de slet ikke anvender testresultaterne. Samlet set kan det altså konstateres, at testen anvendes til vurdering af elevernes faglige niveau af et stort flertal af lærerne.

Den lille forskel mellem anvendelse på klasse- og elevniveau kan ikke bekræftes fuldt ud af case-studierne. De interviewede lærere fortæller, at de først og fremmest orienterer sig om resultatet på klasseniveau, når en test er gennemført, og resultatvisningen er tilgængelig, sådan som det udtrykkes i dette citat:

Det er rart at se, hvordan klassens faglige niveau er – også i forhold til landsgennemsnittet! Jeg kunne tydeligt aflæse, at min klasse var dygtig til afkodning og tekstforståelse, men at det kneb lidt med sprogforståelsen. Det fik mig til i højere grad at gå i dybden med enkelte ord fra de tekster, vi arbejder med – og jeg blev meget forbavset over deres uvidenhed!

Kilde: Kommentarfelt i spørgeskemaundersøgelsen for lærere.

⁵⁵ Brug testresultaterne – inspiration til pædagogisk brug af resultater fra de nationale test, Skolestyrelsen, 2011.

⁵⁶ Bekendtgørelse om elevplaner, elev- og uddannelsesplaner samt uddannelsesplaner i folkeskole, § 3, stk. 2.

Mange lærere giver udtryk for, at de kun ser på enkeltelevers resultater, hvis det overrasker i forhold til deres egen vurdering, ikke mindst hvis det overrasker negativt. Lærere, der har klasser, hvor det går godt, og hvor resultaterne er langt over middel, giver generelt udtryk for, at de bruger mindre tid på opfølgning og vurdering af eleverne.

På baggrund af ikke mindst casestudierne er det evaluators vurdering, at testresultaterne oftest ses som et supplement af lærerne. Flere peger på, at nationale test mest af alt bekræfter et allerede eksisterende billede af elevernes faglige niveau, der er dannet på baggrund af anden løbende evaluering. Dette synspunkt understøttes af kommentarer i spørgeskemaundersøgelsen:

"Bruger ikke udelukkende nationale test som evalueringsredskab, der er jo flere aspekter i omløb – synes det giver et alt for entydigt billede – har jo andre evalueringsredskaber."

"De nationale test bruges sammen med de andre papirtest, som en helhed sammen med mine egne observationer."

Kilde: Kommentarfelter i spørgeskemaundersøgelsen.

At evalueringsresultaterne ses som supplement til anden løbende evaluering stemmer overens med de overordnede målsætninger i lovbemærkningerne, hvor det som nævnt netop pointeres, at nationale test er ét ud af flere pædagogiske værktøjer. Det er også en del af denne evalueringens forandringsteori (og model for evalueringskultur), at der er et samspil mellem nationale test og den øvrige evalueringsindsats.

Ser man på anvendelse af resultatet til vurdering af elevernes faglige niveau fordelt på de enkelte testfag, så er der imidlertid tydelige forskelle, som det fremgår af tabellerne nedenfor.

Tabel 4.2: Tænk tilbage på, da du fik resultaterne af de nationale test i foråret 2012. Hvad har du brugt resultaterne af de nationale test til? Til at få et samlet overblik over klassens faglige niveau. Angivet i pct.

		Slet ikke	I mindre grad	I nogen grad	I temmelig høj grad	I meget høj grad	I alt (N)
Testfag:	Matematik	3	8	31	40	18	100 (499)
	Læsning	1	6	33	42	19	100 (1043)
	Engelsk	1	3	31	44	21	100 (234)
	Geografi	5	9	34	38	13	100 (97)
	Fysik/kemi	7	17	34	37	5	100 (202)
	Biologi	6	10	36	32	16	100 (87)

Kilde: Spørgeskemaundersøgelse til lærere, 2013.

Tabel 4.3: Tænk tilbage på, da du fik resultaterne af de nationale test i foråret 2012. Hvad har du brugt resultaterne af de nationale test til? Til at få et overblik over de enkelte elevers faglige niveau. Angivet i pct.

		Slet ikke	I mindre grad	I nogen grad	I temmelig høj grad	I meget høj grad	I alt (N)
Testfag:	Matematik	3	12	32	38	15	100 (499)
	Læsning	1	7	34	40	18	100 (1043)
	Engelsk	2	6	33	44	16	100 (234)
	Geografi	5	20	34	30	11	100 (97)
	Fysik/kemi	6	25	36	27	6	100 (202)
	Biologi	7	15	43	24	12	100 (87)

Kilde: Spørgeskemaundersøgelse til lærere, 2013.

Tabellen viser, at 65 pct. af lærerne i engelsk angiver, at de i meget eller temmelig høj grad bruger resultatet på *klasseniveau*. Dette tal er kun 42 pct. for fysik/kemi. Tilsvarende er tallene noget lavere for de to andre naturfag på 8. klassestrin.

Billedet er endnu tydeligere, når der spørges til, hvorvidt resultatet bruges til at få overblik over enkelte elevers faglige niveau. Her er svarer kun 33 pct. af fysik-/kemilærerne 'I meget høj grad' eller 'I temmelig høj grad', hvor det i både dansk, læsning og engelsk er ca. 60 pct. Omvendt svarer 31 pct. af fysiklærerne, at de slet ikke eller i mindre grad bruger resultaterne. For dansk, læsning og engelsk er dette tal kun ca. 8 pct.

En antagelse kunne være, at resultaterne af en eller anden grund er sværere at orientere sig i for nogle faggrupper end andre, men i spørgeskemaundersøgelsen har 70 pct. af lærerne på tværs af fag angivet, at de er helt – eller overvejende – enige i, at præsentationen af resultaterne i testsystemet giver et godt overblik over klassens faglige niveau i de dele af faget, der testes⁵⁷.

Det er evaluators vurdering, at de store forskelle i lærernes anvendelse af testresultaterne afspejler graden af testenes samspil med denne øvrige evaluering. Det vil sige, at i nogle fag indgår testen og testresultaterne i højere grad i lærernes (evaluerings)praksis/monitorering end i andre.

Ikke mindst flere af de interviewede naturfagslærere, herunder fysik/kemi, giver udtryk for, at det ikke opleves som meningsfuldt/udbytterigt at inddrage resultaterne i deres evaluering og planlægning af undervisningen. Dette vil blive uddybet i afsnittene om den fremadrettede anvendelse nedenfor.

Nationale test i samspil med anden løbende evaluering

Som nævnt indledningsvis forventes det, at testresultaterne fra nationale test anvendes fremadrettet sammen med resultatet af den øvrige løbende evaluering. Evaluator har derfor spurgt ind til brugen af øvrige evalueringsværktøjer for at få et billede af den supplerende evalueringsspraksis, testen indgår i et samspil med.

Tabel 4.4: I hvilken grad gør du brug af nedenstående øvrige evalueringsværktøjer udover de nationale test? Angivet i pct.

	Slet ikke	I mindre grad	I nogen grad	I temmelig høj grad	I meget høj grad	I alt (N)
Portfolio	52	20	17	7	3	100 (2162)
Lærerlogbog	38	21	25	13	3	100 (2162)
Elevlogbog	46	24	21	7	2	100 (2162)
Elevplan	3	5	22	42	28	100 (2162)
SMTTE og kvalitetsstjerne	61	18	14	6	1	100 (2162)
Delphi-metoden	91	6	2	1	0	100 (2162)
Evalueringssamtale	7	7	32	38	17	100 (2162)
Hverdagsevaluering	5	9	31	39	15	100 (2162)
Selvevaluering (blandt eleverne)	13	23	39	20	5	100 (2162)
Aktionslæringsforløb	69	15	12	4	1	100 (2162)
Undervisningsobservation i samarbejde med teammedlemmer, faglige vejledere el.lign.	47	24	19	7	3	100 (2162)
Selvproducerede test, prøver og opgaver	7	12	30	35	17	100 (2162)
Kommunale test	58	11	15	11	5	100 (2162)
Andre standardiserede prøver	21	12	27	27	13	100 (2162)
Andet	59	7	12	11	12	100 (2162)

Kilde: Spørgeskemaundersøgelse til lærere, 2013.

⁵⁷ Resultatvisningen analyseres selvstændigt i kapitel 8.

Det fremgår af tabellen, at de mest anvendte evalueringsværktøjer er selvproducerede test, prøver og opgaver samt standardiserede test⁵⁸. Dette bekræftes af casestudierne. Enkelte af skolerne er dog ved at udfase de standardiserede læse- og matematiktest på de klassetrin, hvor der er nationale test. Det er evaluators antagelse, at ledelsen her vurderer, at nationale test giver et lige så godt overblik over elevernes faglige udvikling, som de standardiserede test. To test samme år er derfor ikke nødvendigt.

Størstedelen af alle lærere på tværs af fag peger desuden på afleveringsopgaver, selvproducerede test og prøver eller prøver, der indgår som en del af et lærebogssystem, som noget af det, der fylder mest, når de skal beskrive deres egen evalueringspraksis. Netop sidstnævnte typer evalueringer er noget, som lærerne i naturfag på 8. klassetrin fremhæver i casestudierne. Dette kan være en del af forklaringen på, at denne faggruppe bruger testresultaterne i mindre grad.

I forlængelse heraf skal det fremhæves, at stort set alle de interviewede lærere på tværs af fag peger på den løbende evaluering af eleverne i den daglige undervisning på baggrund af dialog med den enkelte elev, classesamtaler, fremlæggelser mv. som en meget anvendt evalueringsform⁵⁹. At denne tilgang kan fylde meget, kommer klart til udtryk i en kommentar fra spørgeskemaundersøgelsen:

“De nationale test fortæller ikke noget, som den gode lærer ikke vidste på forhånd. Ingen evalueringsform kan måle sig med den løbende dialog mellem lærer og elev sammenholdt med de observationer, læreren gør sig i undervisningen.”

Kilde: Kommentarfelt i spørgeskemaundersøgelse til lærere.

Den dialog- og samtalebaserede evalueringspraksis har en mere kvalitativ karakter og bygger på lærerens egen subjektive vurdering. Nogle lærere vælger at supplere denne evalueringsform med en lærerlogbog eller andre lignende skriftlige værktøjer. Det er dog evaluators indtryk, at det er de færreste lærere, der benytter sig af dette.

På trods af, at de nationale test i vid udstrækning bliver brugt af lærerne i den danske folkeskole til en vurdering af elevernes faglige niveau, ikke mindst på klasseniveau, er det evaluators vurdering, at mange lærere ikke oplever, at nationale test gør en forskel i relation til det pædagogiske arbejde i klassen. Det vil sige inddragelse af resultatet som grundlag for *planlægning* af den videre undervisning. Dette uddybes i næste afsnit.

4.2 Lærernes anvendelse af testresultaterne til planlægning af undervisning

Evaluator gør brug af model for evalueringskultur som referenceramme til at vurdere skolernes praksis inden for evaluering – og dermed deres grad af evalueringskultur. Forenklet indebærer dette, at skoler vurderes af have svag/mindre grad af evalueringskultur, hvis man alene bruger resultaterne af de nationale test til vurdering, altså summativt. Tilsvarende klassificeres evalueringsspraksis som værende på et højt niveau, hvis resultatet af evalueringen (bl.a. på baggrund af kollegial dialog og refleksion) bruges som grundlag for planlægning af en undervisning, der tager udgangspunkt i den enkelte elevs behov og forudsætninger (undervisningsdifferentiering), det vil sige formativt.

Det skal dog understreges, at der er tale om en flydende overgang mellem de forskellige niveauer. En summativ anvendelse af resultaterne, som beskrevet i forrige afsnit, udelukker derfor ikke en supplerende formativ brug, der vil blive analyseret nedenfor.

Evaluator har i spørgeskemaundersøgelsen spurgt lærerne, i hvilken grad de har brugt resultatet af nationale test fremadrettet, jf. tabellen nedenfor.

⁵⁸ Det drejer sig konkret om prøver fra det der tidligere hed Dansk Psykologisk Forlag nu Hogrefe. Som eksempel kan nævnes – MAT-prøverne i matematik samt OS 64, OS 120, ST-prøver mv. i læsning.

⁵⁹ På trods af at ministeriet/styrelsen på uvm.dk definerer denne type evaluering som forskellige former for løbende skriftlig evaluering, er det evaluators vurdering, at hverdagevaluering af mange lærere er blevet fortolket som løbende mundtlig dialog ved besvarelse af spørgeskemaet. I så fald kan tabel 4.4 om evalueringsværktøjer bekræfte, at mange lærere gør brug af denne type evaluering.

Tabel 4.5: Tænk tilbage på, da du fik resultaterne af de nationale test i foråret 2012. Hvad har du brugt resultaterne af de nationale test til? Angivet i pct.

	Slet ikke	I mindre grad	I nogen grad	I temmelig høj grad	I meget høj grad	I alt (N)
Til at tilrettelægge en undervisning, der tager udgangspunkt i den enkelte elevs faglige niveau	9	22	42	23	5	100 (2162)
Til at tilrettelægge en undervisning, der tager udgangspunkt i klassens faglige niveau	9	21	44	22	4	100 (2162)

Kilde: Spørgeskemaundersøgelsen til lærere, 2013.

Det fremgår af tabellen, at 28 pct. af lærerne tilkendegiver, at de på baggrund af resultaterne af de nationale test i meget høj eller i temmelig høj grad tilrettelægger en undervisning, der tager udgangspunkt i *den enkelte elevs* faglige niveau. 42 pct. gør det i nogen grad, mens 31 pct. gør det i mindre grad eller slet ikke. Resultatet er stort set det samme, når der spørges til, hvorvidt lærerne på baggrund af resultaterne tilrettelægger en undervisning, der tager udgangspunkt i *klassens* faglige niveau.

På trods af, at 42 pct. af lærerne svarer, at de i nogen grad bruger resultatet, så er det evaluators overordnede indtryk, på baggrund af casestudierne, at det fortsat kun er en forholdsvis lille del af lærerne, der *konsekvent og systematisk* bruger testresultaterne fremadrettet som grundlag for planlægning af undervisningen.

Direkte adspurgt har de fleste lærere svært ved at komme med konkrete eksempler på anvendelse af testresultater til at understøtte en differentieret undervisning⁶⁰ i forhold til den enkelte elev. Nogle lærere (de fleste matematik- og dansk lærere) nævner, at de udarbejder hæfter eller mapper med individuelle træningsopgaver, så eleverne hver især kan arbejde med de faglige temaer, hvor de har udfordringer. Der er altså først og fremmest tale om en differentiering af indholdet (opgavedifferentiering), sådan som det kommer til udtryk i citatet nedenfor:

“De nationale test er et godt redskab til at få et overblik over elevernes standpunkt inden for det emne, der testes i. Efter enhver test er det (ifølge mig) vigtigt at evaluere med eleverne. De nationale test giver ikke altid et retvisende billede af elevens standpunkt i dansk, men kan vise områder, hvor eleven har brug for mere fokus, træning osv.”

Kilde: Kommentarfelt i spørgeskemaundersøgelsen til lærere.

Undervisningsdifferentiering har imidlertid altid været en udfordring i den danske folkeskole. Som eksempel kan nævnes EVAs undersøgelse fra 2011, der konkluderer, at der fortsat hersker usikkerhed om, hvad dette begreb dækker, samt at lærerne primært anvender evalueringresultater bagudrettet⁶¹. Det er evaluators opfattelse, at de få eksempler på reel undervisningsdifferentiering afspejler en mere generel tendens. Med andre ord kan denne evaluering bekræfte, at der er udfordringer med undervisningsdifferentiering, og brug af nationale test synes ikke umiddelbart *alene* at være løsningen på denne problematik.

Det skal dog nævnes, at man på én af de besøgte caseskoler har forpligtet lærerne på at fokusere på en differentieret tilrettelæggelse af undervisningen. Skoleledelsen har besluttet, at der skal

⁶⁰ Evaluator har i denne sammenhæng valgt en definition inspireret af Patricia A. Koeze som referenceramme: *Undervisningsdifferentiering er en undervisning, hvor læreren varierer indhold, processer og produkter med udgangspunkt i elevens parathed, interesser og læringsbehov*. Differentiated Instruction: Effect on Student Achievement in an Elementary School, s. 4, Patricia A. Koeze, 2007. En anden meget anvendt definition er N. Egelunds, der definerer undervisningsdifferentiering som fem forhold læreren skal gøre sig overvejelser om i planlægningen af undervisningen: Indhold/mål, metoder, organisering, materialer og tid (Undervisningsdifferentiering – Status og fremblik, Dafolo, Egelund, N., 2010).

⁶¹ Undervisningsdifferentiering som bærende pædagogisk princip – En evaluering af sammenhænge mellem evalueringfaglighed og differentieret undervisning, Danmarks Evalueringsinstitut, 2011.

udarbejdes en handlingsplan for de elever, der klarer sig klart over eller klart under middel i de nationale test. Det er altså ikke op til den enkelte lærer at tage stilling til, hvorvidt der skal følges op med individuelle tiltag.

Boks 4.1: Eksempel på handleplan fra caseskole

Den nationale test, som klassen gennemførte d. 17.4., viser, at Nikolaj klarede sig rigtig godt i den nationale test i matematik. Særlig stærk er han inden for områderne 'geometri' og 'matematik i anvendelse' og her vil Nikolaj i løbet af april få udleveret opgaver, der kan bidrage med flere udfordringer. Disse vil blive udvalgt i samarbejde med skolens matematikvejleder. Nikolaj har fået udleveret en ekstra Remaarbejdsbog, som passer til hans niveau, og desuden udfordres han i undervisningen med sværere opgaver og spørgsmål end mange af de øvrige elever. Nikolaj skal i øvrigt sammen med nogle af de øvrige elever lave en skitse over vores mooncarbane i det rigtige målestoksforhold m.m. Han udfordres endvidere med materialet 'snak om... - samtalekort' (for 3.-6. klasse).

Det er evaluators indtryk, at i det omfang, resultater anvendes fremadrettet, sker det oftest i forhold til et specifikt profilområde⁶², hvor klassen samlet set har scoret mindre godt. I sådanne situationer fortæller flere lærere, at de efterfølgende har gennemført et nyt undervisningsforløb med fokus på det konkrete profilområde.

Sammenholdt med evaluators model for evalueringskultur er hovedindtrykket, at der endnu er et stykke vej for mange skoler og lærere. På de fleste skoler arbejdes der, som konstateret i forrige afsnit, summativt med resultatet af testen, hvorimod det fortsat er en udfordring, som det er fremgået, med den formative anvendelse.

Den begrænsede fremadrettede brug kan ses som udtryk for, at en stor del af lærerne, som tidligere nævnt, i højere grad anvender resultatet af *anden løbende evaluering*. Men fag, kultur og ledelse spiller også ind. Førstnævnte tydeliggøres, når lærernes besvarelser deles op på fag.

Tabel 4.6: Tænk tilbage på, da du fik resultaterne af de nationale test i foråret 2012. Hvad har du brugt resultaterne af de nationale test til? Til at tilrettelægge en undervisning der tager udgangspunkt i klassens faglige niveau. Angivet i pct.

		Slet ikke	I mindre grad	I nogen grad	I temmelig høj grad	I meget høj grad	I alt (N)
Testfag:	Matematik	10	21	41	23	5	100 (499)
	Læsning	6	18	47	24	4	100 (1043)
	Engelsk	7	16	48	24	5	100 (234)
	Geografi	14	26	41	16	3	100 (97)
	Fysik/kemi	20	29	39	11	2	100 (202)
	Biologi	17	29	37	13	5	100 (87)

Kilde: Spørgeskemaundersøgelse til lærere, 2013.

⁶² Se kapitel 7 for definition af profilområder.

Tabel 4.7: Tænk tilbage på, da du fik resultaterne af de nationale test i foråret 2012. Hvad har du brugt resultaterne af de nationale test til? Til at tilrettelægge en undervisning der tager udgangspunkt i den enkelte elevs faglige niveau. Angivet i pct.

		Slet ikke	I mindre grad	I nogen grad	I temmelig høj grad	I meget høj grad	I alt (N)
Testfag:	Matematik	8	26	40	23	5	100 (499)
	Læsning	5	17	45	28	6	100 (1043)
	Engelsk	7	22	44	23	4	100 (234)
	Geografi	19	36	35	9	1	100 (97)
	Fysik/kemi	21	35	33	11	0	100 (202)
	Biologi	20	30	37	10	3	100 (87)

Kilde: Spørgeskemaundersøgelse til lærere, 2013.

Ovenstående tabeller viser, at engelsklærerne ligger på niveau med matematik- og dansklærere, når man ser på andelen af lærere, der i meget høj eller i temmelig høj grad angiver, at de bruger testresultatet på klasseniveau (ca. 30 pct.), tilsvarende billede tegner sig, når man ser på brugen i relation til enkelte elever. Engelsklærerne skiller sig dermed klart ud i forhold til gruppen af naturfagslærerne, hvor andelen af lærere, der i meget eller i temmelig høj grad bruger testresultaterne, er noget mindre, mellem 13 og 19 pct.

Denne forskel i anvendelsesgraden fagene imellem uddybes i afsnittene nedenfor.

Læsetesten

På trods af, at ikke mindst casestudierne overordnet set giver anledning til at konkludere, at det fortsat er en mindre andel af lærerne, der anvender de nationale test fremadrettet, så er det samtidig vurderingen, at flere kan se værdien i, at testresultaterne kan være med til at *kvalificere deres undervisning*, som én udtrykker det. Ikke mindst i relation til læsetesten gør denne udvikling sig gældende, hvilket der efter evaluators vurdering er flere grunde til.

For det første er der tale om en afgrænset færdighed inden for faget dansk, jf. kapitel 7. At læsetesten alene afspejler klart definerede elementer af faget, gør det mere overskueligt for den enkelte lærer at se på resultatet og udvælge de elementer, der skal følges op på i undervisningen.

Boks 4.2: Eksempel på opfølgning i relation til enkeltelever

En dansklærer fra mellemtrinnet fortæller, at hun udarbejder læsefokuspunkter for hver elev på baggrund af resultatet i de nationale test i dansk, læsning. Fokuspunkterne skrives ind i elevplanen og udleveres til den enkelte elev på et lamineret ark, så han eller hun kan have det med sig rundt både i skolen og hjemme. Et fokuspunkt kan fx være, at en elev "skal se nyheder og undre sig over nye ord, som han/hun støder på".

For det andet gennemføres der, som tidligere nævnt, fire obligatoriske læsetest på 2., 4., 6. og 8. klassetrin. Dertil kommer de frivillige test året før og efter, hvilket gør det muligt at gennemføre en national læsetest hvert år. Dette sammenholdt med danskfagets timetal skaber mulighed for at følge systematisk og kontinuerligt op på resultaterne – noget lærerne, der underviser i naturfagene på 8. klassetrin, efterlyser (se nedenfor).

For det tredje peger flere af de besøgte skoler/interviewede lærere på, at de har haft stor glæde af *profilberegneren* med tilhørende bog⁶³. Bogen giver konkrete anvisninger til, hvordan testresultatet fordelt på elevprofiler kan indgå i lærerens analyse med henblik på den fremadrettede planlægning af undervisningen. Lærerne på flere skoler fremhæver denne hjælp til at *oversætte* resultatet til konkret pædagogisk praksis som meget værdifuldt. Flere peger også på, at kurser

⁶³ Nationale test og anden evaluering af elevens læsning, L. Pøhler & S. A. Sørensen, Dafolo, 2011.

med fokus på analyse af resultater og forskellige elevprofiler har været en øjenåbner i relation til anvendelse af testresultaterne.

For det fjerde har stort set alle skoler i dag ansat en læsevejleder⁶⁴, hvilket antages at kunne understøtte en fremadrettet anvendelse af testresultaterne. Se nedenfor om brugen af ressourcerpersoner.

Sidst, men ikke mindst, har lærerne i dansk en lang tradition for at gennemføre og anvende test inden for læseområdet.

Evaluatoren har undersøgt om der forskel på graden af anvendelse, når man deler lærernes besvarelse op på henholdsvis 2., 4., 6. og 8. klassetrin, men der kan ikke påvises en sammenhæng. Lærerne anvender resultatet i lige høj grad på alle trin.

Matematik

Hvor der, som nævnt, inden for testen i dansk, læsning sker en udvikling frem mod en højere grad af fremadrettet anvendelse, er det evaluators vurdering, at denne udvikling er mindre udtalt i relation til matematiktesten, om end resultatet fra spørgeskemaundersøgelsen viser, at faget ligger på niveau med dansk, læsning og engelsk, se tabel 4.6. I det omfang testresultaterne bruges fremadrettet, så sker det som oftest, som nævnt ovenfor, på klasseniveau. Undervisning, der tager udgangspunkt i den enkelte elev, sker i langt højere grad på baggrund af anden løbende evaluering.

Som begrundelse for ikke at bruge testresultaterne peger flere af de interviewede lærere på, at ikke mindst timingen er uhensigtsmæssig. På 6. klassetrin, hvor den anden af de to matematiktest er placeret, afgiver mange lærere deres klasse til en udskolingslærer. Dette medfører, at den lærer, der afgiver klassen, ikke er motiveret for eller har tid til at følge op på resultatet. Den lærer, der modtager klassen, bruger i nogle tilfælde resultatet til at orientere sig om klassens faglige niveau, hvor andre giver udtryk for, at de slet ikke har adgang til resultaterne⁶⁵. Hovedindtrykket er, at resultaterne fra testene på dette klassetrin i mindre grad kommer i anvendelse.

For andre lærere er udfordringen, at de ikke mener, at testen afspejler den daglige undervisning og undervisningsmateriale:

"De nationale test bør være medvirkende til en meget kompetent evaluering. Jeg mener dog ikke, at [opgaverne i] de nationale test i matematik på tilfredsstillende vis ligner de opgaver, eleverne møder i relevante grundmaterialer for årgangen. Dette bør man i højere grad tage udgangspunkt i. Grundmaterialerne lever op til trinmålene."

Kilde: Kommentarfelt i spørgeskemaundersøgelsen til lærere.

At opgaverne ikke afspejler gældende trinmål er en udfordring, som også de faglige eksperter påpeger (jf. kapitel 7). Flere matematiklærere giver udtryk for, at det generelt er en udfordring at udtrække resultatet af testen og oversætte det til konkret praksis. En lærer udtrykker det på følgende måde:

*"Jeg kan ikke samle op og bruge testen fremadrettet. Det ville tage alt for lang tid, fordi man skal se samtlige opgaver for hver elev. Hvad er fx **matematik i anvendelse**?⁶⁶ Det giver mig ikke nok information. Jeg vil kunne se, at jeg fx skal undervise mere i trekantsgeometri. Ligesom man kan med MG-testene."*

Kilde: Citat fra lærerinterview i casestudie.

Denne henvisning til specifikke standardiserede test gentages af flere lærere i casestudierne. De fremhæver MG-prøverne eller MAT-prøverne, der i dag findes i elektronisk form, som et eksempel på en test, der automatisk genererer en tilbagemelding til læreren, og som umiddelbart kan danne grundlag for en videre planlægning/opfølgning i undervisningen både på klasse- og elevni-

⁶⁴ 97,4 pct. af skolerne svarer i spørgeskemaundersøgelsen ja til, at de har en læsevejleder ansat på skolen.

⁶⁵ Dette skyldes i nogle tilfælde, at eleverne skifter til en overbygningsskole.

⁶⁶ Der henvises her til ét af de tre profilområder i matematik.

veau. Det er evaluators vurdering, at denne form for resultatvisning, sammenholdt med at der er en test på hvert klassetrin, er én af hovedårsagerne til, at mange lærere i højere grad bruger disse test som udgangspunkt for den videre planlægning⁶⁷.

Engelsk

På trods af at engelsk, ligesom naturfagene, kun har én test sent i skoleforløbet, så er der klar forskel på andelen af lærere, der angiver, at de bruger resultatet fremadrettet, som det fremgik af tabel 4.6 og 4.7.

Som nævnt gælder det for både engelsk og naturfagsgruppen, at der kun er én obligatorisk test i skoleforløbet. Resultatet indikerer derfor, at det ikke alene er antallet af test, der er afgørende for graden af anvendelse. Evaluator har derfor ud fra spørgeskemaundersøgelsen til lærerne undersøgt sammenhængen mellem lærernes holdning til de nationale test med deres grad af anvendelse, og der er en klar positiv sammenhæng. Jo mere positiv, lærerne er over for de nationale test, i jo højere grad anvender de resultatet af dem.

Dette resultat understøttes af det generelle indtryk fra casestudierne, hvor de fleste af engelsklærerne udtrykker, at de er tilfredse med testen. Dette skyldes ikke mindst, at der ikke har været standardiserede test til rådighed forud for indførelsen af de nationale test, som det har været tilfældet i matematik og dansk. Engelsklærerne har derfor nu fået muligheden for et mere objektivt billede af klassens faglige niveau. Inden indførelsen af de nationale test har lærerne i høj grad været afhængige af deres egen subjektive vurdering, og den har ikke altid været helt retvisende.

En lærer nævner som eksempel, at det kom bag på hende, hvor dårligt et resultat klassen havde inden for profilområdet ordforråd. Fremadrettet var det derfor en naturlig konsekvens i højere grad at målrette klassens undervisning i den retning.

Et andet spørgsmål i spørgeskemaundersøgelsen understøtter ligeledes billedet af engelsklærerne som en faggruppe, der bruger testresultatet i højere grad end andre. På spørgsmålet om, i hvilken grad man som lærer har dialog med hver enkelt elev om resultatet af testen, svarer 71 pct. af engelsklærerne, at de i meget høj eller i temmelig høj grad har denne dialog. For dansk og matematik er tallet lidt over 50 pct., og for naturfagene vedkommende ligger tallet lidt under. Det skal dog understreges, at en dialog med eleven om testresultatet ikke nødvendigvis er ensbetydende med, at det anvendes fremadrettet, men resultatet kan indikere, at netop engelsklærerne generelt har en særs tiltalt interesse i opfølgning på og anvendelse af testen.

Enkelte lærere har peget på, at testen ligger sent i skoleforløbet. En test på et tidligere klassetrin vil kunne understøtte en højere grad af fremadrettet anvendelse/opfølgning. Ikke mindst set i lyset af at engelsk vil blive introduceret som fag i 1. klasse fra 2014.

Naturfagene på 8. klassetrin

Som det fremgik af tabel 4.6 og 4.7 ovenfor, så er andelen af lærere i naturfag, der bruger resultatet af testen på klasseniveau, markant mindre end andelen af lærere i engelsk, dansk og matematik.

Tilbage meldingen fra naturfagslærerne på tværs af de skoler, der har deltaget i casestudiet, er stort set enslydende. De fleste giver udtryk for, at de ikke bruger testresultaterne. Det skyldes, dels at der er meget få timer i fagene, så der ikke er tid til at følge op, dels at undervisningen i fagene i høj grad er tilrettelagt ud fra faglige temaer/emner fordelt over de tre år, eleverne har fagene. (Dette er ikke tilfældet i engelsk, hvor de tre profilområder samt opgaveemnerne⁶⁸ i højere grad afspejler faglige elementer, der arbejdes med løbende). Konsekvensen af den temaopdelte undervisning i naturfagene fremgår klart af dette citat:

⁶⁷ Det har været fremhævet af flere datakilder, at der opleves utryghed ved, at der i matematiktesten kan være store udsving i resultaterne inden for samme profilområde, hvis elever tager både den obligatoriske og frivillige test. Evaluator har ikke haft mulighed for at validere disse udsagn.

⁶⁸ For definition af opgaveemner, se kapitel 7.

“Efter min mening skal eleverne testes i ting, som er gennemgået. Men da testen ikke er tilrettet den enkelte klasse, kan der ofte forekomme spørgsmål, hvor eleverne ikke har fået den fornødne undervisning. Muligvis fordi andre emner er valgt, eller fordi at emnet først er planlagt til senere. Eleverne er under og efter testene frustrerede, da de føler, at de ikke har viden om nødvendige ting.”

Kilde: Kommentarfelt i spørgeskemaundersøgelse til lærere.

Lærerne oplever altså, at testresultatet i højere grad afspejler de temaer, eleverne er blevet undervist i, frem for at give et retvisende billede af elevernes faglige niveau, og det er ikke motiverende for en fremadrettet brug. Nogle lærere har valgt at prioritere en hurtigere gennemgang af faglige temaer, så klassen er nået igennem flest muligt inden den obligatoriske test. Konsekvensen er, at tempoet er højt, og det går ud over forståelsen hos nogle elever.

På én af de skoler, hvor ledelsen havde klare forventninger til gode faglige resultater, lagde naturfagslærerne ikke skjul på, at de målrettede undervisningen frem mod den obligatoriske test. Det har resulteret i, at den praktiske del af fagene, som feltbiologi (fx ture ud af huset til skove og åer) er blevet nedprioriteret og stort set ikke finder sted mere, hvilket ærgrer dem. Målrkning af undervisningen frem mod testen uddybes yderligere nedenfor.

Samlet set er tilbagemeldingen fra denne faggruppe, at de i langt højere grad er orienteret mod afgangsprøverne end de nationale test. Sidstnævnte spiller en meget lille rolle i tilrettelæggelsen af det videre arbejde hos langt de fleste.

Skoleledernes vurdering

Til brug for vurderingen af, i hvilken grad lærerne anvender testresultatet, har evaluator også spurgt skolelederne om deres vurdering:

Tabel 4.8: I hvilken grad anvender skolens lærere resultater af de nationale test på følgende måde? Angivet i pct.

	Slet ikke	I mindre grad	I nogen grad	I temmelig høj grad	I meget høj grad	Ved ikke	I alt (N)
Bidrag til at få et overblik over klassens/elevens faglige niveau	0	4	28	38	29	1	100 (728)
Skolens lærere bruger resultaterne af de nationale test til fremadrettet at tilrettelægge en undervisning, der tager udgangspunkt i den enkelte elevs faglige niveau	0	11	46	30	11	2	100 (728)

Kilde: Spørgeskemaundersøgelse til skoleledere, 2013.

Skoleledernes besvarelse viser, at de generelt er mere positive i relation til lærernes anvendelse. 41 pct. vurderer, at lærerne i meget høj eller i temmelig høj grad anvender resultaterne fremadrettet i forhold til den enkelte elev. 46 pct. vurderer, at de gør det i nogen grad. Forskelligheden i vurderingen af anvendelsen af testen kan bl.a. hænge sammen med, at 62 pct. af skolelederne svarer, at de først og fremmest leder efter mål og rammer. Det vil i hovedtræk sige at udstikke overordnede visioner og mål samt fokusere på tillid, inddragelse og delegering. Pædagogisk ledelse i form af løbende dialog og sparring samt jævnlig klasserumsobservation fylder mindre i den daglige praksis.

Den divergerende opfattelse lærere og ledere imellem indikerer, at det i relation til *implementeringen* af de nationale test kan være en udfordring, at skoleledelsen ikke alle steder er tæt på praksis. På nogle skoler vil der være risiko for, at ledelsen har en misvisende opfattelse af, i hvilken grad lærerne arbejder med og anvender testene til vurdering og planlægning af undervisning.

gen. Konsekvensen kan være, at skoleledelsen ikke i tilstrækkelig grad får skabt nogle organisatoriske rammer og strukturer, der understøtter den fremadrettede anvendelse af testen. Der henvises til kapitel 6 for en detaljeret analyse af de deltagende skolars ledelsespraksis.

Det er evaluators vurdering, at langt de fleste lærere og skoleledere ikke længere sætter spørgsmålstegn ved, om man skal evaluere systematisk. Det er nu mere et spørgsmål om *hvordan*. Både i relation til hvilke evalueringsværktøjer man ønsker at anvende, og i relation til hvilke systematikker og rammer der skal tilrettelægges med henblik på at understøtte en fremadrettet anvendelse.

At resultatet af testene anvendes fremadrettet i begrænset omfang i forhold til den fremadrettede anvendelse kan for en del læreres vedkommende umiddelbart tolkes som et udtryk for et fravalg af nationale test som evalueringsværktøj til fordel for andre.

Ressourcepersoner

Afslutningsvist skal evaluator pege på ressourcepersonernes rolle i relation til anvendelse af testresultater. I evaluators model for evalueringskultur er det en afgørende faktor for en stærk kultur, at der finder et systematisk/kollegialt samarbejde sted med henblik på at tolke og bruge testresultatet fremadrettet. Denne del er tilsvarende fremhævet i den anvendte forandringsteori, bl.a. på baggrund af resultater fra nyere forskning⁶⁹.

Evaluator har i spørgeskemaundersøgelsen til lærerne spurgt til graden af brug af klasse- og fagteam samt ressourcepersoner. Med ressourcepersoner menes der læsevejledere, matematikvejledere mv.

Tabel 4.9: Drøftelse af brugen af resultater med klasse- eller fagteam samt ressourcepersoner. Angivet i pct.

	Slet ikke	I mindre grad	I nogen grad	I temmelig høj grad	I meget høj grad	I alt (N)
Jeg drøfter brugen af resultaterne af de nationale test i mit klasseteam eller årgangsteam	11	24	40	18	7	100 (2162)
Jeg drøfter brugen af resultaterne af de nationale test i et fagteam	24	32	30	12	3	100 (2162)
Jeg drøfter brugen af resultaterne af de nationale test med en faglig ressourceperson (læsevejleder, matematikvejleder mv.)	31	21	24	16	7	100 (2162)

Kilde: Spørgeskemaundersøgelse til lærere, 2013.

Resultatet viser, at knap halvdelen af lærerne slet ikke eller i mindre grad inddrager fagteam og faglige ressourcepersoner i drøftelse af testresultaterne. Kollegial sparring med klasseteamskolleger viser sig at være den mest anvendte form for dialog.

Casestudierne underbygger ovenstående og viser, at praksis er meget forskellig fra skole til skole. Evaluator har besøgt en skole, hvor inddragelse af fagteam/faglige ressourcepersoner i tolkning og analyse af testresultater er meget systematiseret og tydelig, først og fremmest på læseområdet, men med ambitioner om at udvide det til matematik og naturfag.

Andre af skolerne fortæller, at de er i en opstartsfasen, og at man fra næste skoleår vil arbejde systematisk med at inddrage i første omgang læsevejlederen i en fælles tolkning og opfølgning på læseresultaterne i danskfagteamet. De fleste af de besøgte caseskoler har dog ikke nogen fast praksis for inddragelse af ressourcepersoner/faglige vejledere.

⁶⁹ Se bl.a. Synlig læring – for lærere, Dafolo, John Hattie, 2013.

Casestudier på de skoler, der påtænker eller allerede har påbegyndt det systematiske arbejde med inddragelse af ikke mindst læsekonsulenter, giver evaluater grundlag for at vurdere, at inddragelse af faglige ressourcepersoner kan gøre en forskel i relation til den fremadrettede anvendelse af de nationale test som en del af den løbende evaluering. Den faglige vejleder kan netop hjælpe den enkelte lærer med at oversætte testresultatet til pædagogisk praksis, og man kan i den fælles dialog sammen med fagteamet finde frem til konkrete indsatser målrettet klassen, grupper af elever eller enkelte elever. Det er tydeligt, at de lærere, der har deltaget i et fælles fortolkningsarbejde, har en høj grad af faglig indsigt og viden om læsning – en indsigt og viden, som de bl.a. vil kunne sætte i spil i planlægningen af undervisningen.

Boks 4.3: Inddragelse af faglige ressourcepersoner på en skole (baseret på casestudie)

Skolen har to matematikvejledere, som hver har 250 vejledningstimer på årsbasis. Skolens læsevejleder er den pædagogiske afdelingsleder. Hun har 1400 timer om året til vejledningsfunktionen. Skolen har prioriteret de faglige vejledere og finansierer timerne ved at nedlægge tolærerordningen på skolen. Næste år bliver man nødt til at halvere antallet timer til de faglige vejledere grundet generel stram økonomi.

Efter gennemførelsen af nationale test er der stor forskel fra matematik til dansk på, hvordan de faglige vejledere benyttes.

Matematikvejlederen taler som udgangspunkt kun med lærerne, hvis der er udsving, som undrer ham. I det tilfælde sætter han sig med den pågældende lærer og analyserer testresultatet. Målet er bl.a. at finde frem til, hvorvidt eleven bare har klikket sig gennem opgaverne, eller om han/hun har taget testen seriøst. Matematikvejlederen bliver også brugt til at lave konkrete undervisningsforløb i klasser som opfølgning på testen (sidemandsoplæring). Det bruges meget på skolen. Dog pointeres det, at sidemandsoplæring ikke er en del af den danske skolekultur, hvorfor det har krævet en tilvænningsperiode.

Læsevejlederen har én gang årligt en klassekonference med alle klasser. Her tages alle klassens fag op med udgangspunkt i resultaterne af den nationale læsetest og med inddragelse af resultatet af andre standardiserede test. Målet er, at lærerne skal være gode til at bruge hinanden på tværs af fag. Et dårligt resultat i en matematiktest kan fx skyldes elevens ringe læsekundskaber.

På klassekonferencen redegør læsevejlederen indledningsvis for klassens resultat i læsetesten. Lærerne kigger derefter på tværs af fag og sammenligner testresultater. Efterfølgende tages der beslutning om en indsats over for de svage elever. Da der kun er klassekonferencer én gang årligt, kan der godt gå tid fra testen er gennemført til resultatet drøftes i plenum, men læsevejlederen påpeger, at lærerne og ledelsen er gode til at tage tingene løbende. Generelt gælder det, at ressourcerne især bruges, der hvor der er udfordringer (i form af elever med vanskeligheder).

Læsevejlederen har i indskolingen arbejdet målrettet med at ensrette evalueringspraksissen, således at dansklærerne i indskolingen i dag evaluerer ens, det vil sige, at de bl.a. er enige om, hvad der kendetegner specifikke faglige niveauer hos eleverne. Udviklingen er ifølge lærerne sket inden for de seneste 5 år. Læsevejlederen har holdt fast i fagudvalgsmøder, hvor dansklærerne bl.a. har diskuteret materialer og evalueringspraksis. Ifølge dansklærerne har ensretningen den værdi, at lærerne kan bruge hinanden som sparringspartnere på en ny og bedre måde, fordi de har udviklet et fælles sprog og en fælles forståelse af danskfaget og dets faglige elementer.

Vejlederne påpeger generelt, at det ligger i vejlederfunktionen at være opsøgende, og det er også deres opfattelse, at lærerne forventer det. Vejlederne forventer altså ikke, at lærerne opsøger dem på baggrund af resultaterne fra nationale test, det er den anden vej rundt. Begge vejledere pointerer, at nationale test er et positivt redskab, der kan understøtte deres funktion, da testene bidrager til at kvalificere den faglige dialog.

Det skal i denne sammenhæng nævnes, at i relation til engelsk og inddragelse af faglige vejledere er der en særskilt udfordring, som flere lærere i casestudierne peger på. Det er ofte sådan, at engelsklæreren i 7. klasse er alene på årgangen eller selv er fagansvarlig. Dertil kommer, at der kun er én test i hele skoleforløbet (de frivillige test nævnes ikke i denne sammenhæng). Engelsklærerne giver derfor udtryk for, at det kan være svært at finde et fagligt forum, hvor det giver mening at drøfte resultatet af testen.

4.3 De nationale tests indflydelse på lærernes prioritering af undervisningens indhold

Vurderingen af, i hvilken grad de nationale test har indflydelse på lærernes prioritering af undervisningens indhold, tager udgangspunkt i begrebet **teaching to the test**, der ofte refereres til som en utilsigtet konsekvens af brugen af test, samt det relaterede begreb **washback**, der defineres som *the effects of tests on teaching and learning, the educational system, and the various stakeholders in the education process*⁷⁰. Det vil sige, at testen bliver styrende for, hvad der kommer til at foregå i undervisningen.

Hvor definitionen af *teaching to the test* har en negativ klang, så åbner definitionen af washback op for, at man kan tale om både negativ og positiv påvirkning af/indflydelse på undervisningen. Hvorvidt der er tale om det ene eller det andet, vil i nogen udstrækning afhænge af øjnene, der ser⁷¹, som det fremgår nedenfor. Det er denne nuancerede tilgang, der ligger til grund for evaluators analyse.

Man kan tale om washback når:

- Der er forskel på undervisningen i et konkret fag, der leder frem mod en test, og den undervisning i samme fag, der ikke leder frem mod test
- Den undervisning, der leder frem mod testene, ligner hinanden hos flere forskellige lærere, og det, der undervises i, afspejler testen indhold.

Med inspiration fra Liying Cheng, Yoshinori Watanabe og Andy Curtis (red.):
Washback in language testing: Research contexts and methods, 2004, s. 37.

I forlængelse af ovenstående vil man også kunne tale om en stærk eller svag påvirkning af tilrettelæggelsen og gennemførelsen af undervisningen, herunder om testen påvirker hele eller dele af undervisningen, i kort eller lang tid, om testen påvirker alle lærere eller få lærere, i alle testfag eller i enkelte testfag.

Evaluator har ligeledes undersøgt, hvorvidt man kan tale om *washback to learners* i relation til elevernes måde at lære/forberede sig frem mod testene, idet der kan forventes en vis washback-effekt for nogle elevgrupper, der vil komme til udtryk ved en prioritering af deres arbejde og forberedelse af visse elementer, som de ved, de skal testes i, sammenlignet med andre elementer, som de ved, de ikke skal testes i. Dette besvares kort nedenfor.

⁷⁰ Washback in language testing: Research contexts and methods, s.37, Liying Cheng, Yoshinori Watanabe og Andy Curtis (red.), 2004.

⁷¹ De nationale test i et pædagogisk perspektiv, i: Test i folkeskole, Hans Reitzels Forlag, Stæhr, L. S., 2009.

Lærerne

Evaluators har i spørgeskemaundersøgelsen blandt lærerne spurgt ind til, i hvor høj grad deres undervisning påvirkes af de nationale test (obligatoriske og de frivillige), samt i hvilket omfang det sker, det vil sige, hvor meget af skoleåret der evt. påvirkes.

Table 4.10: I hvilken grad er du enig eller uenig i følgende udsagn om de nationale tests eventuelle indflydelse på din undervisning? Angivet i pct.

	Slet ikke	I mindre grad	I nogen grad	I temmelig høj grad	I meget høj grad	Ved ikke	I alt (N)
Jeg målretter min undervisning frem mod de obligatoriske test ved at inddrage elementer af faget, der testes i	18	27	36	13	6	1	100 (2162)
Jeg målretter min undervisning frem mod de frivillige test ved at inddrage elementer af faget, der testes i	24	31	30	10	4	1	100 (2162)
Jeg opprioriterer undervisningen i de faglige områder, der testes i, hele året	22	29	31	11	6	1	100 (2162)
Jeg opprioriterer undervisningen i de faglige områder, der testes i, de sidste uger inden testens gennemførelse	26	29	25	13	6	1	100 (2162)
Der er dele af fagets trinmål, som jeg vælger ikke at tilgodese i undervisningen, fordi det faglige område ikke testes i de nationale test	68	19	8	2	1	2	100 (2162)
Jeg målretter min undervisning frem mod de nationale test for at sikre, at resultaterne er så gode som muligt i forhold til landsgennemsnittet (som offentliggøres én gang årligt på www.uvm.dk)	42	28	20	6	4	1	100 (2162)

Kilde: Spørgeskemaundersøgelse til lærere, 2013.

Tabellen viser, at ca. halvdelen af lærerne giver udtryk for, at de slet ikke eller i mindre grad målretter undervisningen frem mod de frivillige og de obligatoriske test. Stort set samme andel giver udtryk for, at de ikke opprioriterer de elementer af faget, der testes i, hverken hele året eller de sidste uger op til testen. Modsat giver den anden halvdel udtryk for, at de både målretter og opprioriterer elementer af faget. Størstedelen dog kun i nogen grad.

Deler man lærerbesvarelserne op på indskoling, mellemtrin og udskoling, tegner der sig et tydeligt billede. Her viser det sig, at andelen af indskolingslærere, der i meget høj eller i temmelig høj grad målretter undervisningen frem mod de obligatoriske test, er 23 pct., hvor den kun er 13 pct. for udskolingslærernes vedkommende. Blandt udskolingslærerne er der lidt over halvdelen, der svarer, at de slet ikke eller i mindre grad målretter undervisningen. For indskolingslærernes vedkommende er der kun en tredjedel.

En sandsynlig årsag til ovenstående forskelle er, at udskolingslærerne først og fremmest er orienteret mod afgangsprøverne frem for testene, hvilket flere interviewpersoner har givet udtryk for.

Det er disse resultater, der først og fremmest tæller, og det er derfor, man målretter sin indsats i den retning.

Grupperinger blandt lærerne

Resultaterne i spørgeskemaundersøgelsen vedrørende testens indflydelse på lærernes undervisning, understøttes af casestudierne. Det er evaluators vurdering, at lærerne deler sig i to grupper. De lærere, der klart melder ud, at de aldrig kunne finde på at lade deres undervisning påvirke af de nationale test, og de lærere, der tilkendegiver, at de målretter undervisningen i ét eller andet omfang. Citatet afspejler den gruppe af lærere, der ikke lader sig påvirke:

"Nationale test skal ikke styre mig, og hvordan jeg laver god undervisning. Skal jeg sætte mig ned og kigge efter, hvad det er, jeg skal undervise efter? Nej! Jeg er med til at danne mennesker, og det er den der test ikke – den tester kun faglighed."

Kilde: Citat fra lærerinterview i casestudie.

Andre lærere sammenligner målretning af undervisningen med Tour de France. Man vinder ikke hele løbet ved kun at træne spurt. På samme måde bliver man ikke god til hele faget ved kun at fokusere på udvalgte elementer.

Blandt de lærere, der målretter deres undervisning, finder man yderligere en opdeling i to grupper. Den ene gruppe er af den opfattelse, at de faglige emner, der testes, er helt legitime. Testen afspejler Fælles Mål, og det giver derfor god mening at målrette dele af sin undervisning mod disse elementer, altså en positiv washback. Flere lærere og skoleledere fortæller, at de inddrager resultatet af de frivillige test i efteråret og bruger det til at tage beslutning om, hvilke faglige elementer der skal være fokus på i foråret. Flere peger dog ligeledes på, at der er tale om en hårfin balance, der ikke må tage overhånd. På baggrund af casestudierne er det ikke evaluators indtryk, at denne målrettethed udelukker undervisning i ikke-testbare kundskabs- og færdighedsområder i Fælles Mål.

Den anden gruppe lærere lader taktiske overvejelser veje tungt; *"Vi er jo ikke dumme"*, som én udtrykker det. Dette skal først og fremmest tolkes i retning af, at lærerne er bevidste om, at testresultaterne fra de nationale test har en anden form for fokus end andre standardiserede test. De oplever, at både forældre og ikke mindst skoleledelse er optaget af, at eleverne klarer sig godt. Læreren føler sig derfor nødsaget til at give denne test ekstra opmærksomhed. For en række lærere er der tale om en klart negativ påvirkning af undervisningen, hvilket understreges i dette citat:

Jeg oplever desværre, at lærere, deriblandt mig selv, føler sig nødsaget til at tilrettelægge undervisningen med det ene formål: At eleverne får et godt resultat i de nationale test, for at skolen fremstår som en "god skole."

Kilde: Kommentarfelt i spørgeskemaundersøgelse til lærere.

Nogle skoleledere oplever på samme måde, at de nationale test påvirker undervisningen negativt. De giver udtryk for, at testen har medført *teaching to the test* på en måde, der ikke er hensigtsmæssig, da de nationale test ikke i sig selv bidrager med noget nyt rent evaluermæssigt. Lærerne bliver på trods af dette nødt til at fokusere på indholdet af dem, da de er obligatoriske, og det ses som et forstyrrende element i en allerede eksisterende evalueringsspraksis.

En anden gruppe skoleledere oplever modsat, at testen på positiv vis har været med til at sætte fokus og øge bevidstheden om fagligheden⁷².

I forlængelse af ovenstående skal det også påpeges, at nogle lærere er bevidste om, at testen i ét eller andet omfang afspejler deres egen indsats, hvilket også har indflydelse på lærernes prioriteringer af undervisningen. Dette fremgår bl.a. af følgende citat:

⁷² Kommentarfelt i spørgeskemaundersøgelse til ledere.

“De nationale test er politikernes måde at flytte ansvaret væk fra sig og over på skolen samt øge kontrollen af den enkelte underviser. Testene er klart en stressfaktor på lærerværelserne, da man jo kan vurderes sammen med andre... Det bør være nok med tillid og ikke test.”

Kilde: Kommentarfelt i spørgeskemaundersøgelse lærere.

Andre oplever det som positivt, at lederen kan følge resultatet af ens praksis. Nogle lærere fortæller, at hvis ikke skolelederen selv er opmærksom på et godt testresultat, så gør de selv opmærksom på det.

Ikke mindst på de skoler, hvor skoleledelsen har tydelige forventninger til, at skolen klarer sig godt i de nationale test, fortæller flere lærere, at de målretter undervisningen. For en del lærere er det helt uproblematisk. For andre giver det anledning til bekymring. Ikke mindst for de lærere, der underviser i naturfag, da de som tidligere nævnt oplever, at det går ud over deres metodefrihed, da de ser sig nødsaget til at nedprioritere det praktiske element i faget. En naturfagslærer udtrykker det på følgende måde:

“De er i bedste fald spild af tid og ressourcer. Det bliver ”teaching to test”, hvor vi kun underviser i fx energiomsætning, fordi vi ved, at det bliver testet, selvom det intet har at gøre med det emne, vi har på det tidspunkt. Eleverne hader det, og de kan ikke bruge resultatet til noget. Få motiveres af test, de fleste bliver nervøse og frygter det. I sådanne tilfælde yder de ikke deres optimale. Elever med sproglige vanskeligheder, som måske er gode i det praktisk-eksperimentale, bliver fejlvurderet, så det basker.”

Kilde: Kommentarfelt i spørgeskemaundersøgelse til lærere.

Sammenholder man casestudierne med definitionen på washback ovenfor, er der ingen tvivl om, at det fag, hvor der i højeste grad sker en påvirkning af undervisningen, er i dansk læsning. På stort set alle skoler peges der på, at det element af profilområdet sprogforståelse, der vedrører opgaveemnet ”Faste vendinger”, har fået markant mere opmærksomhed i danskundervisningen. Evaluator har hørt flere eksempler på klasser, der dagligt træner dagens ordsprog eller talemåde. For nogle lærere er det uproblematisk, da det er et element, der inden testen havde en tendens til at blive overset. For andre er der tale om en skævvridning, da det kommer til at fylde for meget i forhold til faget samlet set.

Denne konstatering i relation til læsetesten understøttes af spørgeskemaundersøgelsen, når man ser på fagene enkeltvis. 62 pct. af dansklærerne har angivet, at de i nogen, i temmelig eller i høj grad målretter undervisningen frem mod de obligatoriske test. Dette tal er mellem 10-15 pct. lavere i de andre testfag.

Afslutningsvist skal der peges på, at mange lærere, som en del af målretningen frem mod frivillige og obligatoriske test, nævner gennemførelse af demotest og dialog om/træning af opgavetyper.

Ikke mindst mange indskolingslærere går op i at sørge for, at deres elever er forberedt på test-situationen. De er meget optaget af, at eleverne ikke får en dårlig oplevelse. Flere fortæller om situationer, hvor elever blive kede af det og frustrerede, når de ikke forstår, hvordan de helt konkret skal agere i skærbilledet.

Som eksempel nævner mange lærere ordkædeopgaverne under profilområdet afkodning i læsetesten samt fokus på regneudtryk i relation til matematik. Det er begge opgavetyper, hvor lærerne bruger tid på at gennemgå eksempler samt løsningsstrategier som fx at starte bagfra, når opgaver med ordkæder skal løses.

På baggrund af ovenstående kan det konstateres, at lærerne er meget delte i deres holdning til, hvorvidt man kan tale om positiv eller negativ washback, men det er evaluators hovedindtryk, at testen alene påvirker *dele* af undervisningen. Undervisningen i testfagene rummer fortsat en lang række andre elementer, der ikke er målrettet indholdet i de nationale test. En mindre undtagelse er de konkrete eksempler fra casestudierne i naturfag, hvor hele faget og dets metode ændrer fokus.

Eleverne

Dansk Clearinghouse for uddannelsesforsknings review (2009) viser, at der var stærk evidens for, at den enkelte elevs testresultater har indflydelse på fremtidig motivation såvel som selvværd. Generelt styrkes højt præsterende elevers motivation og selvværd, mens det modsatte er tilfældet hos svagt præsterende elever. Elevernes oplevelse af de nationale test vil blive behandlet i kapitel 5.

På spørgsmålet om, hvorvidt der har været en washback-effekt for nogle elevgrupper i form af en prioritering af deres arbejde og forberedelse på visse elementer, som de ved, de skal testes i, så har evaluator kun mødt elever på én skole, der klart tilkendegav, at de i ugerne op til testen arbejdede målrettede med at læse op, udarbejde noter mv. Der var tale om udskolings elever, der gik på en specifik bogligt/sprogligt orienteret linje, og de var generelt meget ambitiøse.

4.4 Delkonklusion

I dette kapitel er der foretaget en analyse af den pædagogiske anvendelse af de nationale test, dvs. lærernes anvendelse af testene i deres pædagogiske arbejde. Det skal understreges, at analysen i nærværende kapitel sammen med kapitel 5 og 6 dækker flere aktører og niveauer i folkeskolens styringskæde. Samlet set bidrager delanalyserne i kapitlerne 4, 5 og 6 til en vurdering af de nationale tests bidrag til målretning af skolernes arbejde.

I tabellen nedenfor er indsat en oversigt over de nøglehypoteser, der har været undersøgt vedrørende den pædagogiske anvendelse under det overordnede evalueringstema *testenes anvendelse*. Ved læsning af nedenstående tabel skal det understreges, at "status" er udtryk for evaluators samlede vurdering af den indsamlede empiri. Når en hypotese ikke bekræftes, er det således udtryk for, at der ikke er grundlag for at bekræfte de givne antagelser om sammenhænge vedrørende nationale test. Dette betyder imidlertid *ikke*, at hypoteserne (og den bagvedliggende forandringssteori) ikke er relevant/valid, men at der kan konstateres en række implementeringsudfordringer i forhold til at realisere de mål, der blev formuleret i lovgivningen. Samtidig ønsker evaluator at påpege, at casestudierne har dokumenteret flere eksempler på, at de nationale test i høj grad anvendes som tiltænkt i lovgivningen og hver for sig bekræfter udvalgte hypoteser.

Tabel 4.11: Oversigt over nøglehypoteser (den pædagogiske anvendelse)

Hypotese	Data	Status
Lærerne		
Resultatet af de nationale test styrker lærerens løbende evaluering af elevernes faglige styrker og svagheder (bidrager til et nuanceret billede af elevernes kundskaber og færdigheder)	Spørgeskemaundersøgelser til skoleledere og lærere Casestudier Ekspertinterview	Delvist bekræftet +
Resultatet af de nationale test målretter den enkelte lærers planlægning af undervisningen i relation til klassen og den enkelte elev	Spørgeskemaundersøgelser til skoleledere og lærere Casestudier Ekspertinterview	Ikke bekræftet -
Resultatet af de nationale test styrker den enkelte lærers differentiering af undervisningen	Spørgeskemaundersøgelser til skoleledere og lærere Casestudier Ekspertinterview	Ikke bekræftet -
Inddragelse af resultater fra de nationale test styrker og kvalificerer dialogen i lærerteams om udvikling af undervisning	Spørgeskemaundersøgelse til skoleledere og lærere Casestudier Ekspertinterview	Ikke bekræftet -
Inddragelse af ressourcepersoner styrker fortolkningen af resultatet af de nationale test og målretter lærerens opfølgning i relation til den enkelte elev/klassen	Spørgeskemaundersøgelse til skoleledere og lærere Casestudier Ekspertinterview	Ikke bekræftet -

Evalueringen dokumenterer, at langt størstedelen af **lærerne** gør brug af de nationale test. Denne overordnede konklusion nuanceres imidlertid ved, at lærerne i overvejende grad bruger testene til skabe overblik og dermed monitorerer klassens og elevernes faglige niveau. Derimod er brugen af de nationale test mindre udbredt i forhold til den fremadrettede tilrettelæggelse af undervisningen. Evalueringen viser, at mange lærere (indtil videre) primært anvender de nationale testresultater *summativt* og i langt mindre grad *formativt*. Derudover er der markante forskelle i anvendelsen mellem fagene. Anvendelsen af testresultaterne er således mere udbredt i dansk, læsning, matematik og engelsk, end det er tilfældet i naturfagene. Endelig er det en hovedkonklusion, at testresultaterne typisk ses som et supplement til viden indhentet fra øvrig evaluering. De nationale test står sjældent alene som redskab i en evalueringskultur.

Evalueringen dokumenterer, at det fortsat kun er en forholdsvis lille del af lærerne, der konsekvent og systematisk bruger testresultaterne fremadrettet som grundlag for planlægning af undervisningen. Den manglende formative anvendelse kan forklares ved, at mange lærere bruger andre evalueringsværktøjer. En anden forklaring kan være manglende fokus og rammesætning fra skoleledelsens side. Endelig spiller oplevelsen af manglende relevans ind for en del lærere. Særligt lærere inden for naturfag finder det ikke brugbart med én test relativt sent i folkeskolen (8. klasse), og derudover påvirker en række fagspecifikke forhold graden af testanvendelse.

Evalueringen undersøger, i hvilket omfang de nationale test har indflydelse på lærernes prioritering af undervisningens indhold. Der refereres ofte til begrebet *teaching to the test*, men evaluator anvender i stedet begrebet *washback* som et mere nuanceret udtryk for, at man kan tale om både negativ og positiv påvirkning af/indflydelse på undervisningen. Evalueringen viser et særdeles differentieret billede af omfanget af washback i relation til brugen af de nationale test. Lærerne på de undersøgte skoler fordeler sig i forskellige grupper, hvor halvdelen af lærerne giver udtryk for, at de slet ikke eller i begrænset omfang målretter deres undervisning i forhold til obligatoriske og frivillige test, hvorimod den anden halvdel gør det i forskellig grad. Hovedkonklusionen i forhold til dette spørgsmål er, at evalueringen tegner et meget varieret billede af testenes påvirkning af undervisningen. En gruppe lærere målretter undervisningen, enten fordi de finder det naturligt (da testene afspejler Fælles Mål), eller fordi de gerne vil have, at eleverne klarer sig godt, da der er fokus på resultaterne i de nationale test. Den anden gruppe omfatter de lærere, som meget bevidst ikke vil lade deres undervisning styre af de nationale test. En vigtig konklusion er i øvrigt, at *teaching to the test* (eller *washback*) er mere udbredt dels i udskolingen, dels på skoler hvor ledelsen har fokus på faglige resultater.

5. TESTENES BETYDNING FOR ELEVER OG SKOLE-HJEM-SAMARBEJDE

Nærværende kapitel belyser de nationale tests betydning for eleverne og for skole-hjem-samarbejdet. Kapitlet fokuserer på de tre af evalueringens hypoteser, der knytter sig hertil.

Kapitlet sætter fokus på hypotesen om, at lærerens dialog med eleven om testresultater styrker elevernes bevidsthed om egne læreprocesser og dermed motivation for læring. Dette tema behandles også i andre dele af evalueringen, men dette kapitel har særligt fokus på at belyse elevernes perspektiver herpå. Derudover belyser kapitlet hypotesen om, at dialog med hjemmet om resultatet af de nationale test fremmer forældrenes mulighed for at understøtte deres barns læring samt hypotesen om, at de nationale test styrker skole-hjem-samarbejdet.

Der er gennemført interview med elever og forældre i forbindelse med samtlige casestudier. Disse udgør det primære datagrundlag for kapitlet. Dog vil elevernes og forældrenes udsagn undervejs i kapitlet blive suppleret med og forholdt til lærernes besvarelse af spørgeskemaundersøgelsen.

5.1 Eleverne

I spørgeskemaundersøgelsen er lærerne blevet bedt om at angive, hvor enige de er i en række udsagn om elevernes udbytte af og holdning til de nationale test. Resultatet fremgår af nedenstående tabel.

Tablet 5.1: I hvilken grad er du enig eller uenig i følgende udsagn om de nationale test og eleverne? Angivet i pct.

	Helt uenig	Overvejende uenig	Hverken enig eller uenig	Overvejende enig	Helt enig	Ved ikke	I alt (N)
Resultatet af de nationale test styrker elevernes forståelse af faglige styrker og svagheder	8	21	32	33	4	2	100 (2162)
Dialog med eleverne om resultatet af de nationale test styrker elevernes motivation for at lære	9	19	40	26	3	3	100 (2162)
Dialog med eleverne om resultatet af de nationale test mindsker elevernes motivation for at lære	10	31	47	7	2	4	100 (2162)
Tilbage melding og dialog med eleverne om resultatet af de nationale test medvirker til, at eleverne i højere grad inddrages i formulering af fremtidige læringsmål	6	16	40	30	5	3	100 (2162)
Jeg oplever, at eleverne generelt er glade for de nationale test	13	23	40	19	4	2	100 (2162)

Kilde: Spørgeskemaundersøgelse til lærere, 2013.

Af tabellen fremgår det, at 37 pct. af lærerne er helt eller overvejende enige i, at resultatet af de nationale test styrker elevernes forståelse af faglige styrker og svagheder, mens 29 pct. er helt eller overvejende uenige i dette udsagn. Videre viser tabellen, at tre ud af 10 lærere er helt eller overvejende enige i, at dialog med eleverne om resultatet af de nationale test styrker elevernes motivation for at lære. 9 pct. af lærerne er helt eller overvejende enige i, at dialog med eleverne om de nationale test direkte mindsker elevernes motivation for at lære.

Desuden viser tabellen, at ca. hver tredje lærer er helt eller overvejende enige i, at tilbagemelding og dialog med eleverne om resultatet af de nationale test medvirker til, at eleverne i højere grad inddrages i formuleringen af fremtidige læringsmål.

Endelig viser tabellen, at 36 pct. af lærerne er helt eller overvejende uenige i, at eleverne generelt er glade for de nationale test. 23 pct. af lærerne er helt eller overvejende enige i, at eleverne generelt er glade for de nationale test.

5.1.1 Lærernes tilbagemelding til eleverne og opfølgning på resultaterne

I spørgeskemaundersøgelsen er lærerne tillige blevet bedt om at tilkendegive, på hvilken måde og i hvilken grad de giver eleverne tilbagemelding om resultaterne af en national test. Resultatet fremgår af nedenstående tabel.

Tablet 5.2: I hvilken grad stemmer følgende udsagn overens med din opfølgning på resultatet af de nationale test i relation til den enkelte elev? Angivet i pct.

	Slet ikke	I mindre grad	I nogen grad	I temmelig høj grad	I meget høj grad	I alt (N)
Jeg orienterer eleverne mundtligt om resultatet umiddelbart efter testen	15	14	22	28	22	100 (2162)
Jeg har en dialog med hver enkelt elev om resultatet	8	14	25	31	23	100 (2162)
I forbindelse med dialogen med hver enkelt elev om resultatet, opstiller vi i fællesskab fremadrettede faglige mål	13	21	37	20	9	100 (2162)

Kilde: Spørgeskemaundersøgelse til lærere, 2013.

Tabellen viser, at halvdelen af lærerne i temmelig eller i meget høj grad orienterer eleverne mundtligt om resultatet umiddelbart efter testen. 15 pct. af lærerne tilkendegiver, at de slet ikke orienterer eleverne mundtlig om resultatet, mens 14 pct. gør dette i mindre grad. Godt halvdelen af lærerne vurderer, at de i temmelig eller i meget høj grad har en dialog med hver enkelt elev om testresultatet. For 8 pct. af lærerne gælder, at de slet ikke har dialog med den enkelte elev om resultatet, mens 14 pct. angiver, at det foregår i mindre grad. Endelig viser tabellen, at tre ud af 10 lærere vurderer, at de i temmelig eller i meget høj grad opstiller fælles fremadrettede faglige mål i samarbejde med eleven, mens 13 pct. angiver, at dette slet ikke sker, og 21 pct., at det sker i mindre grad.

Med henblik på at undersøge, om der er forskel på typen og graden af lærernes tilbagemelding til eleverne fra fag til fag, er lærernes besvarelse af ovenstående spørgsmål sammenholdt med, hvilke(t) testfag de underviser i. Det fremgår, at 71 pct. af engelsklærerne vurderer, at de i temmelig eller meget høj grad har dialog med hver enkelt elev om testresultaterne. Til sammenligning vurderer godt halvdelen af matematiklærerne og dansklærerne, at de i temmelig eller i meget høj grad har dialog med hver enkelt elev om testresultaterne. For naturfagene forholder det sig sådan, at 43 pct. af geografilærerne, 47 pct. af fysik/kemilærerne og 46 pct. af biologilærerne har dialog med eleverne om testresultatet. Der er således ikke de store forskelle på tværs af fag – dog med den undtagelse, at engelsklærerne i højere grad angiver at have dialog med den enkelte om testresultaterne.

I casestudierne er eleverne blevet bedt om at beskrive, hvordan og i hvilken grad de har dialog med lærerne om resultaterne af de nationale test. Hovedparten af eleverne oplever, at deres lærere sammen med hele klassen gennemgår klassens resultater. Eleverne fortæller typisk, at læreren i det enkelte fag bruger (dele af) en lektion på at redegøre for klassens samlede resultater, herunder hvilke emner eleverne klarer sig godt og mindre godt i. Der er dog også eksempler fra casestudierne på, at elever ikke kan huske, at deres lærer har gennemgået resultaterne af de nationale test.

I forlængelse af ovenstående viser casestudierne videre, at kun en mindre del af eleverne har oplevet af få individuel tilbagemelding fra deres lærer om resultatet af en national test. De elever, der har haft én-til-én dialog med deres lærer om testresultater, forklarer typisk, at deres lærer sætter en lektion af til en kort samtale med eleverne om testresultaterne. Ifølge eleverne handler samtalerne især om, hvad de skal gøre bedre, og hvad de særligt skal øve sig på. Casestudierne tegner endvidere et billede af, at individuel tilbagemelding fra lærer til elev om resultater af nationale test især forekommer blandt elever i udskoling.

Også lærerne er i forbindelse med casestudierne blevet bedt om at beskrive deres praksis for tilbagemelding om resultater af nationale test til eleverne. Lærerne bekræfter i høj grad elevernes billede heraf. De tilkendegiver ofte, at de især giver tilbagemelding om resultaterne på klasseniveau, mens de sjældent eller aldrig giver individuel tilbagemelding til eleverne. Lærerne begrunder især den fraværende dialog med den enkelte elev med praktiske udfordringer. Lærerne fortæller, at tid er en afgørende faktor, og at tilbagemelding til eleverne ville kræve, at andet blev nedprioriteret og/eller udeladt af undervisningen. Derudover nævner flere lærere, at i de tilfælde, hvor de er eneste lærer/voksen i klasserummet, er det vanskeligt at gennemføre den individuelle dialog, da det vil betyde, at resten af klassen skal være i klassen uden opsyn. Som det fremgår ovenfor, indikerer casestudierne, at individuelle samtaler om resultater af nationale test især gennemføres i udskoling, hvilket ifølge lærerne netop skyldes, at de ældre elever nemmere kan lades alene i klassen.

Som det fremgår af tabellen ovenfor, angiver omkring en tredjedel af lærerne, at de i forbindelse med dialogen med hver enkelt elev om resultatet opstiller fremadrettede faglige mål, mens en tilsvarende andel tilkendegiver, at dette slet ikke sker eller sker i mindre grad. Yderligere analyser viser, at det særligt er engelsklærerne, der vurderer, at de i temmelig eller i meget høj grad opstiller læringsmål med eleven i forlængelse af dialogen om testresultatet. Dette billede kan kun i begrænset omfang bekræftes gennem viden fra casestudierne. Ganske få elever har oplevet, at de sammen med læreren opstiller mål for deres arbejde i et fag. De elever, der tilkendegiver at have haft en sådan dialog med en lærer, forklarer typisk, at dialogen består i, at læreren fortæller, hvilke emner eleven bør fokusere (mere) på.

Boksen nedenfor består af eksempler fra casestudierne på lærernes dialog med eleverne om resultater af de nationale test set fra elevernes perspektiv.

Boks 5.1: Eksempler på dialog om resultater af de nationale test

Dialog om resultater med den enkelte elev

De elever, der fortæller, at de har dialog med lærerne om resultaterne af en national test fortæller, at læreren typisk tager dem ud enkeltvis og taler om, hvad der gik godt og mindre godt i testen. Eleverne fortæller, at de til samtalen får at vide, om de ligger under middel, middel eller over middel, og hvad det betyder. Eleverne fortæller videre, at dialogen med lærerne, når den finder sted, primært drejer sig om, hvordan man kan forbedre sig frem mod næste test. En elev fortæller, at man først taler om, hvordan det gik sidste gang, og derefter hvordan det er gået i de seneste test. En anden elev fortæller videre, at hans samtale med læreren om testresultaterne efterfølgende medførte, at han fik en lang liste med øvelser, som han kunne gennemføre sammen med sine forældre frem mod næste test.

Dialog om resultaterne med en klasse

De elever, der oplever, at dialog om testen primært foregår i klassen, fortæller, at lærerne eksempelvis viser elevernes samlede resultater på smartboard. Eleverne fortæller, at der ikke er navne på, men at det er for, at eleverne kan se, hvor klassen ligger samlet. Lærerne fortæller bagefter, hvis der er et specielt emne, som klassen bliver nødt til at træne mere.

På baggrund af ovenstående er det evaluators vurdering, at eleverne i høj grad modtager kollektiv tilbagemelding om resultater af de nationale test fra læreren. Det er ligeledes evaluators vurdering, at eleverne kun i begrænset omfang modtager individuel tilbagemelding fra lærerne, lige-

som der i denne forbindelse og i fællesskab med læreren sjældent opstilles faglige mål. Begge observationer udfordrer grundlæggende elementer i evalueringens forandringsteori vedrørende dialogen mellem lærer og elev om resultatet af testene.

5.1.2 De nationale tests betydning for elevernes forståelse af egne styrker og svagheder

Som det fremgår af tabel 5.1, fordeler lærerne sig nogenlunde ligeligt i deres vurdering af, hvorvidt de nationale test styrker elevernes forståelse af egne styrker og svagheder, om end der er flest, der forholder sig positivt til dette.

I forbindelse med casestudierne er eleverne ligeledes blevet bedt om at forholde sig til dette. Deres udsagn understøtter i høj grad lærernes vurdering, idet der også blandt eleverne er forskel på deres vurdering af betydningen heraf. En gruppe af elever tilkendegiver, at resultaterne af de nationale test i høj grad bidrager til at tydeliggøre deres egne styrker og svagheder i de enkelte fag. De forklarer, at de både i testsituationen, og når de får resultaterne, bliver opmærksomme på, hvad de er gode og mindre gode til. Flere elever fortæller, at de godt kan mærke, hvornår opgaverne falder dem lette eller svære, og at de bruger dette til at vurdere deres eget faglige niveau i faget. Der er dog også en gruppe af elever, som tilkendegiver, at de nationale test, herunder både gennemførelse og resultater, fylder ganske lidt i deres bevidsthed. Denne gruppe elever forklarer typisk, at de gennemfører testene, "fordi de skal", men at de ikke anvender eller reflekterer over resultater af testene.

Hvorvidt eleverne tilhører den eller den anden gruppe, afhænger af den tilbagemelding, som de modtager fra lærerne om resultaterne. Det er evaluators vurdering, at de elever, der udelukkende modtager kollektiv orientering om resultaterne af de nationale test, *alene* har forståelse af egne styrker og svagheder på et generelt og overordnet niveau. I forlængelse heraf vurderer evaluator, at de elever, der modtager individuel tilbagemelding fra lærerne, får mere specifik og emnenær viden om egne styrker og svagheder. Eleverne tilkendegiver typisk, at i de tilfælde, hvor læreren forklarer, hvilke områder eleven kan forbedre sig på, er det således nemmere for eleven at have fokus herpå fremadrettet.

Casestudierne tegner dog også et billede af, at elevernes vurdering af, hvorvidt de nationale test bidrager til forståelse af egne styrker og svagheder, afhænger af deres opfattelse af testenes validitet. Også her deler eleverne sig i to grupper. En gruppe elever tilkendegiver, at de betragter de nationale test som mere troværdige og objektive set i forhold til andre test. Eleverne forklarer dette med, at de nationale test rettes "af computeren" og ikke af læreren. Dette er ifølge eleverne værdifuldt, fordi lærerne kan overse korrekte svar og/eller være påvirket af deres generelle opfattelse af den enkelte elev. En anden gruppe elever udtrykker dog skepsis over for resultatet af testene, fordi de har oplevet at rykke sig meget fra de frivillige til de obligatoriske test. Nogle elever fortæller eksempelvis, at de har rykket sig fra langt under middel til over middel. Eleverne har både eksempler på at rykke sig i positiv og negativ retning. Eleverne betragter dette som udtryk for, at de nationale test ikke tegner et korrekt billede af deres faglige udvikling, herunder styrker og svagheder.

På tværs af datakilder tegner evalueringen således et billede af, at de nationale test i nogen grad bidrager til at bevidstgøre eleverne om egne styrker og svagheder. Det er dog evaluators vurdering, at denne bevidstgørelse er størst i de tilfælde, hvor læreren har individuel dialog med eleverne om resultaterne og i den forbindelse forklarer og drøfter elevens styrker og svagheder.

5.1.3 De nationale tests betydning for elevernes (motivation for) læring

I forlængelse af ovenstående har evalueringen sat fokus på, hvorvidt lærernes dialog med eleverne om resultatet af de nationale test mindsker eller styrker elevernes motivation for at lære.

Som det fremgår af tabel 5.1, vurderer lidt under halvdelen af lærerne i spørgeskemaundersøgelsen, at de hverken er enige eller uenige i, at dialog med eleverne om resultaterne af de nationale test styrker eller mindsker elevernes motivation for læring. Knap en tredjedel af lærerne er enige eller helt enige i, at denne dialog styrker elevernes motivation for læring.

Også i forbindelse med casestudierne er lærerne blevet bedt om at forholde sig til sammenhængen mellem dialog om resultaterne af de nationale test og elevernes motivation for læring. Læ-

ernes udsagn bekræfter i høj grad resultaterne af spørgeskemaundersøgelsen. En stor del af lærerne vurderer, at der ikke er en sammenhæng mellem dialog om resultater af nationale test og elevernes motivation for læring. De forklarer, at eleverne ikke i særlig høj grad tillægger de nationale test værdi, og/eller at testene ikke fylder meget i elevernes bevidsthed og således heller ikke påvirker deres motivation for læring.

Eleverne er som led i casestudierne blevet bedt om at vurdere, hvorvidt de nationale test styrker dem fagligt. Overordnet set deler eleverne sig i to grupper i forhold til dette spørgsmål. En gruppe elever mener, at de bliver fagligt dygtigere af at tage nationale test, blandt andet fordi de i nogle fag øver sig meget op til testen. Eleverne i denne gruppe fortæller typisk, at de oplever at blive bedre til udvalgte fagområder i et testfag op til en national test, fordi læreren underviser mere intensivt i pågældende område(r). Den anden gruppe elever vurderer, at de nationale test ikke har betydning for deres faglige niveau. Disse elever peger ofte på den manglende mulighed for at få konkretiseret, i hvilke opgaver de har lavet fejl. Eleverne sammenligner gerne nationale test med andre typer af test og fremhæver, at andre typer af test har den fordel, at de kan se deres korrekte såvel som ukorrekte besvarelser. De fortæller, at det er irriterende, at de i de nationale test ikke kan se deres ukorrekte svar. Flere elever efterlyser i denne forbindelse (mere) dialog med lærerne om resultaterne.

Som ovenfor vurderer evaluator, at det afhænger af lærerens dialog med eleven, hvorvidt eleverne tilhører den ene eller den anden gruppe. I de tilfælde, hvor lærerne har individuel dialog med eleverne om resultaterne, kan eleverne i højere grad forstå og anvende viden om egne styrker og svagheder fremadrettet.

5.1.4 Elevernes oplevelse af testsituationen

I forbindelse med casestudierne er eleverne blevet bedt om at beskrive, hvordan de oplever selve testsituationen.

Størstedelen af eleverne fortæller, at de er rimelig afslappede i testsituationen. Mange af eleverne fortæller, at de godt kan være lidt nervøse op til en testafvikling, men at nervøsiteten ofte forsvinder, når de først er i gang med at gennemføre testen.

Evaluator har som led i casestudierne gennemført observation af testafviklingerne. Disse bekræfter i høj grad elevernes beskrivelse af situationen. For hovedparten af de testafviklinger, evaluator har observeret, gælder, at eleverne var stille og koncentrerede om opgavebesvarelsen. Der er dog også enkelte eksempler på, at en elev er blevet ked af det og eventuelt har forladt lokalet. I de situationer har læreren forsøgt at berolige eleven, og i de fleste tilfælde er eleven vendt tilbage og har færdiggjort testen.

Casestudierne viser endvidere en tendens til, at nervøsiteten forud for en test er størst blandt elever i indskolingen. Indskolings eleverne fortæller, at de kan være nervøse for at skulle gennemføre en test, fordi de er bange for ikke at kunne svare på spørgsmålene, og/eller fordi de bare ikke bryder sig om selve situationen. Elever på mellemtrinnet og i udskolingen giver udtryk for, at de efterhånden har gennemført så mange nationale test, at de godt ved, hvad det indebærer. Usikkerheden i forhold til at gennemføre nationale test er derfor langt mindre for disse elever.

Det adaptive princip

I forbindelse med casestudierne er eleverne endvidere blevet bedt om at tilkendegive, om de kender (betydningen af) det adaptive princip, og hvordan det indvirker på deres oplevelse af testsituationen.

Hovedparten af eleverne ved godt, at opgaverne i testsituationen tilpasser sig deres faglige niveau, og at ikke alle i klassen får samme spørgsmål og antal opgaver. De er således bekendte med det adaptive princip, om end det ikke er alle, der kender den betegnelse. Casestudierne viser videre, at elevernes viden herom går på tværs af klassetrin, og således er mange af eleverne i indskolingen også bekendte med princippet. Eleverne, uanset klassetrin, fortæller, at lærerne forklarer dem dette forud for en test. De fremhæver, at læreren har forklaret, at antallet og typen af opgaver vil være forskellig fra elev til elev, og at dette ikke betyder, at nogle klarer sig

dårligere eller bedre end andre. Dog viser casestudierne også, at selvom eleverne *i princippet* forstår betydningen af det adaptive princip, påvirker det i høj grad deres oplevelse af testsituationen – og i mange tilfælde i negativ retning.

Eleverne fortæller i casestudierne, at en af udfordringerne ved det adaptive princip er, at de ikke nødvendigvis bliver færdige med testen på samme tidspunkt. På tværs af casestudier fremhæver eleverne deres frustration over at sidde tilbage blandt de sidste under selve testafviklingen. Det fylder meget hos eleverne på alle klassetrin. Det varierer fra skole til skole, hvorvidt lærerne er lykkedes med at forklare eleverne, at det at sidde tilbage blandt de sidste ikke nødvendigvis er ensbetydende med, at man er fagligt svag. I boksen nedenfor er fremhævet nogle eksempler på dette med elevernes egne ord.

Boks 5.2: Eksempler på elevernes oplevelse af at gennemføre testen i forskelligt tempo

En elev fra 2. klasse fortæller om testsituationen: *”Det bliver bare ved og ved... Der kommer hele tiden nye spørgsmål, og nogle gange kan jeg ikke svare. Det er ikke så sjovt. Og jeg sidder altid tilbage til sidst, men nogle gange sammen med [elev]. [Lærer] siger, at vi bare skal svare så godt, vi kan. Og der til sidst, når alle andre er færdige, vil man jo gerne bare ud til dem. [Elev] og [elev] er jo gået i gang med at lege, men så kan jeg ikke nå at være med.”*

En elev fra 4. klasse udtrykker frustration over at være en af de sidste, der bliver færdig med at gennemføre en national test: *”Gid det var mig, der kunne blive hurtigt færdig og ikke sidde alene. ”Find nu mit niveau”, siger jeg til computeren. Så bliver jeg rimelig træt af det. Jeg vil hellere bare blive færdig så hurtigt som muligt.”* Læreren til pågældende elev har forklaret eleverne i klassen, at det ikke er dårligt at sidde tilbage som de sidste, men til trods herfor oplever eleven ubehag ved situationen. Eleven fortæller videre: *”Man prøver jo at koncentrere sig og så går de andre pludselig – så tænker man ’videre, videre’ – det er ret stressende”,* fortæller eleven og afslutter: *”Jeg skynder mig bare, når de andre er færdige.”*

En elev fra 8. klasse bekræfter ovenstående billede: *”Det er træls, hvis man sidder tilbage der som den sidste, helt lonely. Der kan man få stress, hvis man sidder tilbage. [Lærer] har jo sagt, at vi bare skal give os god tid, men det er altså ikke sjovt til sidst. Man vil jo hellere ud til de andre. Og man vil jo heller ikke have, at de tænker, at man er dum eller sådan noget.”*

Eleverne fremhæver også styrker ved det adaptive princip. Flere elever forklarer i forbindelse med casestudierne, at det er motiverende, at spørgsmålene tilpasser sig deres niveau. De tillægger dette princip den værdi, at resultatet heraf er mere retvisende end øvrige, ofte papirbaserede test, hvorfor de gør sig umage for at svare rigtigt.

Mange af eleverne i casestudierne fortæller desuden, at de ved, hvornår de får lettere og sværere opgaver. Flere elever fortæller, at det presser dem, når de kan mærke, at opgaverne bliver lettere, idet de antager, at det betyder, at de har svaret forkert på foregående opgaver. Eleverne forklarer, at dette stresser dem i testsituationen.

Det er evaluators vurdering, at eleverne i høj grad er bekendte med det adaptive princip, men at de til trods herfor i mange tilfælde oplever ubehag ved testsituationen, særligt i de tilfælde, hvor de sidder tilbage blandt de sidste.

En it-baseret test

Eleverne blev i casestudierne bedt om at beskrive, hvad det betyder for deres oplevelse af testsituationen, at nationale test er it-baserede. Eleverne fremhæver både styrker og svagheder herved.

Som styrker fremhæver eleverne, at det er motiverende at gennemføre test på computeren. Eleverne peger på, at det på computeren er nemmere at koncentrere sig om sin egen test, eftersom det ikke er muligt at se, hvor langt dine kammerater er nået. Eleverne sammenligner det med en papirbaseret test, hvor de typisk godt kan se, hvor langt sidemanden er nået.

Som udfordringer fremhæver eleverne især, at det ikke er muligt at gå tilbage og rette en fejl, ligesom det er tilfældet ved en papirbaseret test. Mange elever tilkendegiver, at det er frustrerende og demotiverende at opdage, at man har svaret forkert på et spørgsmål, og at man ikke har mulighed for at gå tilbage og rette denne fejl. Nogle af eleverne fortæller, at det i testsituationen kan påvirke dem så meget, at de har svært ved at koncentrere sig om resten af opgaverne. En anden udfordring, som fremhæves af eleverne, er, at den it-baserede test betyder, at de ikke har overblik over mængden af opgaver og derfor ikke kan disponere deres tid i testsituationen. De forklarer typisk, at ved papirbaserede test orienterer de sig først i den samlede test ved at kigge alle spørgsmål igennem, inden de går i gang med besvarelsen. De beskriver den manglende mulighed herfor i de nationale test som irriterende.

På baggrund af ovenstående er det evaluators vurdering, at eleverne overordnet set er afslappede i forbindelse med gennemførelse af en national test, men at de i nogen grad finder det stressende, hvis de er blandt de sidste til at afslutte testen. Dette til trods for at eleverne i mange tilfælde er bekendte med, hvad det adaptive princip indebærer.

5.2 Skole-hjem-samarbejde

I dette afsnit sætter evaluator fokus på de nationale tests rolle og betydning for skole-hjem-samarbejdet.

Bestemmelserne om anvendelsen af nationale test i forbindelse med skole-hjem-samarbejdet er beskrevet i lovforslaget om de nationale test samt i bekendtgørelse om elevplaner, elev- og uddannelsesplaner samt uddannelsesplaner i folkeskolen⁷³. Af førstnævnte fremgår det, at de nationale test er et pædagogisk værktøj, der bl.a. skal bidrage til at underrette forældre med henblik på at tilrettelægge et forældresamarbejde, der understøtter eleven bedst muligt. Det fremgår desuden, at forældrene skal underrettes skriftligt om resultaterne af de nationale test. Sidstnævnte beskriver, at elevplaner skal indeholde oplysninger om resultater af den løbende evaluering og den besluttede opfølgning. Dog fremgår det eksplicit, at for de nationale test gælder, at alene opfølgningen (og ikke resultaterne) samt eventuelle aftaler om forældres og elevens medvirken til elevens opnåelse af de opstillede læringsmål skal indgå i elevplanen.

Både skoleledere og lærere er i spørgeskemaundersøgelsen blevet bedt om at forholde sig til anvendelse af de nationale test i skole-hjem-samarbejdet. Resultaterne heraf fremgår af følgende to tabeller.

Tabel 5.3: Du bedes angive, hvordan de nationale test indgår i skole-hjem-samarbejdet. Angivet i pct.

	Aldrig	Sjældent	Af og til/jævnligt	Ofte	Altid	I alt (N)
Orienteres forældre om, hvad testene måler, og hvad de skal bruges til inden testafviklingen?	0	8	24	24	45	100 (728)
Får forældrene en skriftlig tilbagemelding om resultatet af de nationale test?	1	3	9	18	69	100 (728)
Indgår resultatet af de nationale test, og lærerens fortolkning af dette, i en dialog med forældrene om fremadrettede læringsmål for deres barn?	0	3	20	40	36	100 (728)

Kilde: Spørgeskemaundersøgelse til skoleledere, 2013.

⁷³ Bekendtgørelse om elevplaner, elev- og uddannelsesplaner samt uddannelsesplaner i folkeskolen (BEK nr. 750 af 13/07/2009).

Tabel 5.4: Du bedes i følgende spørgsmål angive, hvordan de nationale test indgår i skole-hjem-samarbejdet. Angivet i pct.

	Aldrig	Sjældent	Af og til/jævnligt	Ofte	Altid (meget ofte)	I alt (N)
Orienteres forældrene inden testafviklingen om, hvad testene måler, og hvad de skal bruges til?	5	13	19	21	42	100 (2162)
Får forældrene resultatet af de nationale test sammen med elevplanen før skole-hjem-samtalen?	12	11	13	20	45	100 (2162)
Udleverer du resultatet af de nationale test ved selve skole-hjem-samtalen?	35	17	12	15	22	100 (2162)
Gennemgår du resultatet af de nationale test online i testsystemet ved selve skole-hjem-samtalen?	44	17	14	11	14	100 (2162)
Inddrager du resultatet af de nationale test i en dialog med forældrene om læringsmål for deres barn?	8	13	29	31	20	100 (2162)

Kilde: Spørgeskemaundersøgelse til lærere, 2013.

Af den øverste tabel fremgår det, at hovedparten (69 pct.) af skolelederne tilkendegiver, at forældrene ofte eller altid orienteres om, hvad testene måler, og hvad de skal bruges til, inden testafviklingen. Videre viser tabellen, at 87 pct. af skolelederne angiver, at forældrene altid eller ofte får en skriftlig tilbagemelding om resultatet af de nationale test, mens 76 pct. angiver, at resultatet og tolkningen heraf altid eller ofte indgår i dialog med forældrene om fremadrettede læringsmål for deres barn.

Den nederste tabel viser, at 63 pct. af de adspurgte lærere angiver, at forældre altid eller ofte orienteres om, hvad testene måler, og hvad de skal bruges til inden testafviklingen. Tabellen viser videre, at to tredjedele af lærerne angiver, at forældrene får resultatet af de nationale test sammen med elevplanen før skole-hjem-samtalen. For 37 pct. af lærerne gælder, at de udleverer resultatet af de nationale test ved selve skole-hjem-samtalen, mens omkring halvdelen (52 pct.) tilkendegiver, at dette sjældent eller aldrig er tilfældet. Endelig viser tabellen, at flertallet af de adspurgte lærere (61 pct.) angiver, at de sjældent eller aldrig gennemgår resultater af de nationale test online i testsystemet ved skole-hjem-samtalen, mens en fjerdedel angiver, at dette altid eller ofte sker. Ca. halvdelen (51 pct.) af lærerne tilkendegiver, at de inddrager resultatet af de nationale test i en dialog med forældrene om læringsmål for deres barn. 21 pct. angiver, at dette sjældent eller aldrig er tilfældet.

Sammenholder man skolelederes og læreres besvarelser af spørgeskemaundersøgelsen, adskiller de to gruppers besvarelser sig især i forhold til, hvorvidt resultatet af de nationale test inddrages i dialog med forældrene om læringsmål for deres barn. Her tilkendegiver hhv. 87 pct. af skolelederne og 51 pct. af lærerne, at dette er tilfældet. Skolelederes og læreres besvarelser angående, hvorvidt forældre orienteres om, hvad testene måler, og hvad de skal bruges til, fordeler sig nogenlunde ens. Dog er der lidt flere lærere (18 pct.) end skoleledere (8 pct.), der tilkendegiver, at dette aldrig eller sjældent sker. Sammenholder man skolelederes og læreres besvarelser i relation til skriftlig tilbagemelding om resultaterne af de nationale test til forældrene, er skolelederne generelt mere positive i forhold til hyppigheden heraf, end lærerne er. Der er dog 13 pct. af skolelederne, der vurderer, at dette kun sker af og til, sjældent eller aldrig.

Det er evaluators vurdering, at ovenstående tegner et billede af, at (resultaterne af) de nationale test i nogen grad indgår i forældresamarbejdet, men at der er stor variation skolerne og lærerne imellem. Ydermere er det evaluators vurdering, at spørgeskemaundersøgelsen peger på, at resul-

taterne af de nationale test i mange tilfælde ikke på systematisk vis inddrages i skole-hjem-samarbejdet, herunder i forhold til opstilling af læringsmål.

I det følgende uddybes og sammenholdes resultaterne af spørgeskemaundersøgelserne med de kvalitative data, indsamlet via casestudier.

Orientering af forældre

I forbindelse med casestudierne er lærerne blevet bedt om at beskrive skolens praksis i forhold til, om og hvordan forældrene orienteres om (gennemførelsen af) nationale test. Disse beskrivelser nuancerer det billede, der tegnes via spørgeskemaundersøgelsen, jf. ovenfor. Samtlige lærere tilkendegiver, at de orienterer forældrene om tidspunktet for gennemførelsen af en national test forud herfor. Orienteringen foregår typisk skriftligt via forældreintra. I forlængelse heraf peger flere af de adspurgte lærere på, at de nogle gange supplerer den skriftlige information med en mundtlig orientering. Dog tilkendegiver de samtidig, at dette typisk kun foregår i de tilfælde, hvor orienteringen om gennemførelse af en national test falder sammen med et i forvejen planlagt forældremøde.

En del af lærerne forklarer i forbindelse med casestudierne, at forældrenes informationsbehov er størst den/de først(e) gang(e), deres barn skal gennemføre en national test. Det er lærernes vurdering, at forældrene forud for en national test især er interesserede i den praktiske gennemførelse af de nationale test (herunder varighed; at der er tale om en digital test mv.), mens forældrene kun i begrænset omfang efterspørger viden om testenes faglige indhold. Dette betyder ifølge lærerne, at informationsbehovet hos forældrene daler i takt med, at deres børn får erfaring med at gennemføre nationale test.

Forældrene er i interviewene tillige blevet bedt om at beskrive den information, de modtager vedrørende de nationale test. Overordnet set stemmer dette overens med det billede, lærerne tegner. Forældrene forklarer typisk, at de ved den/de første test modtog skriftlig eller mundtlig information om formålet med de nationale test, praktisk gennemførelse mv., men at de ved efterfølgende nationale test udelukkende bliver orienteret om tidspunktet for gennemførelsen heraf. Hovedparten af forældrene tilkendegiver i forlængelse heraf, at dette informationsniveau er tilstrækkeligt. Flere forældre forklarer, at de betragter de nationale test som lærerens redskab, hvorfor de ikke har en forventning om yderligere information herom. Enkelte forældre fortæller, at de har gennemført en demotest med henblik på at få en fornemmelse af, hvad de nationale test består af.

Skolestyrelsen (nu Kvalitets- og Tilsynsstyrelsen) har udarbejdet en pjece med inspiration til pædagogisk brug af resultater fra de nationale test⁷⁴. Af denne fremgår det, at forældrene bør orienteres om, hvad testresultaterne måler, og hvordan de kan bruges over for klassen og den enkelte elev. Det er evaluators vurdering, at denne anbefaling i høj grad synes at være omsat til praksis i skolerne.

Med afsæt i ovenstående og på tværs af datakilder er det evaluators vurdering, at forældrene især orienteres om tidspunktet for gennemførelse af en national test, mens formål med og indhold af testene primært beskrives ved gennemførelse af test i de første skoleår. Evaluator vurderer i forlængelse heraf, at den sparsomme orientering på mellemtrinet og i udskolingen om, hvad testene måler, sjældent udgør en udfordring for forældrene. Endelig er det evaluators vurdering, at dialog om resultater og læringsmål kun i nogen grad indgår i skole-hjem-samarbejdet.

Resultat af de nationale test – tilbagemelding og opfølgning

Som det fremgår af tabel 5.3, tilkendegiver næsten alle skolelederne, at forældre får en skriftlig tilbagemelding på deres barns resultat i en national test. Som supplement hertil angiver flertallet af lærerne, at forældrene får resultatet af de nationale test sammen med elevplanen før skole-hjem-samtalen, mens færre udleverer resultatet af de nationale test ved selve skole-hjem-samtalen. Casestudierne bekræfter dette billede. Disse viser, at skolerne orienterer forældrene om deres barns resultater af de nationale test. Casestudierne viser videre, at den skriftlige orientering sker gennem det forældrebreve, der automatisk genereres ved gennemførelse af en natio-

⁷⁴ Brug testresultaterne – inspiration til pædagogisk brug af resultater fra de nationale test, Skolestyrelsen, 2011.

nal test. Forældrebreve sendes til forældrene eller udleveres i forbindelse med skole-hjem-samtaler. På baggrund af spørgeskemaundersøgelser og casestudier er det evaluators vurdering, at skolerne – i overensstemmelse med lovgivningen – skriftligt orienterer forældre om resultaterne af de nationale test.

Mange af de adspurgte lærere vurderer, at informationen i forældrebreve er utilstrækkelig og kun i begrænset omfang er meningsfuld i forhold til at give forældrene et billede af deres barns faglige niveau i et givent fag. Se desuden kapitel 8 om testsystemets tilbagemelding til lærerne.

Generelt set forholder de adspurgte forældre sig mere positivt til forældrebreve. Mange af forældrene vurderer, at informationen i brevene er forståelig, men påpeger samtidig, at det skal sættes i kontekst og uddybes, hvis forældrene skal kunne anvende resultaterne fremadrettet. Der er dog også enkelte forældre, der tilkendegiver, at de har svært ved at forstå indholdet af forældrebreve, ligesom enkelte forældre efterlyser mere opfølgning på testresultaterne og yderligere forklaring i forældrebreve. Forældrene beskriver typisk, at de orienterer sig i resultaterne af de nationale test, men at det generelt ikke fylder meget.

Lærerne er, som led i spørgeskemaundersøgelsen, blevet bedt om at tilkendegive, i hvilken grad de prioriterer de nationale test i dialogen ved skole-hjem-samtaler. 44 pct. angiver, at dette i nogen grad prioriteres, mens 20 pct. angiver, at det er meget høj eller i temmelig høj grad prioriteres. Dette billede bekræftes i høj grad af viden fra casestudierne. Både lærere og forældre tilkendegiver, at resultaterne af en national test ofte gennemgås i forbindelse med skole-hjem-samtaler, da der, jf. ovenstående, typisk er behov for en mundtlig forklaring/udddykning heraf. Der er dog også flere forældre, der peger på, at resultaterne af de nationale test kun drøftes, hvis de eksplicit beder læreren om det.

For at undersøge om der er forskel på prioriteringen af de nationale test i dialogen med forældre fra fag til fag, er lærernes angivelse af prioritering sammenholdt med, hvilke testfag de underviser i. Denne analyse indikerer, at dialogen om de nationale test prioriteres højest i dansk (25 pct.), engelsk (25 pct.) og matematik (22 pct.), mens den prioriteres lavere i geografi (9 pct.), fysik/kemi (9 pct.) og biologi (8 pct.). Tallene i parentes angiver, hvor stor en andel af lærerne, der har angivet, at de prioriterer dette i meget eller i temmelig høj grad. Det skal dog her bemærkes, at samtlige lærere i spørgeskemaundersøgelsen er blevet bedt om at forholde sig til dette spørgsmål. Det betyder, at også lærere, der ikke nødvendigvis afholder skole-hjem-samtaler, er med i analysen. Det betyder således også, at resultatet skal tolkes med varsomhed.

Dog bekræfter casestudierne ovenstående. Her fremgår det også, at hvorvidt resultaterne af de nationale test drøftes med forældre til skole-hjem-samtaler er afhængig af, om læreren for det/de pågældende testfag deltager ved skole-hjem-samtalen. Det er evaluators vurdering, at dette i praksis betyder, at det typisk er resultater af nationale test i fag som dansk, matematik og evt. engelsk, der gennemgås ved skole-hjem-samtaler, mens resultaterne af nationale test i fagene fysik/kemi, biologi og geografi sjældent eller aldrig drøftes i forbindelse med skole-hjem-samtaler. Evalueringen påpeger således en reel informationsudfordring til hjemmet (egentlig dialog) i de fag, hvor lærerne ikke er repræsenteret ved skole-hjem-samtaler.

Casestudierne viser, at når resultaterne af de nationale test drøftes ved skole-hjem-samtaler, anvendes de til at give forældrene en status for deres barns faglige niveau i pågældende fag. De adspurgte lærere forklarer dog samstemmende, at denne status baserer sig på en samlet vurdering af elevens faglige niveau, som fremkommer ved hjælp af flere evalueringsmetoder og evalueringsredskaber, herunder andre typer af test, resultater ved hjemmeopgaver, elevens daglige indsats mv. Lærerne understreger således, at resultaterne af de nationale test aldrig står alene. Det er evaluators vurdering, at dette er det generelle billede, hvilket understøttes af analysen i afsnit 4.1. Sidstnævnte bekræftes i høj grad af forældrene. De forklarer typisk, at i lærerens vurdering af deres barns faglige niveau indgår resultater af de nationale test på lige fod med resultater af andre test og lærerens mere generelle vurdering af eleven.

Af casestudierne fremgår det endvidere, at mange lærere forsøger at koordinere tidspunktet for gennemførelsen af en national test og afholdelse af skole-hjem-samtaler, så det er muligt at tale om resultaterne af testen i forbindelse med skole-hjem-samtalen.

5.2.1 Inddragelse af hjemmet i dialog om læringsmål

Som det fremgår af tabel 5.4, angiver ca. halvdelen af lærerne, at de inddrager resultatet af de nationale test i dialog med forældrene om læringsmål for deres barn. I forbindelse med casestudierne er forældrene tillige blevet bedt om at beskrive deres erfaringer hermed, og deres udsagn tegner et noget anderledes billede. Langt hovedparten af forældrene fortæller, at de ikke har oplevet, at der er lavet aftaler om opfølgning og/eller opstillet læringsmål for deres barn på baggrund af resultater af nationale test (og andre typer af test/evaluering). Kun ganske få forældre har oplevet, at der er blevet opstillet læringsmål og/eller iværksat tiltag med afsæt i resultater af de nationale test. Både lærernes svar i spørgeskemaundersøgelsen samt i særdeleshed feedback fra forældrene indikerer, at det langt fra er i alle tilfælde, at resultaterne af de nationale test giver anledning til en dialog om fremadrettede læringsmål for eleverne.

Det er evaluators vurdering, at den i nogle tilfælde begrænsede dialog mellem skole og hjem om resultater af de nationale test i høj grad begrænser forældrenes mulighed for at anvende disse til at understøtte deres barns læring, som det er intentionen i lovgivningen. Evaluator vurderer i forlængelse heraf, at den begrænsede dialog udfordrer grundlaget for og dermed indfrielse af evalueringens hypotese om samme. Ydermere adskiller dette sig fra anbefalingen i ministeriets pjece, hvor det fremgår, at testresultaterne kan være udgangspunkt for en dialog med forældre om læringsmål for den enkelte elev⁷⁵. Evaluator vurderer, at evalueringens hypotese om, at dialog med hjemmet om resultatet af de nationale test fremmer forældrenes mulighed for at understøtte deres barns læring, kun i nogen grad bekræftes. Det skal dog understreges, at evaluator ligeledes vurderer, at der synes at være betydelig variation skolerne og lærerne imellem i forhold til praksis herfor, hvilket vidner om et uudnyttet potentiale.

Som tidligere nævnt skal opfølgning på resultater af de nationale test samt eventuelle aftaler om forældres og elevens medvirken til elevens opnåelse af de opstillede læringsmål indgå i elevplanen. Det fremgår af tabel 5.4, at flertallet af lærerne tilkendegiver, at forældrene altid eller ofte får resultatet af de nationale test sammen med elevplanen før skole-hjem-samtalen. Ved casestudierne er forældrene blevet bedt om at forholde sig til dette, og – ligesom det er tilfældet med opstilling af læringsmål – tegner casestudierne et andet billede end det, lærernes besvarelse af spørgeskemaet gør. Størstedelen af forældrene tilkendegiver, at elevplanen for deres barn ikke indeholder information om opfølgning på resultater af nationale test. De interview, der er gennemført med lærere som led i casestudierne, bekræfter, at ikke alle lærere forholder sig til opfølgning på resultater af de nationale test i elevplanen.

Med afsæt i flere datakilder er det således evaluators vurdering, at opfølgning på resultater af de nationale test ikke på alle skoler systematisk indskrives i elevernes elevplan. Det skal nævnes, at det af ministeriets pjece fremgår, at den skriftlige tilbagemelding til forældrene om barnets testresultater kan lægges ind i elevplanen, om end det præciseres, at de nationale test er ét ud af mange evalueringsværktøjer, hvorfor de kun vil fylde en lille del. Spørgeskemaundersøgelse såvel som casestudier peger dog på, at dette i nogle tilfælde slet ikke sker.

5.2.2 Testresultaternes værdi for forældre i skole-hjem-samarbejdet

Som det fremgår indledningsvis, er en hypotese i evalueringen, at de nationale test styrker skole-hjem-samarbejdet. Som led i casestudierne er både lærere og forældre blevet bedt om at forholde sig eksplicit til dette og dermed give en samlet vurdering af, om de nationale test styrker skole-hjem-samarbejdet. For både lærere og forældre gælder, at informanterne overordnet set deler sig i to grupper. Den ene gruppe vurderer, at de nationale test i en vis grad har styrket skole-hjem-samarbejdet. Dette begrundes især med, at resultaterne af de nationale test giver et konkret og objektivt grundlag at drøfte elevens faglige niveau ud fra. Derudover tilkendegiver de, at resultaterne af de nationale test giver dem et grundlag for at sammenligne den enkelte elev/deres barn med andre klasser, ligesom resultaterne giver et billede af elevens/barnets faglige udvikling over tid.

Den anden gruppe lærere og forældre vurderer, at de nationale test ikke bidrager til at styrke skole-hjem-samarbejdet. Særligt forældrene forklarer dette med, at lærerne ikke bruger det ak-

⁷⁵ Brug testresultaterne – inspiration til pædagogisk brug af resultater fra de nationale test, Skolestyrelsen, 2011.

tivt i samarbejdet med forældrene, og at de ser det som lærerens redskab frem for forældrenes. Lærerne påpeger typisk, at de især betragter de nationale test som et styringsredskab for skoleledelsen og forvaltningen, hvorfor det ikke har betydning for samarbejdet mellem skole og hjem.

Nedenstående boks giver eksempler på aspekter af de nationale test, der udgør en værdi for forældrene. De nævnte aspekter er fremhævet af forældre i forbindelse med casestudierne. Det er således evaluators vurdering, at der er tale om generelle tendenser, om end de i boksen er eksemplificeret ved fx én skole og/eller én forælder.

Boks 5.3: Eksempler på testresultaternes værdi for forældre i skole-hjem-samarbejdet

Øget fokus på de fagligt stærke elever

På en skole er der i kommende skoleår fokus på at udvikle de fagligt stærke elever. I casestudierne fortæller en forælder fra skolen, at forældrene er blevet gjort opmærksomme på dette fokus, og at der i hendes barns elevplan står, hvordan der skal arbejdes med barnets faglige udvikling. Forælderen er overbevist om, at de nationale test har medvirket til et øget fokus på de fagligt stærke og fortæller: "Forud for forældresamtalen er der målsætningssamtaler – her bliver der indgået kontrakter mellem lærer og elev. Her aftales, hvor der skal udfordres osv. Både toppen og bunden". Forældrene på pågældende skole vurderer, at de nationale test er med til at øge fokus på de fagligt stærke.

Lærerne er mere åbne for dialog

En del af forældrene oplever, at lærerne efter indførelsen af de nationale test er blevet mere åbne for en løbende dialog med forældrene om elevens faglige niveau. Forældrene fortæller i casestudierne, at det er deres opfattelse, at lærerne er trykke ved de nationale test, fordi de har noget konkret at tale ud fra. En forælder fortæller i relation hertil, at de nationale test til skole-hjem-samtalen kan medvirke til, at samtalen kan dreje sig om noget fagligt fremfor personligt. Dette fremhæves især som en fordel, hvis forældrene og/eller eleven har et mindre godt forhold til lærerne. Her kan de nationale test bruges som konkret omdrejningspunkt for skole-hjem-samtalen. Forældrene giver i den forbindelse udtryk for, at det er med til at styrke indholdet af skole-hjem-samtalen, at der inddrages objektive vurderinger af eleverne.

Sammenligningsgrundlag

En forældregruppe på en skole oplever, at nationale test er et godt redskab for forældrene til at følge deres børns faglige niveau over tid. Særligt læsetesten fremhæves af forældrene som et godt redskab til dette, da man med læsetesten har mulighed for at følge elevernes udvikling over de fire obligatoriske test. Desuden fremhæves den, fordi den kun har fokus på én færdighed, nemlig læsning, hvorfor det er nemt for forældrene at forholde sig til.

På baggrund af ovenstående er det evaluators vurdering, at hypotesen om, at dialog med hjemmet om resultatet af de nationale test fremmer forældrenes mulighed for at understøtte deres barns læring, kun i nogen grad kan bekræftes. Dette skyldes især, at denne dialog kun sker i begrænset omfang og typisk kun i nogle fag. Det er desuden evaluators vurdering, at hypotesen om, at de nationale test styrker skole-hjem-samarbejdet, kun delvist kan bekræftes.

5.3 Delkonklusion

I dette kapitel er der foretaget en analyse af de nationale tests betydning for eleverne og for skole-hjem-samarbejdet. I tabellen nedenfor er indsat en oversigt over de nøglehypoteser, der har været undersøgt i denne forbindelse.

Tabel 5.5: Oversigt over nøglehypoteser (eleven og skole-hjem-samarbejde)

Hypotese	Data	Status
Eleverne		
Lærerens tilbagemelding og dialog med eleven om resultatet af de nationale test styrker elevernes bevidsthed om egne læreprocesser og dermed motivation for læring	Spørgeskemaundersøgelse til lærere Casestudier, herunder interview og observation	Ikke bekræftet -
Skole-hjem-samarbejde		
Nationale test styrker skole-hjem-samarbejdet	Spørgeskemaundersøgelse til skoleledere og lærere Casestudier, herunder interview med forældre m.fl.	Delvist bekræftet +
Dialog med hjemmet om resultatet af de nationale test fremmer forældrenes mulighed for at understøtte deres barns læring	Spørgeskemaundersøgelse til skoleledere og lærere Casestudier, herunder interview med forældre m.fl.	Delvist bekræftet +

I forhold til **elevernes oplevelse af de nationale test** er evaluators samlede vurdering, at eleverne overordnet er positivt indstillede over for de nationale test. De it-baserede og adaptive test er generelt motiverende og udfordrende for eleverne, men der er også eksempler på, at testene skaber stress og usikkerhed hos visse elever. Forskellig praksis på skolerne vedrørende afslutning af testene, hvor nogle elever bliver færdige før andre, skaber en u hensigtsmæssig afvikling for nogle elever. En hovedkonklusion er endvidere, at eleverne hovedsageligt modtager kollektiv tilbagemelding fra lærerne om resultaterne af de nationale test, mens dialog med den enkelte elev om testresultater kun sker i begrænset omfang. Samtidig peger evalueringen på, at netop den individuelle dialog med eleven om resultater af testene er afgørende for testenes betydning for elevernes bevidsthed om egne styrker og svagheder samt deres motivation for at omsætte denne viden til handling.

Det er evaluators samlede vurdering, at de nationale test i nogen grad styrker **skole-hjem-samarbejdet**. Evalueringen viser, at forældrene generelt set orienteres om tidspunkt for gennemførelse af nationale test, samt at yderligere information om formål med, praktisk gennemførelse af de nationale test mv. typisk sker især den/de først(e) gang(e), eleverne skal gennemføre en national test. Forældrene modtager typisk skriftlig information om den enkelte elevs resultater gennem forældrebreve, som enten sendes hjem til forældrene eller udleveres ved skole-hjem-samtaler. En hovedkonklusion er, at resultaterne af de nationale test kun i begrænset omfang anvendes til at opstille læringsmål for eleverne (i dialog med hjemmet), ligesom opfølgningen herpå kun i nogen grad skrives ind i elevplanerne. Der synes således at være potentiale for at inddrage resultater af de nationale test til opstilling af læringsmål i højere grad, end tilfældet er i dag.

6. DE NATIONALE TEST SOM DIALOG- OG STYRINGSREDSKAB

I det følgende kapitel behandles den del af evalueringen, der omhandler anvendelsen af de nationale test som værktøj for kommuners og skolelederes kvalitetsarbejde. Dette er endnu en dimension, relateret til evalueringstemaet *Testenes anvendelse*.

Som tidligere fremhævet blev de nationale test indført med pædagogisk anvendelse som det primære formål (jf. lovforslaget om de nationale test). Imidlertid var det tilsvarende en politisk ambition, at de nationale test skulle anvendes af både skoleledere og kommuner i deres kvalitetsarbejde på henholdsvis den enkelte skole og for det samlede skolevæsen⁷⁶.

Den del af forandringsteorien, der afdækkes og analyseres i dette kapitel, kalder vi for *leder- og styringssporet*. Der skal således foretages en vurdering af, i hvilket omfang og hvordan de nationale test anvendes af henholdsvis *skoleledere* og *kommuner* i deres respektive kvalitetsarbejde, herunder om anvendelsen bidrager til de forventede forandringer og resultater.

6.1 Anvendelse af de nationale test som værktøj for skolelederes kvalitetsarbejde

Som pædagogisk og administrativ ansvarlig for skolens samlede virksomhed har **skolelederen** en central rolle i skolens kvalitetsarbejde, herunder arbejdet med evaluering. I henhold til lovgrundlaget skal skoleledelsen orientere sig i resultaterne af de nationale test, og som det fremgår af såvel model for evalueringsskema som forandringsteori, antages det, at ledelsen går i dialog med lærerne om resultaterne med henblik på fremadrettet handling.

I det følgende afrapporteres evalueringens resultater, for så vidt angår skoleledernes kvalitetsarbejde. Resultaterne er baseret på en spørgeskemaundersøgelse udsendt til samtlige skoleledere i de danske folkeskoler. Over 700 skoleledere har deltaget i spørgeskemaundersøgelsen. Derudover anvendes de gennemførte casestudier, hvor der er foretaget interviews med repræsentanterne fra skoleledelsen.

Evaluatoren har som led i spørgeskemaundersøgelsen til skoleledere spurgt om en række organisatoriske og ledelsesmæssige forhold. Disse data bekræfter i vid udstrækning resultater fra øvrige undersøgelser og forskning⁷⁷. Det gælder bl.a. følgende forhold:

- Meget høj grad af prioritering af formel efteruddannelse blandt skolelederne
- Udbredt brug af forskellige organisationsformer, bl.a. klasseteams, fagteam samt afdelingsopdeling
- Stigende anvendelse af ressourcepersoner/vejledere, særligt inden for læsning
- At *ledelse efter mål og rammer* er den mest udbredte ledelsestilgang
- At *elevernes trivsel* prioriteres markant højere af skolelederne end en række øvrige faktorer (herunder et højt fagligt ambitionsniveau, der dog prioriteres næsthøjest).

Hvor det er relevant, bruges baggrundsplysninger om skolernes organisation og ledelsespraksis til at nuancere analysen i det følgende.

6.1.1 Skolernes evalueringsskema

Forud for evalueringens belysning af skoleledernes brug af de nationale test har vi i spørgeskemaundersøgelsen spurgt ind til skoleledernes syn på skolernes evalueringsskema. Skoleledernes svar, relateret til evalueringsskema, afrapporteres ikke her, idet en dybdegående analyse heraf findes i kapitel 3.

Det er dog værd at gentage – forud for den specifikke analyse af skoleledernes anvendelse af de nationale test – at skolelederne generelt tillægger evalueringsskemaet stor betydning, samt at en

⁷⁶ OECD beskriver dette dobbelte formål i deres review fra 2011: "The Ministry of Education clearly states that the intended purposes of the national tests are two-fold: to provide teachers with diagnostic information about students' strengths and areas of improvement and to monitor school performance", OECD Reviews of Evaluation and Assessment in Education, Denmark, OECD, May 2011.

⁷⁷ Se fx Ledelse af folkeskolerne, SFI, 2011.

række praksisser inden for evaluering ser ud til at være ganske udbredt på de fleste skoler. Endvidere viser analysen i kapitel 3, at de fleste dimensioner af evalueringspraksis er etableret på skolerne inden for de seneste 3-5 år.

Analysen konstaterer dog samtidig, at der er et væsentligt element af evalueringspraksis, der ser ud til at være mangelfuld. Det handler om brug af evalueringsresultater fra skoleledelsens side til at sætte årlige mål. Over en tredjedel af skolerne har angivet, at målfastsættelse på baggrund af evaluering slet ikke var en del af praksis. Endelig tyder det på, at skolelederne benytter en bred vifte af evalueringstværværktøjer i deres ledelse af skolerne.

For så vidt angår skolernes evalueringspraksis, skal det afslutningsvist konstateres, at det ikke på baggrund af evalueringens data har været muligt at bekræfte **hypotesen** om, at formel lederuddannelse hos skoleledelsen betyder stærkere evalueringskultur på skolen. Dette skyldes den begrænsede variation i skoleledernes svar, da stort set alle skoleledere har én eller anden form for skoleledelse.

6.1.2 Skoleledelsens anvendelse af de nationale test

Nedenstående tabel giver et indblik i, om og i hvilket omfang skoleledelsen bruger de nationale test.

Tabel 6.1: I hvilken grad bruger skolens ledelse resultaterne af de nationale test til systematisk at holde øje med og vurdere... Angivet i pct.

	Slet ikke	I mindre grad	I nogen grad	I temmelig høj grad	I meget høj grad	I alt (N)
Skolens niveau ved at se på, hvordan skolen placerer sig samlet set i forhold til den nationale præstationsprofil	5	17	39	30	10	100 (728)
Skolens niveau ved at se på, hvordan skolen klarer sig som helhed i forhold til tidligere års resultater	4	11	37	36	12	100 (728)
Klassens niveau ved at sammenligne de enkelte klassers resultat med hinanden (fx 6.a dansk med 6.b dansk)	19	22	31	21	7	100 (728)
Klassens niveau ved at sammenligne de enkelte klassers resultat over tid (fx ved at sammenholde testenens resultater med, hvordan klassen klarede sig på tidligere klassetrin)	7	15	35	33	11	100 (728)
Enkelte elevers faglige niveau	6	18	29	34	13	100 (728)
Lærerens faglige præstation	16	28	36	17	3	100 (728)

Kilde: Spørgeskemaundersøgelse til skoleledere, 2013.

I evalueringens forandringsteori er det en kritisk antagelse, at skoleledelsen reelt orienterer sig i og analyserer testresultaterne, som det er forudsat i loven, og som der er givet mulighed for⁷⁸. Grundlæggende må denne forudsætning siges at være opfyldt, da resultaterne tyder på, at skoleledelsen generelt orienterer sig i resultaterne. Dog er den hyppigste besvarelse (med en enkelt undtagelse), at de anvender testene i nogen grad. Der er en lille andel af skolerne, der slet ikke bruger de nationale test.

Det er mest udbredt for skoleledelsen at holde øje med resultaterne *over tid*. Det gælder både skolens niveau i forhold til tidligere år (knap halvdelen angiver, at de gør det i temmelig høj grad eller i meget høj grad), samt de enkelte klassers niveau over tid. Derimod er det markant færre,

⁷⁸ I henhold til Bekendtgørelse om anvendelse af test i folkeskolen mv. (BEK nr. 1000 af 26/10/2009) kan skolens leder foretage udtræk af (i) skolens samlede profil, (ii) de enkelte klassers gennemsnitsprofiler; (iii) antal elever, der har deltaget; samt (iv) en liste over de enkelte eleveres resultat i hver klasse.

der orienterer sig i, hvordan de enkelte klasser klarer sig i sammenligning med hinanden (fire ud af 10 skoler gør det slet ikke eller i mindre grad). Den nationale præstationsprofil bringes i anvendelse på stort set alle skoler (i ét eller andet omfang).

Tabellen indikerer meget stor variation i forhold til, om de nationale test fungerer som styringsværktøj for skoleledelsen i relation til lærernes faglige præstationer. De fleste skoleledelser angiver i nogen grad (36 pct.), og det er kun en mindre andel af skolerne, der gør dette i udpræget grad. Som de åbne svar i spørgeskemaundersøgelsen samt casestudierne viser, så er det udbredt praksis at anvende de nationale test i samspil med andre evalueringsredskaber og sjældent alene. Dertil kommer, at det generelt ikke synes at være standard at give lærerne systematisk feedback på deres indsats og præstationer⁷⁹. Der er dog skoler, hvor testene har stor betydning – også i forhold til at give feedback til lærerne. Disse eksempler vender vi tilbage til nedenfor.

Endelig konstaterer evaluatoren, at skoleledelsen på et betydeligt antal skoler tilsyneladende bruger de nationale test til på systematisk vis at holde øje med de enkelte elevers faglige niveau. Dette synes umiddelbart bemærkelsesværdigt, da det kunne være en antagelse, at indblik i testresultater på elevniveau er forbeholdt lærerne. Evalueringen kan ikke svare på, om dette i alle tilfælde sker i et samarbejde mellem lærere og skoleleder, men evaluatoren har set eksempler på dette.

Yderligere analyser af data fra spørgeskemaundersøgelsen viser endvidere, at der er en positiv sammenhæng mellem, hvorvidt skoleledelsen systematisk holder øje med elevernes niveau, og om skoleledelsen prioriterer et højt fagligt ambitionsniveau på vegne af eleverne. Sagt forenklet har de skoler, der angiver at have et højt fagligt ambitionsniveau, større tilbøjelighed til at orientere sig i resultater på elevniveau.

I tabellen nedenfor er det afsøgt, *hvad* de nationale test nærmere anvendes til, når resultaterne foreligger.

Tabel 6.2: I hvilken grad benytter skolens ledelse resultaterne af de nationale test til at... Angivet i pct.

	Slet ikke	I mindre grad	I nogen grad	I temmelig høj grad	I meget høj grad	I alt (N)
Identificere faglige problemstillinger, der skal håndteres	6	16	36	32	10	100 (728)
Prioritere nye pædagogiske indsatsområder	9	21	38	25	7	100 (728)
Prioritere ressourcer (ekstra timer/lærerressourcer/undervisningsmaterialer mv. til klasser eller elever)	11	22	36	24	7	100 (728)

Kilde: Spørgeskemaundersøgelse til skoleledere, 2013.

Tabellen viser relativ stor variation i skoleledelsens anvendelse af testresultaterne. Det hyppigste svar på alle spørgsmål er, at man i nogen grad identificerer faglige problemstillinger, prioriterer nye indsatsområder og prioriterer ressourcer (36-38 pct. af skolerne). Svarene indikerer på ny, at mange skoleledere gør aktiv brug af resultaterne fra de nationale test i deres kvalitetsarbejde.

De kvalitative casestudier vidner om den samme høje grad af variation. I *den ene ende af spektret* er der en række skoler, hvor der nok er fokus på at identificere og handle på faglige problemstillinger, men dette sker af læreren og ikke af skolens ledelse (selv om skoleledelsen orienterer sig i resultaterne). Hvis skoleledelsen har en rolle i fremadrettede handlinger på disse skoler, så er arbejdet centreret om de klasser, der klarer sig mindre godt. Den fremadrettede prioritering af nye pædagogiske indsatser (udover hvad læreren igangsætter i klassen) eller yderligere ressourcer sker ikke eller i meget begrænset omfang. Det bemærkes også, at der er skoler (6 pct.), hvor de nationale test slet ikke anvendes til at identificere faglige problemstillinger.

⁷⁹ Se fx OECD Reviews of Evaluation and Assessment in Education, Denmark, OECD, May 2011.

I *den anden ende af spektret* har evaluatoren besøgt skoler, hvor skoleledelsen i langt større omfang er aktive også i den fremadrettede anvendelse – frem for "blot" at monitorere testresultaterne og delegere den konkrete handling til lærerstaben. På disse skoler har testresultater været anledning til omfattende fremadrettede aktiviteter, som det fremgår nedenfor (eksempler fra udvalgte skoler):

- Prioritering af ressourcer til faglige vejledere, der på den pågældende skole havde ført til et markant fagligt løft
- Indkøb af nye bogsystemer på baggrund af utilfredsstillende resultater i naturfag
- Efteruddannelsesstrategi med henblik på løfte til linjefagsniveau inden for bestemte fagområder
- Ændret fagfordeling.

Ovenstående eksempler er udtryk for, at testresultater kan føre til konkrete ændringer, fx at flytte lærere fra en klasse ved gentagne dårlige resultater i nationale test, og/eller prioritere ressourcer til fag, hvor der er behov for ekstra støtte, vejledning eller opfølgning.

I nedenstående tabel afdækkes en række spørgsmål vedrørende *rammer, dialog med lærergruppen* samt brug af *frivillige test*.

Tabel 6.3: Angiv, i hvilken grad du er enig eller uenig i de følgende udsagn om nationale test. På min skole... Angivet i pct.

	Helt uenig	Overvejende uenig	Hverken enig eller uenig	Overvejende enig	Helt enig	I alt (N)
Har skolens ledelse klart kommunikeret formålet med/værdien af at benytte de nationale test	5	14	33	37	11	100 (728)
Har skolens ledelse klart kommunikeret <i>hvordan</i> lærerne skal bruge resultaterne af de nationale test	6	17	33	35	8	100 (728)
Arbejder ledelsen systematisk og løbende med at inddrage resultatet af de nationale test i en reflektiv dialog med lærergruppen, med henblik på at opstille klare faglige mål for udviklingen af skolens undervisning	8	20	35	31	7	100 (728)
Har vi en målsætning om at måle elevernes faglige progression ved systematisk brug af frivillige test	10	16	29	32	13	100 (728)

Kilde: Spørgeskemaundersøgelse til skoleledere, 2013.

For så vidt angår item 1 og 2, så belyser spørgsmålene, i hvilket omfang der på skolerne er *klare rammer og retningslinjer* for anvendelsen af de nationale test (herunder skoleledelsens rolle i den sammenhæng).

Knap halvdelen af respondenterne angiver, at skolens ledelse klart har kommunikeret formålet med/værdien af at benytte de nationale test, og en lidt mindre andel (43 pct. erklærer sig overvejende enige eller helt enige) svarer, at skolens ledelse klart har kommunikeret, *hvordan* lærerne skal bruge de nationale test. Det noteres endvidere, at en betydelig andel af skoleledere svarer hverken enig eller uenig på spørgsmålene.

I forhold til evalueringens forandringsteori er den store forskel skolerne imellem i forhold til "rammesætning" problematisk. Det vurderes vigtigt, at skoleledelsen tager ejerskab for brugen

af de nationale test, bl.a. ved at kommunikere formålet med at bruge testene. Det gælder ikke mindst sammenhængen til den øvrige evalueringsindsats på skolen. Resultaterne af spørgeskemaundersøgelsen indikerer samtidig, at der i vid udstrækning sker delegation af ansvar til lærerne i forhold til brugen af de nationale test.

Dette billede kan evaluator bekræfte på baggrund af udvalgte casestudier. Selv om skolelederne generelt tillægger de nationale test en vis værdi (se tabellen nedenfor), så er der flere skoler, hvor brugen af nationale test bliver italesat som symbolsk/instrumentel, fordi *”det er jo noget, som vi skal”*, som det blev udtrykt af en skoleleder. Evaluator vurderer samtidig, at resultaterne er udtryk for, at de nationale test nu er indlejret i skolernes praksis, hvorfor nogle skoleledere måske ikke vurderer, at der er behov for (fortsat) kommunikation om formål med og værdi af testene. Det er dog evaluators vurdering, at der er et potentiale for bedre anvendelse af de nationale test på de skoler, hvor dette ikke sker.

Evaluator bygger bl.a. dette på, at resultaterne fra spørgeskemaundersøgelsen viser en positiv sammenhæng mellem, om skoleledelsen klart har kommunikeret formålet med de nationale test, og om skolens lærere (ifølge skolelederne) anvender de nationale test fremadrettet til at tilrettelægge en undervisning, der tager udgangspunkt i den enkelte elevs faglige niveau. Denne sammenhæng fremgår af tabellen nedenfor.

Tabel 6.4: Sammenhæng mellem, om skoleledelsen klart har kommunikeret formålet med de nationale test og om skolens lærere (ifølge skolelederne) anvender de nationale test fremadrettet til at tilrettelægge en undervisning, der tager udgangspunkt i den enkelte elevs faglige niveau. Angivet i pct.

	Skolens lærere bruger resultaterne af de nationale test til fremadrettet at tilrettelægge en undervisning, der tager udgangspunkt i den enkelte elevs faglige niveau						
		Slet ikke	I mindre grad	I nogen grad	I temmelig høj grad	I meget høj grad	I alt (N)
På min skole – Har skolens ledelse klart kommunikeret formålet med/værdien af at benytte de nationale test	Helt uenig	6	28	53	14	0	100 (36)
	Overvejende uenig	0	17	49	28	7	100 (101)
	Hverken enig eller uenig	0	12	48	33	7	100 (232)
	Overvejende enig	0	7	48	32	13	100 (269)
	Helt enig	0	7	33	35	26	100 (77)

Kilde: Spørgeskemaundersøgelsen til skoleledere, 2013.

Dette indebærer, at evaluator delvist kan bekræfte **hypotesen** om, at skoleledelsens formulering af klare retningslinjer og målsætninger for anvendelsen af de nationale test har en positiv betydning på lærernes pædagogiske anvendelse af de nationale test.

Potentialet – og den delvise bekræftelse af hypotesen – bygger evaluator endvidere på kvalitative data, indhentet via casestudierne. Det er således evaluators vurdering, at klare retningslinjer og målsætninger for brug af testene understøtter den pædagogiske anvendelse. På de skoler med stærk evaluingskultur er det bl.a. den fremadrettede anvendelse, som man arbejder med at få til at fungere, og her spiller ledelsen en væsentlig rolle. Omvendt viser evalueringen, at retningslinjer fra skoleledelsens side på nogle skoler har været helt fraværende med den konsekvens, at anvendelsen bliver symbolsk fra lærernes side, og/eller at lærerne vender sig mod andre evaluingsværktøjer. På nogle af disse skoler efterlyser lærerne viden om, hvordan resultaterne kan bruges.

Tabellen viser videre, at skoleledelsen på knap 4 ud af 10 skoler arbejder systematisk og løbende med at inddrage resultatet af de nationale test i en refleksiv dialog med lærergruppen, med henblik på at opstille klare faglige mål for udviklingen af skolens undervisning. Også her angiver en stor andel af skolelederne (35 pct.), at de hverken er enige eller uenige. Dette kan enten repræsentere indifferens (at man ikke beskæftiger sig med temaet/ikke finder det relevant), eller at de

pågældende skoleledere har "uddelegeret" arbejdet med de nationale test til lærerne og derfor ikke er en del af denne dialog.

Evaluators har da også besøgt skoler, hvor "den reflekterende dialog med lærergruppen" slet ikke finder sted. I bedste fald finder dialogen sted, hvis der er brændende platforme, dvs. klasser med meget dårlige testresultater. Og i den anden ende af skalaen er der eksempler på skoler, hvor de nationale test spiller en central rolle i dialogen mellem skoleledelse og lærer. Der er ikke nødvendigvis tale om dialoger med hele lærergruppen på én gang, men snarere målrettede dialoger med specifikke faggrupper eller lærerne enkeltvis. Evaluator har også set eksempler på, at den årlige klassekonference er platform for dialog om klassens og enkeltelevers resultater samt lærerens indsats (klassekonference skal her forstås som en given klassens lærere/klasse team samt repræsentanter for ledelsen). Nogle bruger den årlige MUS-samtale som platform til at få læreren til at reflektere over egen praksis: Hvorfor ser resultatet ud, som det gør, og hvad kan gøres for at ændre på resultatet fremadrettet? Denne anvendelse er helt i tråd med ministeriets egen vejledning til skolelederne⁸⁰. Samlet set er der således stor variation i forhold til, om skoleledelsen arbejder systematisk med resultaterne i en dialog med lærerne.

I boksen nedenfor har vi set nærmere på for skolernes brug af ressourcerpersoner.

Boks 6.1: Skolernes brug af ressourcerpersoner (ifølge skolelederne)

Ressourcerpersoner på udvalgte faglige områder (særligt læsning) bruges i stigende grad i de danske folkeskoler. I spørgeskemaundersøgelsen til skoleledere er det afdækket nærmere, hvordan ressourcerpersoner anvendes i forhold til de nationale test. Evalueringen viser følgende resultater:

- Der er en formaliseret procedure for inddragelse af ressourcerpersoner som hjælp til tolkning af testresultater på knap halvdelen af skolerne (46 pct. af skolelederne er i temmelig høj grad eller i meget høj grad enige heri)
- En tilsvarende andel (46 pct.) er enige i, at skolen prioriterer ressourcer til, at lærerne kan sparre med skolens ressourcerpersoner om de nationale test.

Data viser således betydelig variation på skoleniveau i forhold til brugen af ressourcerpersoner, når der skal tolkes på resultater fra de nationale test. Yderligere analyser viser, at de skoleledere, der angiver, at skolen har en formaliseret procedure for inddragelse af ressourcerpersoner i højere grad, vurderer, at skolen er blevet bedre til, dels at foretage systematiske og løbende evalueringer af elevernes udbytte af undervisningen, dels at anvende evaluering fremadrettet med henblik på at tilrettelægge en undervisning, der tager udgangspunkt i den enkelte elev. Dette indikerer en positiv sammenhæng mellem formaliseret brug af ressourcerpersoner og styrket fremadrettet anvendelse af evalueringens resultater.

Endelig har evaluator afdækket skoleledernes **vurderinger af/holdninger til de nationale test**. Resultaterne fremgår nedenfor⁸¹.

⁸⁰ Brug testresultaterne – inspiration til pædagogisk brug af resultater fra de nationale test, Skolestyrelsen, 2011.

⁸¹ Det bemærkes, at skoleledernes oplevelse af effekterne af de nationale test er afrapporteret i kapitel 3.

Tabel 6.5: Angiv venligst, hvor enig eller uenig du/skoleledelsen er i relation til følgende udsagn om de nationale test. Nationale test... Angivet i pct.

	Helt uenig	Overvejende uenig	Hverken enig eller uenig	Overvejende enig	Helt enig	I alt (N)
giver ny viden om elevernes faglige niveau	2	11	29	49	10	100 (728)
understøtter lærernes arbejde med at differentiere undervisningen	2	9	28	50	12	100 (728)
påvirker undervisningens indhold uhensigtsmæssigt meget	10	28	41	18	3	100 (728)
giver skoleledelsen anvendelig information om skolens faglige niveau	2	7	27	52	13	100 (728)
fungere som et nyttigt dialogredskab til brug for samarbejdet mellem skole og forvaltning	13	22	41	19	5	100 (728)
fungere som et nyttigt dialogredskab mellem skole og hjem om elevernes faglige udvikling	2	6	21	57	14	100 (728)
skaber for meget bureaukrati på skolen	13	27	38	16	6	100 (728)
fungere alene som et dokumentationsredskab for forvaltningen over for kommunalbestyrelsen	39	23	28	8	2	100 (728)

Kilde: Spørgeskemaundersøgelse til skoleledere, 2013.

Ovenstående resultater giver et relativt klart og overvejende positivt billede af de nationale test, hvis man spørger skolelederne. Et udpluk af svarene illustrerer dette:

- Seks ud af 10 skoleledere er enige i (overvejende eller helt enige), at testene giver ny viden om elevernes faglige niveau. Der er altså *ikke* tale om, at testene kun bekræfter/reproducerer eksisterende viden
- En tilsvarende andel mener, at de nationale test understøtter arbejdet med at differentiere undervisningen – en opgave, der notorisk er en stor udfordring i folkeskolen
- 71 pct. vurderer, at testene er et nyttigt værktøj i skole-hjem-samarbejdet
- 65 pct. angiver, at testene giver anvendelig information om skolens faglige niveau.

Alt i alt vurderer evaluator, at størstedelen af skolelederne i de danske folkeskoler har taget testene til sig og finder dem anvendelige. På trods heraf er det samtidig tydeligt (baseret på flere datakilder), at der stadig er en betydelig andel af skolelederne, der ikke finder dem relevante, og/eller overlader ansvaret til lærerne ud fra devisen, at der er tale om et pædagogisk værktøj (ses også ved, at en stor andel svarer hverken enig eller uenig til udsagnene ovenfor). Lignende tilbagemeldinger er tidligere givet fra skoleledere i andre undersøgelser, fx om Fælles Mål, hvor en del skoleledere ikke ønskede at blande sig i arbejdet med Fælles Mål, da dette ansås som "lærernes domæne"⁸².

Det bemærkes i øvrigt, at blot én ud af fire anser de nationale test for at være et nyttigt dialogredskab i samarbejdet mellem skole og forvaltning (se nedenfor).

⁸² Fælles Mål i folkeskolen, Danmarks Evalueringsinstitut, 2012.

6.1.3 Brug af de nationale test som dialog- og styringsredskab

Som led i styringsdelen af forandringsteorien (se bilag 1) er det en antagelse, at dialog mellem skoleledere og kommune om skolens testresultater er væsentlig i forhold til fremadrettet at kunne opstille faglige mål for skolen.

I tabellen nedenfor har vi bedt skolelederen om at vurdere, hvordan de oplever kommunens arbejde med de nationale test.

Tablet 6.6: I hvilken grad stemmer nedenstående udsagn om kommunalbestyrelsens arbejde i relation til nationale test overens med oplevelsen på jeres skole? Angivet i pct.

	Slet ikke	I mindre grad	I nogen grad	I temmelig høj grad	I meget høj grad	Ved ikke	I alt (N)
Kommunalbestyrelsen tillægger resultaterne af de nationale test stor vægt i arbejdet med de kommunale kvalitetsrapporter og vurderingen af den enkelte skoles faglige niveau	5	13	26	18	16	22	100 (728)
Kommunalbestyrelsen bruger den nationale præstationsprofil i den løbende vurdering af det faglige niveau på skolerne	7	14	23	18	13	26	100 (728)
Kommunalbestyrelsen har formuleret klare mål til elevernes faglige niveau, herunder resultaterne af de nationale test	25	19	20	9	6	21	100 (728)
Kommunerne udarbejder en handleplan til skolen, hvis resultatet af elevernes faglige niveau på en skole, herunder resultatet af de nationale test, ikke er tilfredsstillende	31	14	13	7	3	32	100 (728)
Kommunalbestyrelsen bruger aktivt resultaterne af de nationale test i en løbende og systematisk dialog med skoleledelsen om skolens faglige niveau og behovet for opfølgning	27	22	20	8	6	18	100 (728)

Kilde: Spørgeskemaundersøgelse til skoleledere, 2013.

Note: Respondenterne er blevet gjort opmærksomme på, at kommunalbestyrelsen kan læses som forvaltningen.

Skoleledernes svar på ovenstående udsagn giver et sammensat billede af skoleledernes opfattelse af kommunernes arbejde med de nationale test. Der er således stor variation i forhold til, om skolelederne vurderer, at kommunerne tillægger testresultaterne stor vægt. Skolelederne oplever i begrænset omfang, at kommunerne opstiller klare mål til elevernes faglige niveau (herunder testresultater), idet næsten halvdelen slet ikke eller kun i mindre grad vurderer, at dette er tilfældet.

Mest markant viser tabellen, at 49 pct. i mindre grad eller slet ikke finder, at kommunen aktivt bruger testresultaterne i en løbende og systematisk dialog med skoleledelsen. Denne relativt kritiske (eller "blot" nøgterne) vurdering af kommunernes brug af de nationale test stemmer *ikke* overens med kommunernes egen vurdering af testresultaterne. Her svarer en del flere kommuner på samme spørgsmål, at de aktivt bruger resultaterne. Der er dog stadig 4 ud af 10 kommuner, der slet ikke eller i mindre grad bruger resultaterne aktivt i dialog med skoleledelsen. Det er sandsynligvis udtryk for, at testresultaterne ses som en del af en større sammenhæng, jf. afsnit 6.2.

Under alle omstændigheder indebærer ovenstående, at forandringsteorien på dette punkt (den systematiske og løbende dialog mellem skoleledelse og kommune) udfordres på en lang række skoler. Dette bekræftes i flere af casestudierne, hvor vurderingen er, at de nationale test er vigtige, men kun på det overordnede niveau. Kommunerne/forvaltningerne efterstræber gode resultater i de nationale test, men der er sjældent direkte samarbejde om disse. En repræsentant fra en af de interviewede forvaltninger anførte, at man fra kommunal side med vilje *ikke* opstillede egentlige mål for skolerne, da man mente, at der dermed ville være risiko for *teaching to the test*.

Omvendt er der eksempler i evalueringen på, at nogle kommuner spiller en meget aktiv rolle, hvor der foretages en løbende og systematisk dialog med skolerne om resultaterne, og hvor man også inddrager resultaterne fremadrettet i kommunernes strategiske planlægning. Det bemærkes dog samtidig, at der er en tendens til, at de kommunale forvaltninger mest er interesserede i de skoler, der klarer sig dårligt. Der kan derfor typisk konstateres en form for "slagside", eller alternativt prioritering, fra kommunal side.

Væsentligt er det også at fremhæve, at evaluator har oplevet eksempler, hvor skoler leverer stærke faglige resultater (bl.a. i de nationale test), men hvor forvaltningen ikke spiller en aktiv rolle. Det tyder på, at forvaltningerne ikke i alle tilfælde er udslagsgivende for positive faglige resultater. Omvendt vurderer evaluator, at de nationale test netop kan være afgørende som styringsværktøj i de tilfælde, hvor der *ikke* er tale om en stærk/dygtig skoleledelse.

I forhold til evalueringens **hypoteser om skoleledelse** kan evalueringen delvist bekræfte, at skoleledelsens formulering af klare retningslinjer og målsætninger for testanvendelse har en positiv betydning for lærernes pædagogiske anvendelse. Der er eksempler på, at rammesætningen slet ikke finder sted, og i de tilfælde kan hypotesen indlysende ikke bekræftes. Omvendt viser analyse af spørgeskemadata samt udvalgte casestudier, at den positive sammenhæng findes. Tilsvarende kan hypotesen om, at anvendelse af testresultater bidrager til et bedre grundlag for skoleledelsens kvalitetsarbejde, delvist bekræftes. På skolerne med stærk evalueringskultur er der en klar sammenhæng med den vigtige tilføjelse, at skolelederne anvender testresultater som ét ud af flere evaluerings-/monitoreringsværktøjer. Den sidste hypotese, relateret til skoleledelse – at inddragelse af testresultater (sammen med øvrig evaluering) giver et styrket grundlag for aktiv pædagogisk skoleledelse – betoner den formative og fremadrettede dimension af testanvendelsen. Denne hypotese kan evalueringen kun i begrænset omfang bekræfte. Det skyldes, at de fleste skoleledere (endnu) anvender testresultaterne til at skabe et overblik (summativt), og det er således et mindretal, der anvender resultaterne fremadrettet, som hypotesen lægger op til.

6.2 De nationale test som værktøj for kommuners kvalitetsarbejde

Det kommunale niveau – kommunalbestyrelse og den kommunale skoleforvaltning – var i begrænset omfang i fokus i det lovforslag, der lå til grund for indførelse af de nationale test. Heri nævnes det blot, at kommunerne har ansvar for at sikre, at de obligatoriske test gennemføres (kommunens ansvar som skoleejer), samt at kommunerne ville kunne orientere sig i testresultaterne.

Det var det efterfølgende lovforslag fra 2006, der specifikt koblede de nationale test sammen med et politisk ønske om at præcisere det kommunale ansvar for folkeskolen (lovforslaget om indførelse af elevplaner, kvalitetsrapporter mv.). I lovforslaget introduceres kravet om udarbejdelse af kommunale kvalitetsrapporter som svar på en række udfordringer vedrørende tilsyn, kvalitetsudvikling og evalueringskultur på kommunalt niveau.

Med udgangspunkt i lovændringerne af folkeskoleloven fra 2006 kan det konstateres, at de nationale test primært har til hensigt at fungere som et styringsværktøj på kommunalt niveau, når de ses i sammenhæng med de kommunale kvalitetsrapporter. Således fremgår det af lovbetragtningerne til lovforslaget om elevplaner, kvalitetsrapporter mv., at kvalitetsrapporten skal anvendes til at vurdere skolernes faglige niveau "*i forhold til resultaterne fra de obligatoriske nationale test samt skolernes evalueringspraksis*". Det forventes endvidere, at de nationale test kan være omdrejningspunktet for en løbende dialog om kvalitet mellem kommune og skoleledere med henblik på løbende og systematisk opfølgning og eventuel handling.

Nedenfor afrapporteres evalueringens resultater, for så vidt angår kommunernes kvalitetsarbejde. Resultaterne er baseret på en spørgeskemaundersøgelse til de kommunale skoleforvaltninger samt de gennemførte casestudier, hvor der er foretaget interviews med repræsentanter fra skoleforvaltningerne.

6.2.1 Kommuners generelle kvalitetsarbejde på skoleområdet

I tabellen nedenfor har kommunerne i spørgeskemaundersøgelsen angivet, hvad der kendetegner kommunens *generelle* kvalitetsarbejde på skoleområdet. Dette vurderes relevant, da de nationale test, jf. lovgrundlaget, indgår som ét ud af flere værktøjer i arbejdet med tilsyn og kvalitetsudvikling af det samlede skolevæsen.

Tabel 6.7: I hvilken grad er følgende forhold kendetegnende for kommunens generelle kvalitetsarbejde på skoleområdet? Angivet i pct.

	Slet ikke	I mindre grad	I nogen grad	I temmelig høj grad	I meget høj grad	Ved ikke	I alt (N)
Kommunen har fastsat klare mål til skolernes faglige resultater	5	13	39	27	17	0	100 (78)
Kommunen har faste procedurer for dialog med skolerne om opfyldelse af de faglige mål	3	4	19	32	42	0	100 (78)
Kommunen følger løbende og systematisk op på skolernes opfyldelse af de faglige mål	1	4	26	39	31	0	100 (78)
Skolers manglende efterlevelse af faglige mål fører til krav om, at skolerne skal forbedre sig fagligt	1	12	28	35	22	3	100 (78)

Kilde: Spørgeskemaundersøgelse til skoleforvaltninger, 2013.

Som det fremgår af tabellen, angiver størstedelen af kommunerne *både* at opstille klare mål for skolernes faglige præstationer *og* at arbejde løbende og systematisk med opfyldelsen af disse mål. Således angiver 8 ud af 10 kommuner, at de har fastlagt klare faglige mål, om end det skal understreges, at flest svarer 'I nogen grad'. Langt størstedelen af kommunerne (70 pct.) svarer, at de i temmelig høj grad eller i meget høj grad følger løbende og systematisk op på målopfyldelsen på skolerne, og en betydelig andel stiller tilsvarende krav til forbedringer, hvis der kan konstateres manglende efterlevelse af mål⁸³.

På baggrund af spørgeskemaundersøgelsen ser det således ud til, at de kommunale forvaltninger generelt stiller klare krav til skolernes faglige resultater, og at de – for de flestes vedkommende – ikke mindst følger op på, om dette er tilfældet.

Det fremgår endvidere af spørgeskemaundersøgelsen, at de lovpligtige kvalitetsrapporter er det centrale styringsværktøj for kommunerne på skoleområdet. Alle kommuner anvender kvalitetsrapporter for det samlede skolevæsen, og godt 7 ud af 10 kommuner vurderer, at kvalitetsrapporter i temmelig høj grad eller i meget høj grad bidrager til at styrke kvalitetsudviklingen af skolevæsenet. En stor andel gør tilsvarende brug af kvalitetsrapporter for de enkelte skoler.

⁸³ Selv om der er begrænset viden om betydningen af kommunale krav og mål for skoler og elevers præstationer, så har flere undersøgelser peget på den positive betydning af klare mål – også på lokalt niveau. Således konkluderer AKF i deres rapport fra 2010, at højt præsterende skoler bl.a. er kendetegnede ved en tydelig ledelse med klare krav fra forvaltningens side i dialog med skoleledelsen, jf. Den højt præsterende skole, AKF, 2010. Derudover peger formandskabet for Skolerådet i deres beretning fra 2012 på en række internationale undersøgelser, der understøtter denne konklusion.

Endelig viser spørgeskemaundersøgelsen et meget stort fokus på resultater generelt i kommunerne. Flest kommuner (halvdelen) prioriterer skolens resultater som den vigtigste faktor foran skolernes rammebetingelser (31 pct. angiver dette som vigtigst) og de pædagogiske processer (19 pct. finder dette vigtigst). Disse svar fremgår af tabellen nedenfor.

Table 6.8: Hvis du tager udgangspunkt i behandlingen af den seneste kvalitetsrapport for kommunens skolevæsen: Hvilke af nedenstående faktorer har størst vægt i kommunalbestyrelsens drøftelse og behandling af de kommunale kvalitetsrapporter? Angivet i pct.

	Vigtigste faktor	Næstvigtigste faktor	Tredje vigtigste faktor	I alt (N)
Skolernes rammebetingelser (fx antal elever, timer, linjefagsdækning, økonomi)	31	37	32	100 (78)
Skolernes pædagogiske processer	19	23	58	100 (78)
Skolernes resultater (fx i de nationale test, afgangsprøver, overgang til ungdomsuddannelse)	50	40	10	100 (78)

Kilde: Spørgeskemaundersøgelse til skoleforvaltninger, 2013.

Note: Respondenterne er blevet bedt om at prioriterer de tre udsagn.

Der er således i spørgeskemaundersøgelsen flere indikationer på, at der fra kommunal side generelt er et stærkt fokus på skolernes faglige resultater – også når denne faktor skal prioriteres i forhold til andre forhold⁸⁴. Nærværende evaluerings kvalitative data, indhentet via casestudierne, nuancerer de data, der er indhentet ved hjælp af spørgeskemaundersøgelsen. Det generelle indtryk er, at kommunerne har stærkt fokus på faglige resultater, men i flere tilfælde orienterer de kommunale forvaltninger sig i højere grad mod skolernes afgangskarakterer (der offentliggøres) end mod resultaterne i de nationale test. Her er testresultaterne i højere grad én ud af flere indikatorer i kvalitetsrapporten (se næste afsnit). I nogle tilfælde er der tale om ingen eller begrænset anvendelse af de nationale test fra kommunal side.

6.2.2 Kommunernes anvendelse af de nationale test

Nedenstående tabel giver et indblik i, hvilken vægt de nationale test tillægges af kommunerne i forhold til andre faktorer, der relaterer sig til kommunens kvalitetsarbejde.

Table 6.9: I hvilken grad er følgende faktorer væsentlige for kommunens kvalitetsarbejde i dialogen med skolelederne? Udvalgte items. Angivet i pct.

	Slet ikke	I mindre grad	I nogen grad	I temmelig høj grad	I meget høj grad	Ved ikke	I alt (N)
Skolernes resultater i afgangsprøver	1	4	33	31	31	0	100 (78)
Skolernes resultater i de nationale test	6	12	41	19	22	0	100 (78)
Skolernes resultater i andre standardiserede test	5	15	31	28	19	1	100 (78)
Skolernes overgangsfrekvens til ungdomsuddannelser	1	6	26	37	30	0	100 (78)
Skolernes inklusionsgrad	3	9	21	23	45	0	100 (78)

Kilde: Spørgeskemaundersøgelse til skoleforvaltninger, 2013.

⁸⁴ Anden dokumentation peger på, at der også er udfordringer i forhold til kommunernes arbejde med kvalitetsudvikling, herunder målformulering og opfølgning, jf. bl.a. OECD Reviews of Evaluation and Assessment in Education, Denmark, OECD, May 2011, Delrapport 3: Kvalitetsrapporten – undersøgelse af kvalitetsrapportens betydning for praksis på skoler og i kommuner, Danmarks Evalueringsinstitut, 2011, samt Beretning om evaluering og kvalitetsudvikling af folkeskolen 2012, Formandskabet for Rådet for Evaluering og Kvalitetsudvikling af Folkeskolen, maj 2012.

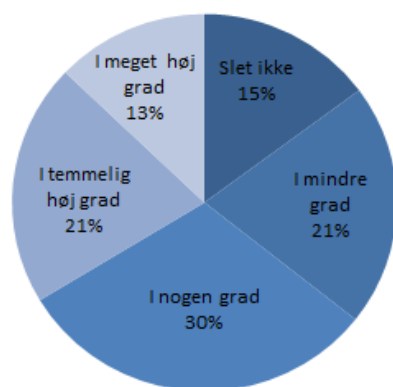
Tabellen viser, at 8 ud af 10 kommuner finder skolernes resultater i de nationale test væsentlige i nogen grad, i temmelig høj grad eller i meget høj grad. Det bemærkes imidlertid, at den klart største andel af kommunerne (41 pct.) angiver, at testene i nogen grad er væsentlige i kommunens kvalitetsarbejde. Samtidig fremgår det, at andre indikatorer – resultater i afgangsprøver, øvrige testresultater, overgangsfrekvens og ikke mindst inklusionsgraden – anses for væsentlige fra et kommunalt perspektiv. Dette samlede billede, hvor resultaterne af de nationale test tilægges en vis, men ikke nødvendigvis afgørende, vægt, understøttes af casestudierne.

Hovedbudskabet fra casestudierne er, at resultater fra de nationale test er "en del af paletten", som det udtrykkes. Kommunerne interesserer sig for skolernes faglige resultater, men de nationale test er én ud af mange brikker til at sammensætte et billede af en skoles faglige niveau. De fleste kommuner betragter de nationale test som væsentlige, især fordi de skaber det hurtige overblik. De gør det samtidig muligt at sammenligne på tværs af skoler. Flere kommuner angiver, at de bruger resultaterne fra de nationale test over for skoler lidt på samme måde, som lederne anvender dem over for lærerne – til at identificere gode og *i særdeleshed* mindre gode skoler. Og til det formål er de nationale test særdeles anvendelige.

Der er imidlertid også kommuner, der slet ikke eller i begrænset omfang gør brug af de nationale test, bl.a. fremhæves udfordringer i forhold til kravet om fortrolighed om testresultaterne. Kravet om fortrolighed medfører, at der i nogle tilfælde er usikkerhed om, hvordan testresultater kan anvendes. Det gælder særligt, når testresultater aggregeres og skal anvendes som dokumentation af de kommunale forvaltninger. Konsekvensen bliver i nogle tilfælde, at testresultater *ikke* eller *i begrænset omfang* anvendes med henvisning til fortrolighedsprincippet.

Med ovenstående resultat in mente viser figuren nedenfor kommunernes svar på, om de anvender resultaterne af de nationale test i kvalitetsarbejdet *udover* i arbejdet med kvalitetsrapporten. Svaret er stillet for at få et billede af *bredden* i anvendelsen af testene på kommunalt niveau.

Figur 6.1: Indgår resultaterne af de nationale test i forvaltningens kvalitetsarbejde udover i arbejdet med udarbejdelse og opfølgning på kvalitetsrapporten? Pct.



Kilde: Spørgeskemaundersøgelse til skoleforvaltninger, 2013.

Figuren viser stor variation i anvendelsesgraden blandt kommunerne, når der ses bort fra den obligatoriske anvendelse som indikator i sammenhæng med kvalitetsrapporten. Resultatet tyder umiddelbart på, at de nationale test af en række kommuner anses som et relevant værktøj i det bredere arbejde med kvalitetsudvikling, selv om der også er en betydelig andel af kommunerne, der svarer 'Slet ikke' (15 pct.) eller 'I mindre grad' (21 pct.).

Hvis lovgrundlaget bag såvel nationale test som kvalitetsrapporter læses meget snævert, skal den primære anvendelse af de nationale test på kommunalt niveau ske i forbindelse med udarbejdelsen af (og opfølgning på) kvalitetsrapporten, jf. også tidligere i rapporten. Figuren ovenfor indikerer imidlertid en mere varieret brug af resultaterne af de nationale test, og at værktøjet i nogle kommuner også bruges i andre sammenhænge.

Den store variation mellem kommuner bekræftes af de casestudier, som evaluatoren har gennemført. De 10 casestudier er eksempler på en kommunal anvendelsesgrad, der rækker over et kontinuum fra *ingen/begrænset anvendelse* over *brug af test til at skabe overblik* til *at udgøre et vigtigt element i de løbende feedbacksamtaler mellem forvaltning og skoleledere*.

I nedenstående tabel er de kommunale respondenter blevet bedt om at tage stilling til en række udsagn vedrørende kommunens arbejde med de nationale test.

Table 6.10: I hvilken grad stemmer nedenstående udsagn overens med kommunens arbejde med de nationale test? Udvalgte items. Angivet i pct.

	Slet ikke	I mindre grad	I nogen grad	I temmelig høj grad	I meget høj grad	Ved ikke	I alt (N)
Kommunen har fastsat klare mål til skolernes resultater i de nationale test	42	23	22	8	4	1	100 (78)
Kommunen følger løbende og systematisk op på, om skolerne opnår de forventede resultater i de nationale test	18	24	32	14	12	0	100 (78)
Kommunen gør brug af de forskellige muligheder for resultatvisninger i testsystemet til at følge udviklingen i kommunens og skolernes testresultater	12	23	22	24	17	3	100 (78)
Kommunen bruger den nationale præstationsprofil i den løbende vurdering af det faglige niveau på de enkelte skoler	17	23	24	21	13	3	100 (78)
Kommunen bruger aktivt resultaterne af de nationale test i en løbende og systematisk dialog med skoleledelsen om skolens faglige niveau og behovet for opfølgning	14	26	27	17	15	1	100 (78)
Kommunen bruger aktivt testresultaterne som grundlag for iværksættelse af nye initiativer målrettet faglig udvikling, kvalitetsudvikling og lign. i kommunen (samlet for alle skoler)	15	24	28	18	13	1	100 (78)
Kommunen bruger aktivt testresultaterne som grundlag for iværksættelse af nye initiativer målrettet faglig udvikling, kvalitetsudvikling og lign. på enkelte skoler	14	24	32	18	10	1	100 (78)

Kilde: Spørgeskemaundersøgelse til skoleforvaltninger, 2013.

Tabellen indeholder en række interessante resultater, der sammenfattes i punkterne nedenfor:

- For det første viser tabellen en betydelig forskel mellem det generelle kvalitetsarbejde i kommunerne og anvendelsen af de nationale test. Således er det relativt få kommuner (12 pct. angiver 'I temmelig høj grad' eller 'I meget høj grad'), der angiver at have opstillet klare mål til skolernes resultater i de nationale test. Dette skal sammenholdes med, at kommunerne i større grad angiver at have klare mål til skolernes faglige resultater *generelt* (jf. tabel 6.7). Tilsvarende finder den specifikke opfølgning i forhold til testresultater i langt mindre grad sted fra kommunal side, end hvad der er tilfældet på det generelle niveau. Svarene indikerer, at mange kommuner opfatter de nationale test som skolernes (herunder lærernes værktøj), men samtidig, at manglende målformulering og opfølgning ikke nødvendigvis er ensbetydende med fravær af faglige ambitioner fra kommunal side. Denne pointe kan evaluator underbygge ved hjælp af de kvalitative data. Mange kommuner bruger de nationale test til at skabe "*et hurtigt overblik*", men betragter dem ellers som lærernes værktøj. Det er opfattelsen fra kommunerne, at det er skolernes ansvar at opstille mål for resultatopnåelse i de nationale test – ikke kommunernes.
- For det andet viser tabellen, at langt størstedelen i ét eller andet omfang gør brug af de muligheder, som kommunerne har, for at gå ind i testsystemet og få overblik via forskellige resultatvisninger. Der er i øvrigt stor variation i, hvor meget kommunerne gør brug af denne mulighed, og 12 pct. af kommunerne gør det slet ikke. Stort set den samme variation gør sig gældende ved spørgsmålet om, hvorvidt kommunerne gør brug af den nationale præstationsprofil (det gennemsnitlige landsresultat). Evaluator konstaterer, at de fleste kommuner reelt gennemgår skolernes resultater i testsystemet, hvilket i evalueringens forandringsteori vurderes som en kritisk antagelse for at kunne indgå i dialog med skolerne om deres resultater.
- Endelig viser tabellen for det tredje, at det ikke er udbredt blandt kommunerne at bruge resultaterne af de nationale test til dels aktivt at gå i dialog med skoleledelsen, dels aktivt at iværksætte nye initiativer målrettet faglig udvikling. De tre sidste spørgsmål i ovenstående tabel viser, at kun 3 ud af 10 kommuner i betydeligt omfang gør aktivt brug af testresultaterne i den fremadrettede dialog. Dette tyder på, at kommunerne primært *orienterer sig* i testresultaterne (jf. lovbemærkningerne) og i mindre omfang bruger testene fremadrettet. Med andre ord er en summativ anvendelse af testene mest udbredt i kommunerne, mens en formativ anvendelse enten ikke vurderes relevant, eller også kan svaret være udtryk for, at testene ikke *isoleret set* giver anledning til fremadrettet handling. Testene skal snarere ses som et delement i den bredere styringsdialog mellem kommune og skoleledelse. Samlet set er det generelle billede – vurderet op imod evaluators model for evalueringskultur – at de fleste kommuner bør placeres på "de lavere niveauer". Omvendt har evaluator også mødt eksempler på kommuner med meget aktiv anvendelse af testresultaterne i dialogen med skoleledelsen, jf. eksemplet nedenfor.

Det bemærkes endvidere, at detaljerede analyser af resultaterne fra spørgeskemaundersøgelsen ikke har afsløret statistiske sammenhænge mellem kommunestørrelse og arbejdet med de nationale test.

Boks 6.2: Et eksempel på aktiv kommunal anvendelse af de nationale test

I en kommune indgår resultatet af de nationale test som et vigtigt element i de løbende feedbacksamtaler mellem forvaltning og skoleledere. Centerchefen for skoleområdet orienterer sig i resultaterne af testene på hver enkelt årgang, men understreger samtidig, at testresultaterne ikke kan stå alene. Det er en del af det samlede billede af skolernes faglige niveau. Der er en forventning om, at skolelederne selv peger på de klasser/årgange, som centerchefen skal forholde sig særskilt til, og at skolelederen i forlængelse heraf har reflekteret over, hvorfor resultatet er, som det er, og hvad der kan gøres. Det er centerchefens erfaring, at en dialog om resultatet af de nationale test kan afføde, at skolerne rykker sig fagligt, eksempelvis ved at der iværksættes en styrket læseindsats. En sådan indsats skal skolerne beskrive i den handleplan, der indgår som en del af kvalitetsrapporten.

På kommunalt niveau har man også fokuseret sin indsats og ressourcer på baggrund af testresultater og karakterer i folkeskolens afgangsprøve. Resultaterne inden for naturfag var generelt ikke tilfredsstillende, og man har derfor besluttet at ansætte en naturfagskonsulent.

Centerchefen understreger afslutningsvis, at det er vurderingen, at nationale test kan styrke det faglige niveau, men det forudsætter, at testene først og fremmest bruges som det pædagogisk redskab, de er tænkt som, altså i mødet mellem læreren og eleven. Det styringsmæssige element er sekundært, men absolut væsentligt i det fortsatte systematiske arbejde med at understøtte skolernes faglige resultater.

Evaluator har tilsvarende mødt et par andre kommuner, hvor resultatet af de nationale test og opfølgning herpå indgår som et væsentligt element i enten LUS-samtaler (Lederudviklingssamtaler) eller i dialogen med bl.a. ressourcepersoner/vejledere i kommunale netværk.

Ovenstående observationer kan nuanceres yderligere ved hjælp af kommunernes svar på nedenstående udsagn, som udtrykker kommunernes *holdninger* til de nationale test⁸⁵.

⁸⁵ Heller ikke i relation til kommunernes holdninger til de nationale test har det været muligt at konstatere signifikante sammenhænge mellem kommunestørrelse og besvarelse af de holdningsbaserede udsagn om de nationale test.

Tabel 6.11: Angiv i hvilken grad du er enig eller uenig i nedenstående udsagn om de nationale test. Udvalgte items. Angivet i pct.

	Helt enig	Overvejende enig	Hverken enig eller uenig	Overvejende uenig	Helt uenig	Ved ikke	I alt (N)
Resultaterne af de nationale test udgør en væsentlig indikator til at vurdere skolernes faglige resultater	14	54	23	8	0	1	100 (78)
Indførelsen af de nationale test har styrket det politiske fokus i kommunen på faglige resultater	10	37	30	15	5	3	100 (78)
De nationale test er et nyttigt værktøj til at følge udviklingen i det faglige niveau i kommunen og på de enkelte skoler	23	53	19	4	0	1	100 (78)
Resultaterne af de nationale test prioriteres ikke højere end andre indikatorer til at vurdere skolernes faglige resultater	23	47	15	9	3	3	100 (78)
Resultaterne fra de nationale test udgør et godt grundlag for dialog mellem forvaltning og de enkelte skoler i forhold til kvalitetsudvikling af kommunens skoler	21	46	22	8	1	3	100 (78)

Kilde: Spørgeskemaundersøgelse til skoleforvaltninger, 2013.

Samlet set peger svarene på, at kommunerne anser de nationale test som havende en væsentlig betydning. Således er 68 pct. af kommunerne helt enige eller overvejende enige i, at resultaterne af de nationale test udgør en væsentlig indikator til at vurdere skolernes faglige resultater. Tilsvarende er 3 ud af 4 kommuner helt eller overvejende enige i, at de er et nyttigt værktøj til at følge udviklingen i det faglige niveau. Størstedelen erklærer sig samtidig enige i, at resultaterne af de nationale test ikke prioriteres højere end andre indikatorer.

Det er evaluators vurdering, at ovenstående svar fra spørgeskemaundersøgelsen kan fortolkes i samme retning, som evaluator fremhævede ovenfor. Resultaterne af de nationale test tillægges betydelig værdi, men primært som *netop* en indikator, der kan bidrage til at skabe overblik over faglige præstationer. I forhold til fremadrettet anvendelse og handling kan resultaterne af de nationale test ikke stå alene og vurderes derfor ikke at have større værdi end andre indikatorer/værktøjer. Dette indtryk understøttes af de gennemførte casestudier.

Evaluator har gennemført omfattende statistiske analyser af data fra spørgeskemaundersøgelsen til kommunerne for at finde relevante sammenhænge. Analysen viser en meget stærk sammenhæng mellem, om kommunerne angiver at finde de nationale test væsentlige for kommunens kvalitetsarbejde i dialogen med skolelederne, og om kommunerne har opfattelsen af, at de nationale test har medvirket til at styrke elevernes faglige resultater på kommunens skoler. Evaluator har også via casestudierne set udvalgte eksempler på betydningen af kommunens rolle i forhold til at skabe effekt, men det er langt fra det generelle billede, at kommunerne har en aktiv dialog med skolelederen om resultaterne af de nationale test (som forudsat på de øverste niveauer i modellen for evalueringskultur). De kommunerelaterede **hypoteser** kan på den baggrund kun i begrænset omfang bekræftes. Der er dog indikationer på, at kommunerne kan spille en væsentlig rolle i arbejdet med at målrette skolernes arbejde på baggrund af input fra de nationale test.

6.2.3 Samarbejde og dialog mellem forvaltning og skoler

I forhold til samarbejde og dialog mellem forvaltning og skoler har langt hovedparten af kommunerne (95 pct.) skoleledermøder, fællesledermøder og lign. på månedlig basis. Derudover har de fleste kommuner formaliseret dialogen og samarbejdet i årlige MUS-samtaler mellem skolechef (e.l.) og skoleledere, og i kvartalsvise/månedlige møder, dialog og lign. i forbindelse med kommunale udviklingsprojekter. Endelig har en stor del af kommunerne månedlig/ugentlig sparring, coaching e.l. fra skolechef til skoleledere, samt ad hoc dialog.

På kommunalt niveau har de fleste ansat pædagogiske konsulenter og psykologer (fx i PPR) og omkring 2/3 af kommunerne har også ansat faglige vejledere og udviklingskonsulenter. I forhold til hvilke fagområder, der dækkes af kommunens konsulenter og vejledere, svarer 90 pct., at de dækker læsning. Omkring 60 pct. angiver tosprogsområdet, og 30 pct. angiver matematik samt natur og teknik. Derudover har 27 pct. af kommunerne ansat konsulenter eller vejledere inden for evaluering.

I forhold til de nationale test angiver forvaltningerne, at omkring 25 pct. af de kommunale konsulenter og vejledere rådgiver den enkelte skole om, hvordan de kan fortolke og anvende resultatet af de nationale test, samt hvordan man på kommunalt niveau kan følge op på resultatet af de nationale test. Modsat svarer omkring 40 pct. af kommunerne, at de kommunale konsulenter og vejledere i mindre grad eller slet ikke rådgiver herom. Der er dog en positiv sammenhæng mellem ansættelse af kommunale konsulenter og vejledere inden for læsning og graden af rådgivning fra konsulenterne til skolen om, hvordan de skal anvende resultatet af de nationale test. Samtidig er der en positiv sammenhæng mellem ansættelse af kommunale konsulenter og vejledere inden for evaluering, og i hvor høj grad konsulenterne indgår i en dialog med skolerne om, hvordan man på kommunalt niveau kan følge op på resultaterne af de nationale test.

På baggrund af data fra såvel spørgeskemaundersøgelse som casestudier vurderer evaluatoren, at der er begrænset dialog mellem skole og forvaltning om de nationale test, hvorfor sidste del af **hypotesen** om aktivt samarbejde (se bilag 1) udfordres. I de kvalitative interview fremgår det dog, at det i højere grad er den kommunale læsekonsulent, som kommer rundt på skolerne, der har dialogen om skolens resultater, herunder resultater i de nationale test, hvilket bakkes op af spørgeskemabesvarelsene.

Evalueringsens **hypoteser om kommunen/forvaltningen** (at aktivt samarbejde/dialog mellem forvaltning og skoleledere om testresultater fører til en målretning af skolernes arbejde, *samt* at et klart kommunalt fokus på faglige resultater har en positiv indflydelse på skolernes faglige niveau) kan kun i begrænset omfang bekræftes. Dette tilskrives primært, at de fleste kommuner bruger testresultaterne til at skabe overblik og således ikke arbejder med testresultater på et niveau, som det var antaget på "de øverste niveauer" i evaluators model for evalueringskultur (dvs. formativ anvendelse). Men der er eksempler på, at denne form for dialog finder sted, hvorfor evaluatoren kan pege på et potentiale for stærkere kommunal anvendelse af de nationale test.

6.3 Delkonklusion

I dette kapitel er der foretaget en analyse af de nationale test som værktøj for skolelederes og kommuners kvalitetsarbejde, dvs. brugen af testene som dialog- og styringsredskab. I tabellen nedenfor er indsat en oversigt over de nøglehypoteser, der har været undersøgt i denne forbindelse.

Tabel 6.12: Oversigt over nøglehypoteser

Hypotese	Data	Status
Skoleledelse		
Formel lederuddannelse hos skoleledelsen betyder stærkere evalueringskultur på skolen	Spørgeskemaundersøgelse til skoleledere	NA (for lidt spredning i data)
Skoleledelsens formulering af klare retningslinjer og målsætninger for anvendelsen af de nationale test har positiv betydning for lærernes pædagogiske anvendelse af de nationale test	Spørgeskemaundersøgelse til skoleledere Casestudier, herunder interview med skoleledere, lærere m.fl.	Delvist bekræftet +
Anvendelse af resultaterne fra de nationale test bidrager til, at skoleledelsen får et bedre grundlag for kvalitetsudvikling af skolen (ved fx at formulere mål og strategier for skolens arbejde)	Spørgeskemaundersøgelse til skoleledere Casestudier, herunder interview med skoleledere, lærere m.fl.	Delvist bekræftet +
Inddragelse af resultaterne af de nationale test giver – sammen med resultaterne af den øvrige evalueringsindsats – et styrket grundlag for aktiv pædagogisk skoleledelse (fx ved at følge op på elevpræstationer og i dialog udvikle undervisningen)	Spørgeskemaundersøgelse til skoleledere Casestudier, herunder interview med skoleledere, lærere m.fl.	Ikke bekræftet -
Kommunerne/forvaltningerne		
Skolerne og forvaltningens aktive samarbejde/dialog om skolens testresultater fører til en målretning af skolernes arbejde (og dermed forbedrede faglige resultater)	Spørgeskemaundersøgelse til skoleledere og forvaltninger Casestudier, herunder interview med forvaltninger, skoleledere m.fl.	Ikke bekræftet -
Et klart kommunalt fokus på faglige resultater har en positiv indflydelse på skolernes faglige niveau	Spørgeskemaundersøgelse til skoleledere og forvaltninger Casestudier, herunder interview med forvaltninger, skoleledere m.fl.	Delvist bekræftet +

Evalueringen viser, at **skolelederne** generelt gør brug af de nationale test, om end med betydelig variation. De orienterer sig i og analyserer testresultaterne *over tid* primært på skole- og klasseniveau. På de skoler, hvor skolelederen udtrykker høje faglige forventninger til eleverne, er der en tendens til, at skolelederne også orienterer sig i resultater på elevniveau. De nationale test indgår således til en vis grad som en del af skoleledernes kvalitetsarbejde.

Evalueringen dokumenterer stor variation blandt skolerne i forhold til, dels om der er klare retningslinjer for brug af de nationale test, dels *hvordan* testresultater anvendes. Nogle skoleledere bruger de nationale test i meget begrænset omfang og fungerer primært som støttende for lærerne i forhold til det praktiske arbejde, men uden at indgå i dialog med lærerne om testresultaterne. Andre skoleledere bruger resultaterne til monitorering og overblik over det faglige niveau, mens en tredje gruppe skoleledere *handler aktivt* og lægger testresultaterne til grund for tildeling af ressourcer, ændret fagfordeling, sparring til lærere mv. Med udgangspunkt i de eksempler, hvor skoleledelsen meget aktivt rammesætter og anvender de nationale test som en del af dokumentationsgrundlaget (sammen med øvrig evaluering), vurderer evaluator, at der er et uudnyttet potentiale for nationale test på en stor andel skoler, der med mere aktivt fokus på anvendelse af nationale test vil kunne stimulere yderligere positive tiltag til at fremme af elevernes faglige niveau. Endelig viser evalueringen, at der blandt de danske skoleledere generelt er en positiv holdning til de nationale test og testenes potentialer. Det er dog de færreste, der finder, at testene er et værktøj til dialog med de kommunale forvaltninger.

Evalueringen dokumenterer, at **kommunerne** generelt stiller krav til skolernes faglige resultater. Resultaterne af de nationale test er for kommunerne en væsentlig indikator for skolernes faglige resultater, men hovedkonklusionen er, at resultater fra de nationale test er "en del af paletten", som det udtrykkes af flere kommuner. Kommunerne interesserer sig således for skolernes faglige

resultater (og vurderer også, at resultater er vigtigere end rammebetingelser og pædagogiske processer), men de nationale test er blot ét ud af flere værktøjer til at sammensætte et billede af en skoles faglige niveau. Dette kan være en hovedårsag til, at få kommuner angiver at bruge testresultater aktivt i dialog med skoleledelsen. Evalueringen viser endvidere stor variation på tværs af kommuner i forhold til anvendelsen af de nationale test – uden at disse forskelle dog kan knyttes systematisk til kommunestørrelse. Det samlede billede er, at kommunerne overvejende bruger de nationale test som et monitoreringsredskab, der kan skabe et overblik over faglige resultater. Omvendt er testresultater sjældent omdrejningspunktet for kommunernes kvalitetsarbejde, hvor de inddrages i dialog med skolelederne med henblik på fremadrettet pædagogisk handling. Evalueringen viser, at der er et potentiale for at fremme evalueringskulturen og den positive udvikling i elevernes faglige niveau ved en mere aktiv forvaltningsadfærd i dialog med skoleledere om, hvordan lederne fremmer evalueringskulturen.

7. TESTENES INDHOLD

Dette kapitel indeholder en analyse af de nationale tests indhold. Kapitlets to første afsnit sætter fokus på indhold, udformning og kvalitet i fire udvalgte testfag (dansk, læsning, fysik/kemi, engelsk samt matematik), mens det sidste afsnit belyser tekniske og økonomiske fordele og ulemper ved det adaptive princip.

For begge dele af kapitlet gælder, at analyserne er gennemført i samarbejde med faglige eksperter på de pågældende områder.

Analysen af indhold, udformning og kvalitet baserer sig på viden, indhentet gennem seminarer med faglige eksperter, og individuelle interview med nøglepersoner. Både eksperter og nøglepersonerne er udvalgt, fordi de besidder viden om og erfaring med nationale test i relation til et af de fire udvalgte fag. Evaluatoren har udarbejdet et fagnotat for hver af de fire udvalgte fag, som sammenfatter viden fra ekspertseminarerne. Nærværende kapitel sætter fokus på de generelle tendenser på tværs af de fire fag. For en mere detaljeret og fagspecifik beskrivelse af opmærksomhedspunkterne, se bilag 5, som således udgør et supplement til nærværende kapitel.

Analysen af tekniske og økonomiske fordele og ulemper ved det adaptive princip baserer sig på et internationalt litteraturstudie og analyse udarbejdet til nærværende evaluering af professor Peter Allerup fra Aarhus Universitet. Denne analyse afrapporteres i et selvstændigt notat vedlagt som bilag 4 til evalueringsrapporten.

Som supplement til de faglige eksperters analyser inddrages viden fra evalueringens øvrige datakilder til belysning af temaet i nærværende kapitel.

7.1 Testenes indhold, udformning og kvalitet

Dette afsnit sætter fokus på indhold, udformning og kvalitet af de nationale test. Mere specifikt behandler afsnittet fag og profilområder, trinmål, opgaveemner, antal opgaver, opgavetyper, testopgavernes kvalitet samt udvikling af opgaver.

7.1.1 Fag og profilområder

De faglige eksperter er i forbindelse med de gennemførte seminarer blevet bedt om at forholde sig til, hvorvidt det fag, de hver især repræsenterer, er egnet som nationalt testfag. Alle fire udvalgte testfag vurderes at være relevante og testegnede. Dog kan der også peges på en række fagspecifikke svagheder, som beskrives nærmere nedenfor og i bilag 5.

Hver test er bygget op om tre profilområder, der afspejler testbare indholdsområder af faget. Disse fremgår af tabellen nedenfor.

Tablet 7.1: Oversigt over profilområder

Dansk, læsning	Matematik	Fysik/kemi	Engelsk
Afkodning	Tal og algebra	Anvendelse og perspektiver	Læsning
Sprogforståelse	Geometri	Energi og energiomsætning	Ordforråd
Tekstforståelse	Matematik i anvendelse	Fænomener, stoffer og materialer	Sprog og sprogbrug

De faglige eksperter vurderer på tværs af fag, at profilområderne generelt set afspejler nogle, for lærerne, relevante og genkendelige kerneelementer. Eksperterne understreger, at der – i overensstemmelse med intentionen med testene – netop er tale om elementer af pågældende fag og ikke hele faget. Eksperterne tilkendegiver desuden, at de valgte profilområder i høj grad rummer fagenes testbare områder.

I forlængelse af ovenstående fremgår det dog også, at der eksisterer nogle fagspecifikke udfordringer i relation til profilområderne i matematik, engelsk og fysik/kemi.

Eksperterne i matematik peger på, at der er et efterslæb i de nationale test i matematik. I forbindelse med revisionen af Fælles Mål (fagenes indholdsbeskrivelse) i 2009 skete der store ændringer af fagets indhold, herunder at det centrale kundskabs- og færdighedsområde *matematik i anvendelse* blev erstattet af *statistik og sandsynlighed*. Eksperterne vurderer, at denne ændring endnu ikke er slået igennem indholdsmæssigt i de nationale test i faget.

Eksperterne i engelsk tilkendegiver, at det er en styrke, at *ordforråd* har fået en selvstændig status i testen, da dette element er afgørende for udviklingen af de fire færdigheder (lytte, læse, tale og skrive), der tegner faget. Det kan virke ulogisk, at *ordforråd* og *sprog og sprogbrug* udgør hver deres profilområde, da *ordforråd* i Fælles Mål er et delelement af det centrale kundskabs- og færdighedsområde *sprog og sprogbrug*. Det er dog eksperternes vurdering, at det ikke giver problemer i praksis for lærerne.

I relation til testen i fysik/kemi påpeger eksperterne, at det potentielt kan skabe forvirring blandt lærerne, at testen indeholder tre profilområder, når dette sammenholdes med, at faget i Fælles Mål har fire centrale kundskabs- og færdighedsområder⁸⁶. Eksperterne peger derudover på, at der knytter sig en særlig udfordring til profilområdet *energi og energiomsætning*. De forklarer det med, at dette tema, modsat de to andre profilområder, ikke afspejler faget bredt. En konsekvens af dette kan være, at *energi og energiomsætning* kan få uforholdsmæssig stor vægt i undervisningen, hvis undervisningen tilrettelægges efter testens indhold.

At der ikke er fuld overensstemmelse mellem profilområder og centrale kundskabs- og færdighedsområder, eller at *energi og energiomsætning* har fået uforholdsmæssig stor vægt, er kun i meget begrænset omfang blevet italesat af lærerne i faget, når de er blevet interviewet i forbindelse med de gennemførte casestudier. Evaluator kan derfor ikke bekræfte, at det ud fra en lærersynsvinkel udgør et problem. Dette er dog ikke ensbetydende med, at undervisningen aldrig tilrettelægges efter testenes indhold, som det fremgår af afsnit 4.1.

Eksperterne i dansk, læsning vurderer, at profilområderne i dette fag er præcise og dækkende. Sidstnævnte begrundes med, at testen i dansk, læsning er en læsetest, og at den derfor tester en afgrænset færdighed.

Som supplement til eksperternes vurdering er lærerne i forbindelse med spørgeskemaundersøgelsen blevet bedt om at forholde sig til, hvorvidt de tre profilområder for det testfag, de underviser i, afspejler faget på en genkendelig måde. Lærernes besvarelser fremgår af tabellen nedenfor.

Tabel 7.2: I hvilken grad er du enig eller uenig i følgende udsagn om de nationale test i relation til det testfag, du underviser i? Angivet i pct.

	Helt uenig	Overvejende uenig	Hverken enig eller uenig	Overvejende enig	Helt enig	Ved ikke	I alt (N)
De tre profilområder afspejler faget på en genkendelig måde	5	19	31	38	5	2	100 (2162)
Testopgaverne afspejler væsentlige faglige kompetencer i faget	4	13	32	43	7	1	100 (2162)

Kilde: Spørgeskemaundersøgelse til lærere, 2013.

Tabellen viser, at der generelt set er flere lærere, der er overvejende enige eller helt enige i, at de tre profilområder afspejler faget på en genkendelig måde. Tilsvarende fordeling gør sig gæl-

⁸⁶ De fire centrale kundskabs- og færdighedsområder er:

- 1) Fysikkens og kemiens verden
- 2) Udvikling i naturvidenskabelig erkendelse
- 3) Anvendelse af fysik og kemi i hverdag og samfund
- 4) Arbejds måder og tankegange.

dende i relation til lærernes vurdering af, hvorvidt testopgaverne afspejler væsentlige faglige kompetencer i faget.

Evaluators har endvidere undersøgt, om der er forskelle på lærernes besvarelse af ovenstående, alt efter hvilke fag de repræsenterer. Denne analyse viser en forskel i lærernes vurdering af profilområdernes genkendelighed. Det fremgår, at lærerne i fysik/kemi i langt højere grad er overvejende eller helt uenige i, at de tre profilområder afspejler faget på en genkendelig måde. Dette understøtter således det billede, eksperterne tegner af, at der især i fysik/kemi kan være udfordringer med genkendeligheden af profilområderne i faget.

7.1.2 Trimål

De nationale test afspejler Fælles Mål. Fælles Mål omfatter fagformål, slutmål (centrale kundskabs- og færdighedsområder), trimål samt læseplaner for skolens fag og emner. Testopgaverne relaterer sig inden for hvert af de tre profilområder til udvalgte trimål. Eksperterne er blevet bedt om at vurdere, i hvilket omfang testene afspejler alle relevante testegnede trimål for de enkelte fag.

På tværs af fag vurderer eksperterne, at hovedparten af de testegnede trimål indgår i de nationale test. Det tilkendegives dog samtidig, at der er et mindre antal testegnede trimål, der ikke testes i fagene. Det skyldes bl.a., at de nationale test ikke i alle tilfælde afspejler gældende trimål, jf. ovenfor. Eksperterne peger på, at når dette er tilfældet, skyldes det eksempelvis, at der er elementer af faget, som det ikke er teknisk muligt at teste (fx lytteelementer).

7.1.3 Opgaveemner

Opgaveemnet (tidligere benævnt kendeord) fremgår af den enkelte opgaves metadata, det vil sige de baggrundsinformationer, der er knyttet til hver opgave. Opgaveemnet bruges til at beskrive de konkrete kundskaber eller færdigheder fra Fælles Mål, der afprøves med opgaven inden for hvert af de tre profilområder. Der er tale om et redskab til både opgaveudviklere samt lærere og forældre. Læreren kan se den enkelte opgaves opgaveemne via resultatvisningen på computerskærmen, når en elev har gennemført en test. Eksempler på opgaveemner kan være navneord, udsagnsord, specifikt indhold i tekster samt addition/subtraktion.

Ekspertseminarernes deltagere peger på, at der i forbindelse med testens indførelse blev "bundet" et bestemt antal opgaveemner til fagene, hvilket vurderes at have medført nogle fagspecifikke uhensigtsmæssigheder. Dette er eksempelvis, at der er for få eller for mange opgaveemner til rådighed. Eksperterne forklarer, at hvis der er for få opgaveemner til rådighed, giver det upræcis information til lærere og forældre, der efterfølgende ønsker at undersøge, hvilke opgavetyper eleven har haft udfordringer med eller klarer sig særlig godt i. Er der omvendt for mange opgaveemner til rådighed, er risikoen, at opgaveemnerne ikke kan underbygges fagligt.

Afslutningsvist skal lærernes brug af opgaveemner nævnes. På tværs af fag peger flere af eksperterne på risikoen for, at der sker en overfortolkning af resultaterne, hvis lærerne lægger for meget vægt på netop opgaveemner, når de skal danne sig et billede af elevens faglige udfordringer/styrker. Testen kan alene sige noget om elevens faglige niveau inden for det enkelte profilområde. Eksperterne understreger, at der er for få testopgaver inden for de enkelte opgaveemner til at kunne sige noget mere specifikt om elevens kundskaber og færdigheder, og at det er vigtigt, at lærerne kender til denne begrænsning.

I forbindelse med casestudierne er lærerne blevet bedt om at beskrive, hvordan de analyserer resultater af de nationale test. På baggrund heraf er det evaluators vurdering, at lærerne ikke i særlig høj grad gør brug af muligheden for at se det enkelte testspørgsmåls opgaveemne, jf. afsnit 6.3.

Med afsæt i ovenstående vurderer evaluators, at det kan være meningsfuldt at revidere rammerne for brugen af opgaveemner, så der i højere grad bliver overensstemmelse mellem opgaveemnets ordlyd, og det som den enkelte opgave rent faktisk tester. Det gælder ikke mindst inden for dansk, læsning og matematik.

7.1.4 Antal opgaver

Opgaverne i opgavebanken fordeler sig inden for hvert profilområde på fem niveauer, hvor 5 er det højeste og sværeste niveau, og 1 er det letteste niveau.

REVIEW-panelets rapport fra 2007 fastslår, at minimumsforventningen til opgavebanken er, at der skal være lige mange opgaver til de 10 pct. svageste elever som til de 10 pct. stærkeste elever, hvilket langt fra er tilfældet i dag, da der er mange flere opgaver på niveau 1 end på niveau 5. På tværs af fag tilkendegiver deltagerne på ekspertseminarerne enighed om, at jo bredere opgavebanken repræsenterer stoffet, herunder det svære stof, jo bedre er det.

Evaluators har i forbindelse med evalueringen haft adgang til oplysninger om antallet af opgaver i opgavebanken for hver af de fire testfag⁸⁷. Disse er blevet forelagt eksperterne, der er blevet bedt om at forholde sig til, om der er nok opgaver i henhold til at skabe bredde samt meningsfulde testforløb set fra en lærer-/elevsynsvinkel. Eksperterne vurderer, at der særligt er en udfordring med antallet af opgaver inden for det sværeste niveau (niveau 5), hvor der er flere eksempler på profilområder med kun én til to opgaver eller slet ingen på det sværeste niveau.

Eksperterne peger på, at for få svære opgaver bl.a. medfører, at fagligt stærke elever "tømmer" opgavebanken på niveau 5 og dermed løber tør for opgaver af en sværhedsgrad, der er relevant for netop disse elever. Det kan også medføre, at den enkelte elev møder den samme opgave i de obligatoriske og frivillige test, eller at mange elever i klassen får de samme opgaver.

Sidstnævnte bekræftes af de casestudier, evaluator har gennemført som led i evalueringen. Lærerne fortæller, at de i forbindelse med testafviklinger i klasser med højt fagligt niveau og en forholdsvis fagligt homogen elevgruppe har oplevet, at mange af eleverne får den samme opgave/de samme opgaver i løbet af testforløbet, dog ikke samtidigt. Lærerne fortæller, at de har elever, der er stødt på den samme opgave mere end én gang.

Afslutningsvist skal det fremhæves, at eksperterne peger på, at der mangler viden om, hvad der definerer en opgaves niveau, herunder parametre for hvad der definerer en henholdsvis let, mellem og svær opgave. Det vurderes, at det vil være en viden, der kan kvalificere arbejdet med at udvikle nye opgaver, der tilgodeser de aktuelle behov, herunder at man kan målrette arbejdet med at udforme opgaver på niveau 5. Se endvidere afsnit 7.1.7 om udvikling af nye opgaver.

7.1.5 Opgavetyper

Der er i dag 15 opgavetyper til rådighed for opgaveudviklerne. Disse er fx multiple choice, indsættelsesopgaver, del ord, cloze-test samt farv felter. Evaluator har til brug for evalueringen opgjort antallet af opgaver. Opgørelsen viser, at der er stor variation i antallet af opgaver inden for den enkelte opgavetype. Multiple choice og del ord er nogle af de absolut mest anvendte.

Evaluator har ydermere bedt eksperterne om at forholde sig til fordelingen af opgavetyper samt vurdere, hvorvidt de fungerer hensigtsmæssigt i relation til de færdigheder og kundskaber, man ønsker at afdække inden for det enkelte profilområde.

På tværs af fag vurderer eksperterne, at det, at nogle typer af opgaver anvendes langt hyppigere end andre, skal betragtes som et udtryk for, at det er testens indhold, der kommer først og ikke formen. Eksperternes vurdering er, at man anvender netop de opgavetyper, der tester specifikke kundskaber og færdigheder på den mest hensigtsmæssige måde.

Ser man på tværs af de fire fag peger eksperterne på, at det er problematisk, at de opgaveskabeloner, der er til rådighed i dag, ikke afspejler den teknologiske udvikling, der er sket siden 2007 på både hard- og softwaresiden. Et eksempel er den tidligere nævnte manglende mulighed for at inddrage lytteelementet, men også muligheden for, at eleven fx selv kan tegne streger i koordinatsystem, fremhæves af eksperterne.

Eksperterne er dog enige om, at anvendelsen af mange forskellige opgavetyper i et enkelt testforløb ikke nødvendigvis er en kvalitet i sig selv. De vurderer, at der kan være en risiko for, at

⁸⁷ Opgørelsen bygger på data udleveret af Kvalitets- og Tilsynsstyrelsen december 2012.

eleven kommer til at fokusere for meget på at afkode forskellige opgavetyper, frem for det faglige indhold.

I forlængelse heraf har både eksperterne og flere af de lærere, der er interviewet som led i case-studierne, peget på, at ikke alle opgavetyper indgår i demotesten på ministeriets hjemmeside. De påpeger, at dette er uhensigtsmæssigt, da flere lærere fortæller, at de gør brug af denne mulighed for at gøre eleverne fortrolige med de forskellige typer af opgaver, de vil kunne møde. Ikke mindst på de yngste klassetrin.

7.1.6 Testopgavernes kvalitet

Dette afsnit belyser testopgavernes faglige kvalitet i forhold til en række af de kvalitetsfaktorer, som REVIEW-panelet forholdt sig til i 2007, så som layout, grafik og illustrationer samt distraktorer (svarmuligheder)⁸⁸.

Udgangspunktet for vurderingen har været en række repræsentativt (i relation til opgavetyper, antal opgaver mv.) udvalgte opgaver fra hvert af de fire fag. Dette udvalg af opgaver har på ekspertseminarerne fungeret som afsæt for eksperternes analyse og vurdering af testopgavernes kvalitet.

Ekspertene vurderer, at der er en lang række udfordringer i **layout** af testopgaverne. På tværs af fag peger eksperterne på, at der er flere eksempler på layout af opgaver, der er direkte forstyrrende for elevernes opgaveløsning. De fremhæver eksempelvis opgaver, hvor der er flere overskrifter og linjer med forskellige forklaringer på, hvad opgaven går ud på, samt opgaver, hvor brugen af faglige begreber er inkonsekvent.

Sammenholder man de fire udvalgte testfag, synes udfordringerne størst for engelsk, matematik og fysik/kemi. Det layoutmæssige behov er mindst udtalt i dansk, læsning. Deltagerne i dette ekspertseminar forklarer, at man i længere tid har arbejdet med opgavernes layout og tilstræbt, at overskrift og opgavespørgsmål er én og samme ting, så eleven kun skal orientere sig et sted. Deltagerne konstaterer, at det er slået igennem i langt de fleste opgaver. Indtrykket er derfor, at mange opgaver har en stringent, ensartet og overskuelig opbygning. Det er derfor i læsetesten, man finder de færrest layoutmæssige udfordringer, når man ser bort fra opgavetekster med spal-tebrud.

En del af testopgaverne i opgavebanken indeholder **grafik og illustrationer** i form af tegninger, fotografier, affotograferinger, tabeller mv. Også her fremhæver eksperterne, at der er mange eksempler på brug heraf, der virker forstyrrende frem for hensigtsmæssige. Dette kan eksempelvis være en utidssvarende tegnestil og forældede illustrationer. Eksperterne tilkendegiver, at grafik og illustrationer er relevant i de testopgaver, hvor det indgår som en naturlig del af opgaveløsningen, mens det bør fjernes i de tilfælde, hvor det kun er til pynt.

Distraktorer er de svarmuligheder, som eleven præsenteres for som en del af testspørgsmålet. De er et afgørende element i forhold til testens validitet. Deltagerne på tværs af ekspertseminarer peger på en række udfordringer ved distraktorerne. Eksperterne fremhæver især, at det er problematisk, at det i nogle opgaver kan konstateres, at flere svarmuligheder rent faktisk er korrekte, men kun ét af svarene accepteres som rigtigt. Det skaber frustration hos fagligt stærke elever, da de ikke har mulighed for at afgive det svar, som de ønsker. Desuden er der eksempler på inkonsekvent brug af faglige begreber både i opgaveformuleringer og svarmuligheder.

Som supplement til eksperternes vurderinger af testopgavernes kvalitet er lærerne i spørgeskemaundersøgelsen blevet bedt om at vurdere, hvorvidt testopgavernes faglige kvalitet er tilfredsstillende, og om testopgaverne er formuleret entydigt og klart. Besvarelsene fremgår af tabellen nedenfor.

⁸⁸ Rapport fra REVIEW-panelet, 2007.

Tabel 7.3: I hvilken grad er du enig eller uenig i følgende udsagn om de nationale test i relation til [TESTFAGET]? Angivet i Pct.

	Helt uenig	Overvejende uenig	Hverken enig eller uenig	Overvejende enig	Helt enig	Ved ikke	I alt (N)
Testopgavernes faglige kvalitet er tilfredsstillende	7	24	28	34	6	1	100 (2162)
Testopgaverne er formuleret på en måde, så de er entydige og klare	9	27	31	28	3	1	100 (2162)

Kilde: Spørgeskemaundersøgelse til lærere, 2013.

Tabellen viser, at der generelt set er flere lærere, der er overvejende enige eller helt enige i, at testopgavernes faglige kvalitet er tilfredsstillende. Modsat er der flere lærere, der er overvejende eller helt uenige i, at testopgaverne er formuleret på en måde, så de er entydige og klare.

I forhold til, om testopgaverne er formuleret på en måde, så de er entydige og klare, adskiller lærerne sig fra hinanden. Her svarer kun 23 pct. af matematiklærerne, at de er overvejende eller helt enige, hvor det er hele 40 pct. af hhv. biologi-, geografi- og engelsklærerne, der svarer, at de overvejende er enige eller helt enige.

Lærernes vurderinger understøtter således eksperternes vurderinger af, at der knytter sig en række udfordringer til testopgavernes kvalitet.

7.1.7 Udvikling af opgaver

Testopgavernes faglige kvalitet og udformning hænger tæt sammen med rammerne for udviklingen af de enkelte opgaver. Tilbagemeldinger fra eksperterne giver grundlag for at pege på en række opmærksomhedspunkter i relation hertil. Eksperterne peger på, at opgaveudviklerne i dag ikke har adgang til opgaveproduktionens maskinrum. Det vil sige adgang til at se de opgaver, der kasseres i Rasch-analyse⁸⁹. De tilkendegiver, at hvis det var en mulighed, kunne fremtidig opgaveudvikling bygge på konkret viden om, hvorfor en opgave kasseres eller ej. Opgaveudviklerne ville fremover kunne være mere præcise og i højere grad undgå faldgruber.

Derudover vurderer eksperterne, at der på nuværende tidspunkt i høj grad udvikles opgaver ud fra den enkelte udviklers mavefornemmelse, hvilket ikke betragtes som optimalt. Der er, som tidligere nævnt, heller ikke lavet analyser af, hvad der definerer sværhedsgraden inden for de fem niveauer, hvilket vanskeliggør udviklingsarbejdet.

På baggrund af ovenstående vurderer evaluator, at der synes at være et behov for i endnu højere grad at klæde opgaveudviklerne fagligt på, end det sker i dag. Eksperterne nævner som eksempel England og Holland, hvor opgaveudviklere gennemgår en systematisk uddannelse, så de er opdateret med den nyeste viden inden for digitale/adaptive test. Det er ligeledes evaluators vurdering, at der synes at være et behov for, at opgaveudviklingen i langt højere grad knyttes til forskning, sådan at testen og ikke mindst udviklingen af opgaverne bygger på faktuel viden. Et eksempel på relevant forskning kunne være viden om, hvordan elever agerer foran en skærm i en testsituation, eller hvilke specifikke opgavetyper der bedst tester konkrete færdigheder og kompetencer, som fx ordkæder i relation til afkodning.

7.2 Fordele og ulemper ved det adaptive princip

I dette afsnit præsenteres evaluators vurdering af tekniske og økonomiske fordele og ulemper ved det adaptive princip, der benyttes i de nationale test. Vurderingen er fortrinsvis baseret på et internationalt litteraturstudie og en analyse udarbejdet til nærværende evaluering af professor Peter Allerup fra Aarhus Universitet⁹⁰. Dertil kommer indsamlede data om ressourceforbrug til brug for vurdering af økonomiske fordele og ulemper samt enkelte referencer til evaluators egne

⁸⁹ Se afsnit 7.2.2 for definition.

⁹⁰ Analysen afrapporteres i selvstændigt notat vedlagt som bilag 4 til evalueringsrapporten.

datakilder (spørgeskemaundersøgelse og casestudier). De pædagogiske fordele og ulemper ved det adaptive princip behandles særskilt i afsnit 8.2.

7.2.1 Kort beskrivelse af det adaptive princip

Det adaptive princip, som de nationale test er funderet på, er en procedure for en aktiv sammensætning af opgaver til eleverne, hvor hver opgave udvælges af testsystemet på baggrund af elevens besvarelse af foregående opgaver. Testsystemet beregner således for den enkelte elev, hvilken opgave eleven skal have ud fra et princip om at give en opgave med en passende sværhedsgrad beregnet ud fra elevens tidligere besvarelser. Alle elever får altså et individuelt sammensat opgavesæt, kreeret undervejs i testafviklingen i modsætning til eksempelvis folkeskolens afgangsprøve, hvor alle får samme opgaver.

Det adaptive princip fungerer i praksis ved, at eleven indledningsvis stilles en middelsvær opgave. Afhængigt af, om eleven besvarer opgaven rigtigt eller forkert, udvælges den næste stillede opgave, som så er sværere eller lettere end den først stillede opgave. Herefter fortsætter systemet med at udvælge opgaver efter samme princip, indtil den statistiske Standard Error of Measurement (SEM) (som bliver systematisk mindre med antallet af opgaver eleven stilles) i beregningen af elevdygtigheden er reduceret så meget, at eleven bør kunne nå et resultat for alle tre profilområder på 45 minutter. De 45 minutter anvendes som alternativt stopkriterium. Teoretisk vil en stor andel elever afslutte testen væsentligt hurtigere end 45 minutter. Tidspunktet for, hvornår elevens test afsluttes, bestemmes af, hvor hurtigt eleven opnår et konsistent svarmønster, som gør systemet i stand til at fastslå elevens faglige niveau. Litteraturstudiet fastslår, at dette vil ske med anvendelse af ca. 50 pct. færre opgaver end det, der kræves med lineære test. Litteraturstudiet fastslår samtidig, at det ikke er muligt at konstruere et testsystem, hvor stopkriteriet nås for samtlige elever. Dette bekræftes af de gennemførte casestudier, hvor evaluator har erfaret, at testen for et mindre antal elever fortsætter ud over de 45 minutter.

7.2.2 Teknisk om det adaptive princip

Det adaptive princip udmøntes ved to parametre: itemsværhed og elevdygtighed. Det adaptive princip betyder, at testsystemet søger at udvælge opgaver, der kan fastlægge elevens faglige niveau hurtigt. Systemet vurderer elevens dygtighed baseret på besvarelse af de stillede opgaver og beregner konstant, hvilken ny opgave eleven skal have. Det sker på baggrund af rigtige og forkerte besvarelser af tidligere opgaver. Med den rette sammensætning af itemsværhed⁹¹ til den demonstrerede elevdygtighed kan det adaptive princip oftest hurtigt og præcist fastlægge elevens faglige niveau (som nævnt ovenfor, er det ca. 50 pct. af det antal opgaver, der kræves med lineære test). Idet det adaptive princip betyder, at der opgave for opgave successivt opbygges viden om elevens dygtighedsniveau, sker dette med maksimal information, hvorved det adaptive princip sikrer, at de nationale test fremskaffer sikker viden om elevens dygtighedsniveau med høj præcision og med færrest mulige opgaver.

Adaptive test er baseret på en psykometrisk metode, det vil sige en statistisk model, som betegnes Item Response Theory (IRT). De nationale test er baseret på en Rasch-model, der er en sådan IRT-model, karakteriseret ved at indeholde itemsværhed og elevdygtighed.

Alle opgaver i de nationale test er udviklet med Rasch-analyse. Det vil sige, at alle opgaver udvikles af eksperter og testes blandt et stort antal elever (ca. 700), hvorefter resultaterne gennemgår en Rasch-analyse for at konstatere, om opgaven er anvendelig til at indgå i opgavebanken (måler opgaven det, den skal måle, på hvilket fagligt niveau mv., hvilket sker med diverse analyser af robusthed i elevbesvarelser, standardafvigelse mv.).

Opgaver udvikles efter et såkaldt homogenitetskriterium, det vil sige, at en hvilken som helst opgave i opgavebanken kan erstatte en hvilken som helst anden opgave (dog således at hvert profilområde skal dækkes for den enkelte elev). Samtlige opgaver inden for hvert område adskiller sig således alene ved deres sværhedsgrad, som er testet i afprøvning, hvorefter resultaterne er Rasch-analyseret.

⁹¹ Se afsnit 7.1.4.

Brug af homogenitetskriteriet (hvilket ikke anvendes i alle internationalt anvendte adaptivt baserede test) betyder også, at sammenligninger mellem elever, grupper af elever og mellem elever over år er valide og kan gennemføres på én skala. I modsætning hertil vil test baseret på ikke-homogene opgavebanker ikke kunne sammenligne elever, da eleverne vil være stillet ikke bare forskellige opgaver, men forskellige opgaver uden systematisk etableret ensartethed i sværhedsgrad. Det adaptive princip, baseret på homogenitetskriteriet for opgavebanken, sikrer således de nationale test en stor psykometrisk fordel, men homogenitetskriteriet er ikke en nødvendighed. Litteraturstudiet konstaterer dog: "Hvis man slækker på disse homogenitetskrav, umuliggør man valide sammenligninger af elevernes præstationer i lys af IRT."

Resultatet af den nationale test beregnes og kommunikeres ved, at den enkelte elev placeres i en gruppe, der omfatter hhv. 10 pct., 25 pct., 30 pct., 25 pct. og 10 pct. af samtlige elever, der har taget testen (svarende til intervallerne 1-10, 11-35, 36-65, 66-90, 91-100), der igen svarer til et udmeldt niveau 1-5. Ved tilbagemeldingen til lærerne vises resultatet på en skala fra 1-100 (percentilskala). Læreren oplyses om elevens placering på percentilskalaen for hvert profilområde og for testen som helhed. Besvarelsen af hver enkelt testopgave er endvidere tilgængelig⁹². Systemet oplyser ikke det samlede antal rigtigt besvarede opgaver pr. elev, idet eleverne ikke løser de samme opgaver, hvorved sammenligning af antallet af korrekte besvarelser ikke giver mening.

Litteraturen peger på, at antallet af opgaver af middelsvær karakter skal være størst, mens et mindre antal skal være hhv. svære og lette. Antallet af opgaver i opgavebanken for nationale test er fordelt således. Dog viser evaluators opgørelse, at der er for få meget svære opgaver inden for specifikke profilområder.

7.2.3 Økonomisk om det adaptive princip

Anvendelse af det adaptive princip i test kræver omfattende ressourcer i udviklings-, afprøvnings- og analysefasen sammenholdt med lineære test. Det skyldes, at alle opgaver skal gennemgå en Rasch-analyse med henblik på at etablere et validt grundlag, sådan at opgaverne kan benyttes vilkårligt i testen.

Der udøves således et omfattende arbejde med at:

- Udvikle opgaver i opgavekommissioner og med kvalitetssikringsprocedurer ved eksternt indkøbte eksperter og embedsmænd i ministeriet
- Afprøve opgaverne blandt et stort antal elever. I de nationale test sker dette blandt 700 elever, hvilket er et minimum for at sikre, at afprøvningen foregår på et validt niveau. Afprøvning sker på udvalgte skoler. Opgaverne distribueres digitalt under hensyn til nøje definerede kvalitetssikringsprocedurer og afprøves på skolerne. I 2012 deltog hele 361 skoler i afprøvning af nye opgaver med deltagelse af i alt 15.512 elever
- Udsætte de afprøvede opgaver for en Rasch-analyse i UNI-C, hvor de psykometriske egenskaber nøje vurderes i forhold til kravene i Rasch-modellen. Især dette er stærkt arbejdskrævende og fordrer statistisk faglig indsigt i analyseresultater ud fra kontrol af IRT-modellen. Opgaven forkastes eller accepteres til medlemskab af opgavebanken.

Forkastede opgaver, eller opgaver, der ønskes ændret, selv meget små ændringer i overskrift og lignende, skal gennemgå samme proces som beskrevet i ovenstående. Arbejdet med udvikling af en opgave fra første udkast til den indlemmes i opgavebanken tager ca. et år.

Der er altså tale om en omfattende anvendelse af personaleressourcer i ministeriet og UNI-C, køb (aflønning) af eksterne eksperter til opgavekommissioner og kvalitetssikring, ressourcer på skoler, der afprøver opgaver (lærer-, elev- og administrative ressourcer), samt tidsmæssige ressourcer (et års udviklingstid) for at skabe et validt grundlag for en adaptiv procedure.

Litteraturstudiet peger på, at det er muligt at forenkle dele af processen noget ved at medtage opgaver til afprøvning i ordinær testafvikling, dog således at disse opgaver ikke indgår i beregningen af elevens faglige niveau i testen. Opgaver til afprøvning blandes altså med de opgaver,

⁹² Se kap. 8 for særskilt analyse af resultatvisningen.

der er i opgavebanken, men opgaverne til afprøvning benyttes kun til Rasch-analyse efterfølgende med henblik på at vurdere opgavens egnethed til at indgå i opgavebanken fremover.

Det skal nævnes, at det forhold, at de nationale test er digitale, indebærer betydelige ressourcer i udvikling og vedligeholdelse af it-systemet, men evaluator vurderer, at papirbaserede test formentlig vil medføre endnu højere udgifter i den samlede testafvikling, fordi lærerne herved vil skulle rette, score og foretage rapporteringer af testresultater med en betydelig højere samlet ressourceindsats, end den, der samlet medgår til den model, der benyttes for de nationale test. Litteraturstudiet peger på, at den tid, der spares ved adaptive test, ved at eleverne generelt meget hurtigere end i lineære test når stopkriteriet, altså hvor testsystemet har beregnet og fastslået elevens faglige niveau, generelt "ikke kan kapitaliseres til at udgøre en ressource for andre undervisningsrelaterede aktiviteter", som det hedder. Dette skyldes, at de fleste skoler ikke påbegynder læringsaktiviteter for eleverne i takt med, at de afslutter testen.

Tilsvarende er det ikke muligt via litteraturstudiet at opgøre effekten (i økonomisk værdi) af, at relativt fagligt svagere og stærkere elever opnår højere motivation ved de adaptive test end ved lineære test.

Litteraturstudiet konkluderer, at "det er en klar samlet vurdering, at vedligeholdelsen af de nationale test kræver flere ressourcer end udvikling af fra-gang-til-gang lineære test."

Evaluator vurderer på baggrund af litteraturstudiet, at der samlet set er både fordele og ulemper ved adaptive test. De vigtigste fordele er øget præcision i etablering af vurdering af elevernes faglige niveau med brug af færre opgaver end lineære test og bedre sammenligningsgrundlag mellem elever. Af ulemper viser analysen, at adaptive test er betydeligt mere ressourcekrævende end lineære test, og udviklingstid af nye opgaver og selv simpel tilretning af eksisterende opgaver er på ca. et år, hvilket gør vedligeholdelse af opgavebanken på et anvendeligt niveau meget tung, hvorved systemet har en lav agilitet og fleksibilitet.

7.3 Delkonklusion

Dette kapitel har belyst en række evalueringsspørgsmål under evalueringstemaet: *Testens indhold*. Kapitlet har med udgangspunkt i tilbagemeldinger fra interview med nøglepersoner og eksperter inden for fire testfag (dansk, læsning, matematik, engelsk og fysik/kemi) analyseret en række udvalgte elementer, der afspejler forskellige dimensioner af testens indhold, herunder **fag**, **profilområder**, **faglige emner** og **relationen til Fælles Mål** samt testopgavernes faglige kvalitet i form af fx **opgavetyper**, **layout** og **grafik og illustrationer**. Afslutningsvist er de **tekniske** og **økonomiske** fordele og ulemper ved det adaptive princip blevet analyseret med udgangspunkt i et internationalt litteraturstudie.

Opgavernes indhold og udformning

Generelt kan det konkluderes, at valg af testfag og profilområder samlet set synes hensigtsmæssige, om end der er mindre opmærksomhedspunkter.

Ser man på koblingen mellem testspørgsmål og Fælles Mål, er konklusionen, at testene i hovedtræk afspejler testegnede trinmål. Testspørgsmålene i to af fagene er dog ikke opdaterede (matematik og fysik/kemi), så de afspejler gældende faglige mål, hvilket må vurderes at være uhenigtsmæssigt.

En væsentlig konklusion er endvidere, at der inden for flere testfag er en mangel på svære opgaver i opgavebanken. En optælling af opgaver udført af evaluator viser, at der er ganske få eller ingen svære opgaver inden for flere profilområder i både engelsk, dansk, læsning samt matematik. Antallet af opgaver i opgavebanken synes derfor at være en udfordring.

I forhold til testopgavernes faglige kvalitet, viser analysen, at der er en række opgavetyper til rådighed, og de anvendes i varierende grad. Multiple choice er én af de typer af opgaver, der er meget populær. Eksperterne pointerer dog, at stor variation i brugen af opgavetyper ikke er afgørende. Det væsentligste er, hvordan man tester specifikke kundskaber og færdigheder på den mest hensigtsmæssige måde.

I forlængelse heraf er det en generel udfordring, at opgaveskabelonerne (som er rammen for udvikling og design af opgaver) ikke afspejler den teknologiske udvikling, der er sket siden 2007.

Det kan ligeledes konkluderes, at der er behov for en gennemgang af udvalgte testopgaver i forhold til at sikre hensigtsmæssigt layout, brug af illustrationer samt distraktorer. Behovet for forbedringer er dog mindst udtalt i dansk, læsning.

Endelig viser analysen af testens indhold, at det vil være hensigtsmæssigt at forbedre rammerne for opgaveudviklingen, bl.a. i form af systematisk uddannelse af opgaveudviklere samt stærkere tilknytning til forskning.

Tekniske og økonomiske fordele og ulemper ved det adaptive princip

Analysen af de tekniske og økonomiske fordele og ulemper ved det adaptive princip er som nævnt, sket på baggrund af et litteraturstudie.

Den tekniske analyse giver grundlag for at konkludere, at det adaptive princip sikrer, at de nationale test fremskaffer sikker viden om elevens faglige niveau med høj præcision og med færrest mulige opgaver.

Brug af homogenitetskriteriet sikrer samtidig, at sammenligninger mellem elever, grupper af elever og mellem elever over år er valide og kan gennemføres på én skala.

Den økonomiske analyse viser, at anvendelse af det adaptive princip i test kræver omfattende ressourcer i udviklings-, afprøvnings- og analysefasen sammenholdt med lineære test. Det skyldes, at alle opgaver skal gennemgå en Rasch-analyse. Litteraturstudiet viser dog, at det er muligt at forenkle dele af processen noget ved at medtage opgaver til afprøvning i ordinær testafvikling.

Konklusionen er her, at de adaptive test er betydeligt mere ressourcekrævende end lineære test, og systemet har en lav agilitet og fleksibilitet.

8. TESTENES UDFORMNING

Dette kapitel belyser en række evalueringsspørgsmål, der er samlet under følgende overordnede evalueringstema: *Testens udformning*. Kapitlet vil indeholde en analyse af forskellige elementer i relation hertil i form af **den praktiske tilrettelæggelse** af testafviklingen på skolerne, herunder tids- og ressourceforbrug, testsystemets **resultatvisning** samt en vurdering af **det adaptive princip** set ud fra et **undervisningsmæssigt/pædagogisk perspektiv**.

Også i dette kapitel vil tilbagemeldinger fra seminar med faglige eksperter (jf. kapitel 7) indgå som en del af datagrundlaget, der derudover bygger på spørgeskemaundersøgelserne til skoleledere og lærere, de tidligere nævnte casestudier, herunder observationer af testafviklingen, samt det i afsnit 7.2 omtalte litteraturstudium.

Tilrettelæggelse, gennemførelse og brug af resultatvisningen er alle elementer, der indgår som grundlæggende forudsætninger i evaluators forandringsteori, dvs. de ressourcer, aktiviteter og output, der ligger forud for de forventede resultater og virkninger. Der henvises til forandringsteorien i bilag 1. Dette kapitel er derfor en analyse af disse grundlæggende forudsætninger.

- 8.1 Den praktiske tilrettelæggelse af testene på skolerne, herunder selve gennemførelsen**
Nærværende afsnit indeholder en vurdering af den praktiske tilrettelæggelse af testene på skolerne, herunder booking af test, den tekniske klargøring, selve gennemførelsen samt opfølgningen i form af analyse af resultater alene eller i dialog med klasse-, årgangs- eller fagteam samt faglige vejledere. Analysen vurderer fordele og ulemper set ud fra et tidsmæssigt og teknisk perspektiv.

Evaluators har i spørgeskemaundersøgelsen spurgt ind til lærernes tidsforbrug før, under og efter testafvikling.

Tabel 8.1: Angivelse af gennemsnitlig tidsforbrug pr. test fordelt på aktiviteter. Angivet i pct.

	0 timer	1-3 timer	4-6 timer	7-9 timer	10 timer eller mere	I alt (N)
Planlægning (booking, orientering af forældre mv.)	6	88	5	1	0	100 (2162)
Gennemførelse (til og fra lokale, gennemførelsen af selve testen mv.)	1	91	8	1	0	100 (2162)
Analyse og fortolkning af resultatet	1	63	29	5	2	100 (2162)
Samtaler med fx årgangs- eller fagteam eller med ressourcepersoner i form af læsevejledere, matematikvejledere og lign. om resultatet	16	78	6	1	0	100 (2162)
Elevsamtaler om resultatet	9	58	27	5	2	100 (2162)
Orientering af forældre om resultatet	4	70	18	6	2	100 (2162)
Samtaler med skoleledelsen om resultatet	48	51	1	0	0	100 (2162)

Kilde: Spørgeskemaundersøgelse til lærere, 2013.

Af tabellen fremgår det, at langt størstedelen af lærerne (88 pct.) bruger 1-3 timer på planlægning af testafvikling, herunder booking, orientering af forældre mv. 6 pct. af lærerne bruger ingen tid forud for testafvikling.

Tabellen viser endvidere, at 91 pct. af lærerne bruger 1-3 timer på selve testgennemførelsen, hvilket inkluderer tid til, at eleverne kommer til og fra lokalet samt gennemførelse af selve testen. Der er tale om et forholdsvis stort tidsspænd. På baggrund af de kvalitative data er det muligt at komme med en mere præcis vurdering af tidsforbruget, som det fremgår nedenfor.

Aktiviteter efter testafvikling er i tabellen delt op i analyse og fortolkning af resultat, samtaler med fagteam og/eller ressourcepersoner, elevsamtaler om resultat, orientering af forældre om resultat samt samtaler med skoleledelsen om resultat. Det fremgår, at størstedelen af lærerne bruger 1-3 timer på de forskellige elementer af opfølgning. Besvarelserne, der er fordelt på fag, viser, at lidt over en tredjedel af dansk- og matematiklærerne bruger mellem 4-6 timer på analyse og fortolkning af resultaterne. Dette tal er noget lavere for fysik/kemi og biologi, hvor kun 17-18 pct. af lærerne tilkendegiver, at de bruger dette antal timer.

Ser man på tværs af indskoling, mellemtrin og udskoling, er der en tendens til, at tidsforbruget generelt er mindre i udskoling.

Et spørgsmål i tabel 8.1 skiller sig særskilt ud. Det er spørgsmålet vedrørende lærerens samtaler med skoleledelsen om resultatet. Her tilkendegiver 48 pct. af lærerne, at de bruger 0 timer på en sådan samtale.

Nedenfor sammenholdes resultatet af spørgeskemaundersøgelsen med resultatet af casestudierne, herunder evaluators observationer.

Før testen

Det fremgår af tabellen ovenfor, at 88 pct. af lærerne bruger 1-3 timer på planlægning af testen, herunder booking af test, orientering af forældre mv.

I forbindelse med casestudierne er lærerne dels blevet bedt om at angive det konkrete tidsforbrug til planlægning (i timer), dels at vurdere om gennemførelse af en national test samlet set tager længere eller kortere tid end andre typer af test.

Lærerne tilkendegiver, at de ofte bruger lidt længere tid på den praktiske tilrettelæggelse forud for testafvikling i indskoling sammenholdt med udskoling. Dette skyldes ikke mindst, at eleverne i udskoling er blevet fortrolige med de nationale test og derfor ikke har behov for en længere introduktion til login, opgavetyper mv.

Det samme billede tegner sig i forhold til orientering af forældre. Også her bruges der mest tid i indskoling, da det for mange forældre er første gang, de (og deres børn) stifter bekendtskab med de nationale test. Af samme grund er der også her flere forældrehenvendelser. Dette stemmer overens med den generelle analyse af skole-hjem-samarbejdet, jf. kapitel 5.

Når lærerne bliver bedt om at beskrive, hvad de bruger tid på forud for testafvikling, nævner de fleste på tværs af årgange, at de introducerer eller opsummerer det adaptive princip for eleverne, træner demotest samt skriver praktiske oplysninger vedrørende testafvikling på forældre-intra.

En gruppe af lærerne forklarer, at der i nogle klasser er et stort behov for at gå i dybden med det tekniske (herunder det adaptive princip), da flere elever bliver utrygge og nervøse (se kapitel 5 for uddybning af dette). Andre lærere fortæller, at de bruger lidt umiddelbart inden testafviklingen på eksempelvis at repetere strategier for orddeling sammen med eleverne.

Generelt er det evaluators vurdering, at anvendelsen af tid forud for testen er for nedadgående sammenlignet med tidsforbruget ved testens indførelse, da eleverne på de fleste klassetrin efterhånden er bekendt med testen. Tidsforbruget inden testen italesættes heller ikke som en ulempe.

Booking af tid

I casestudierne har evaluator spurgt ind til, hvem der booker tid til test. På ca. halvdelen af skolerne er det en repræsentant fra ledelsen, der har ansvaret for at booke til samtlige klasser. På den anden halvdel af skolerne er det lærerne selv, der står for denne del.

På de skoler, hvor man har ændret bookingprocedure fra ledelse til lærerne selv, er det ofte sket ud fra den begrundelse, at det er "spildte" ledelsesressourcer. Lærerne er tilfredse med denne løsning. Dette skyldes ikke mindst, at bookingsystemet i dag fungerer langt bedre end tidligere.

Dette bekræftes i spørgeskemaundersøgelsen, hvor 67 pct. tilkendegiver, at de er helt eller overvejende enige i, at bookingsystemet fungerer godt.

Enkelte peger på, at der kan opstå et behov for koordinering lærerne imellem, så man undgår, at eleverne har flere test samme dag. Det er noget af det, som eleverne selv pæger. De oplever det som u hensigtsmæssigt at have op til tre test i træk, da det kan være svært at koncentrere sig.

Enkelte lærere pæger, at de oplever systemet som ufleksibelt, da man ikke kan ændre i bookingen til en elev, hvis det efterfølgende viser sig, at eleven fx skal på ferie, til læge eller lign. Dette beror dog på en misforståelse, da det fremgår af UNI-C's vejledning, at lærere med bookingrettighed kan redigere en booking frem til det tidspunkt, hvor testen starter, eller lærers side til testafvikling åbnes⁹³.

Andre skoler ser det som en fordel at have samme person til at booke test til alle klasser. På én skole er det eksempelvis den pædagogiske leder, der står for al booking. Hendes ansvarsområde er løbende evaluering, herunder de nationale test. Lærerne på skolen udtrykker stor tilfredshed med, at ansvaret er placeret centralt. Ud over at der er kontinuitet i perioden for afholdelsen af test fra år til år, fremhæver lærerne, at det er en fordel, at den, der booker test, har et samlet overblik over ledige computerressourcer på skolen.

På spørgsmålet om, hvorvidt lærerne er enige i, at det er hensigtsmæssigt, at de nationale test skal gennemføres inden for en centralt fastsat periode, svarer 16 pct. af lærerne, at de er helt eller overvejende uenige i dette udsagn. Ca. 50 pct. er helt eller overvejende enige. Det generelle indtryk er således, at de fleste lærere ikke ser noget problem med en på forhånd fastlagt testperiode. Af den del af lærerne, der er uenige, angiver 75 pct. som begrundelse, at testen kan risikere at komme på et tidspunkt, hvor klassen mangler at blive undervist i ét eller flere af de trinmål/områder, der testes. Dette er i overensstemmelse med de tilbagemeldinger, evaluator har fået af lærerne i naturfag, jf. kapitel 4.

Af mere teknisk karakter peger flere lærere fortsat på, at det er en udfordring at fremskaffe nok computere, der er funktionsdygtige, sådan at en hel klasse kan afvikle testen samtidig. Det er heller ikke altid, at der er ledige computere til rådighed i det tidsrum, man har booket, hvilket selvsagt opleves som frustrerende.

Under testen

Selve testafviklingen

I spørgeskemaundersøgelsen svarer 91 pct. af lærerne, som nævnt, at de bruger 1-3 timer på selve gennemførelsen af testen. Det generelle billede præciseres af casestudierne. Evaluators observationer af testafvikling samt tilbagemeldinger i interviewene giver grundlag for at konstatere, at gennemførelse af en test i de fleste tilfælde varer ca. 60 minutter. Ministeriet har i forbindelse med indførelsen af testen meldt ud, at et testforløb forventes at tage ca. 45 minutter⁹⁴. At evaluator har set flere forløb på ca. 60 minutter skyldes bl.a. elever, der kommer senere i gang end resten af klassen, samt elever, hvor det er nødvendigt at forlænge deres testforløb med ca. 15 minutter.

I parentes bemærket oplever flere lærere, at en testafvikling på mellem 45-60 minutter kan være en udfordring for nogle elever. Det gælder ikke mindst de yngre elever i indskolingen, for hvem det kan være svært at bevare koncentrationen.

De 60 minutter afspejler selve testafviklingen foran computeren. Dertil kommer den tid, der må gå med at få eleverne over i test-/it-lokalet samt retur. De klasser, der afvikler testen i deres eget klasselokale på fx en tablet, sparer selvsagt tid på denne del. Tidsforbruget afhænger derudover af, hvor meget tid læreren bruger på at introducere/genopfriske diverse praktiske som ikke-praktiske informationer/oplysninger.

⁹³ Brugervejledning. Testsystemet – De nationale test, UNI-C, 2013.

⁹⁴ Fremme af evalueringskulturen. Evaluering og test, Jakob Wandall, 2006. PowerPoint-show til Uddannelsesforum i Odense den 3. oktober 2006.

I forlængelse af ovenstående skal det nævnes, at de på forhånd definerede tider til testafvikling ofte ikke passer til skolernes ringetider. En test vil derfor ofte løbe over to lektioner.

Flere lærere fortæller, at det kan være nødvendigt at indgå aftaler med kollegaer om, at eleverne kommer for sent til den efterfølgende undervisning. Det gælder ikke mindst i de testfag, hvor man sjældent har to lektioner i træk. Det er evaluators generelle indtryk, at tidsforbruget, herunder at flere test løber over to lektioner, ikke ses som en stor ulempe blandt lærere og skoleledere. Det tolker evaluator som en accept af, at det er det tidsforbrug, der skal til for at sikre en hensigtsmæssig testafvikling.

Hjælpemidler

I forbindelse med casestudierne har flere lærere uopfordret peget på, at det er uklart, hvilken type hjælpemidler der er tilladt under selve testafviklingen. De mener, at der er behov for tydeligere retningslinjer/større grad af ensretning, så resultatet af testen ikke bliver misvisende, fordi det er forskelligt fra skole til skole, hvad der gøres brug af.

Evaluator oplevede ikke hjælpemidler i brug under testafvikling med undtagelse af CD-ord. Det var dog forskelligt, i hvilken grad og på hvilken måde lærerne gik rundt og hjalp de elever videre, der var stoppet op. Nogle lærere var helt afvisende, andre forsøgte at få eleven til at reflektere endnu engang, bruge udelukkelsesmetoden eller blot klikke videre.

CD-ord er et program, der læser teksten på skærmen op under testafviklingen. Evaluator observerede flere elever, der brugte CD-ord helt uden problemer eller gene for andre elever.

Tekniske udfordringer

I forbindelse med indførelsen af de nationale test oplevede skolerne en række tekniske udfordringer. Evaluator har derfor spurgt lærerne i casestudierne, om de fortsat oplever problemer af teknisk karakter. Tilbage meldingen er, at langt de fleste udfordringer er løst, og det generelt ikke opleves som et problem. Det er således evaluators generelle vurdering, på tværs af datakilder, at de tekniske problemer i selve testsystemet er minimeret. Der er nu i højere grad tale om tekniske udfordringer af lokal karakter.

På enkelte skoler er der således stadig udfordringer med trådløs opkobling på grund af ringe kapacitet. En af caseskolerne har derfor set sig nødsaget til at koble hver enkelt bærbar computer op via kablede netværk i et særskilt testlokale for at sikre, at eleverne ikke bliver hængt af under prøven. Andre skoler vælger fortsat at gennemføre testene i computerlokaler med stationære maskiner.

Efter testen

Opfølgning på testen

Som det fremgår af tabel 8.1 ovenfor, er opfølgning på testresultater opdelt i analyse og fortolkning af resultat samt opfølgning på team-, elev-, forældre- og skoleledelsesniveau.

Størstedelen af lærerne tilkendegiver, at de bruger mellem 1-3 timer på analyse og fortolkning. I casestudierne giver lærerne udtryk for, at det er meget forskelligt fra fag til fag og klasse til klasse, hvor lang tid de bruger på dette. Tidsforbruget afhænger ikke mindst af, hvor mange elever der klarer sig under middel, da de kræver ekstra tid, fordi læreren er nødt til at gå tilbage og analysere, hvilke opgaver eleven fik fejl i. For videre analyse af lærernes anvendelse af testresultater, se kapitel 4.

Langt de fleste lærere understreger, at de nationale test overordnet set er mindre tidskrævende end de fleste andre test, alene fordi de rettes automatisk. Enkelte oplever dog, at også de nationale test kan være tidskrævende, hvis man ønsker at gå i dybden med analysen af elevernes resultater, herunder de enkelte opgavetyper.

Andre peger på, at der ligger en værdi i selv at rette de test, som eleverne gennemfører. Det giver mulighed for at få et billede af elevernes svarmønstre og ræsonnementer, samt et godt overblik over den enkelte elevs faglige styrker og svagheder. Det er dog evaluators vurdering, at der er tale om et klart mindretal, der foretrækker dette. Som det fremgår af kapitel 4, foretrækker de fleste lærere en test, der selv genererer et resultat.

I relation til opfølgning i form af kollegial dialog, elevsamtaler og skole-hjem-samtaler er det evaluators vurdering, at der ikke er særskilte tidsmæssige fordele eller ulemper, når man sammenholder nationale test med andre testtyper.

Mest iøjnefaldende er det, at kun ca. halvdelen af lærerne drøfter resultatet med skolens ledelse. Dette bekræfter, at lærerne har et andet syn på graden af opfølgning sammenholdt med skolelejerne.

Sammenfattende er det evaluators vurdering, at hverken det tidsmæssige eller det tekniske perspektiv fylder meget i ledernes og lærernes bevidsthed. Umiddelbart er langt de fleste lærere af den opfattelse, at booking, klargøring og gennemførelse af test ikke opleves som særskilt problematisk eller tidskrævende. I relation til tidsforbruget efter testens gennemførelse er den største tidsmæssige fordel sammenholdt med papirbaserede test, at de nationale test selv genererer et resultat, hvilket de fleste ser som positivt. For en nærmere analyse af omkostninger ved drift af testen henvises til afsnit 7.2.3.

8.2 Det adaptive princip ud fra et pædagogisk perspektiv

Dette afsnit vil analysere fordele og ulemper ved det adaptive princip ud fra et pædagogisk perspektiv. Det vil sige en analyse af lærernes opfattelse/vurdering af, hvordan en adaptiv test påvirker elevernes testafvikling, perspektiveret med konklusioner fra litteraturstudiet.

I spørgeskemaundersøgelsen har evaluator først spurgt ind til lærernes forståelse af det adaptive princip og efterfølgende deres vurdering heraf.

Det kan konstateres, at langt størstedelen af lærerne kender til og forstår, hvad en adaptiv test indebærer. Kun 5 pct. har i spørgeskemaundersøgelsen sat kryds ud for de udsagn, der ikke kendetegner det adaptive princip. Dette stemmer overens med indtrykket fra casestudierne. Lærerne har en klar forståelse af testens adaptive element, og de gør meget ud af at forklare eleverne om det inden testgennemførelsen.

Tabel 8.2: I hvilken grad er du enig eller uenig i følgende udsagn vedrørende det adaptive princip? Angivet i pct.

	Slet ikke	I mindre grad	I nogen grad	I høj grad	I meget høj grad	Ved ikke	I alt (N)
Det giver et mere præcist billede af elevernes faglige niveau end traditionelle papirbaserede/lineære test	11	19	35	19	9	7	100 (2162)
Det bidrager ikke med ny viden om elevernes faglige niveau	12	35	30	8	6	8	100 (2162)
Det påvirker elevernes engagement positivt, at de løbende får udfordrende spørgsmål	7	22	43	17	5	7	100 (2162)
Det påvirker eleverne engagement negativt, at de løbende får fagligt udfordrende spørgsmål	16	36	30	5	4	9	100 (2162)
Det gør det svært at sammenligne elevernes testresultater, at de får forskellige spørgsmål	29	35	20	5	3	8	100 (2162)
Det gør det svært at sammenligne den enkelte elevs testresultater over tid	25	34	19	5	3	15	100 (2162)

Kilde: Spørgeskemaundersøgelse til lærere, 2013.

Det fremgår af tabellen ovenfor, at 43 pct. af lærerne i nogen grad og 22 pct. i høj eller meget høj grad vurderer, at det adaptive princip påvirker elevernes engagement *positivt*, fordi de løbende får fagligt udfordrende spørgsmål. I den modsatte ende af skalaen svarer 30 pct., at det i nogen grad påvirker eleverne *negativt*, at de får fagligt udfordrende spørgsmål. Der er altså umiddelbart tale om tilbagemeldinger fra lærerne, der peger i begge retninger.

Casestudierne understøtter spørgeskemaundersøgelsens billede af, at lærerne vurderer det adaptive princip som havende både styrker og svagheder. Generelt er vurderingen fra de interviewede lærere på tværs af skolerne dog, at styrkerne vejer tungest. De fremhæver først og fremmest, at både de stærke og svage elever udfordres på et niveau, der afspejler deres faglige kunnen. Med en adaptiv test får eleverne mulighed for at få en god testoplevelse, fordi alle undervejs får spørgsmål af en sværhedsgrad, som de kan svare på. Det antages at fremme motivationen og engagementet

Denne konstatering understøttes af litteraturstudiet, der peger på den fordel, at den adaptive test kan kreeres, så alle elever oplever, at ca. 50 pct. af de stillede opgaver løses korrekt. Dette er også tilfældet i de nationale test. Denne tilgang er motiverende for både relativt fagligt svage elever, fordi de ikke oplever en stor faglig utilstrækkelighed, og for relativt fagligt stærkere elever, fordi de oplever at blive udfordret og ikke keder sig pga. for mange for lette opgaver.

Litteraturstudiet peger desuden på, at de adaptive test også er gode til at registrere faglig progression blandt eleverne, særligt for elever med relativt store faglige spring mellem to test. Adaptive test konstaterer hurtigt en udvikling og tilpasser opgaverne herefter.

Det skal i denne sammenhæng anføres, at der fra et fagligt pædagogisk synspunkt kan argumenteres for, at anvendelsen af de nationale test, som et værktøj til måling af den enkelte elevs faglige progression, forudsætter, at testen tages med mindre tidsintervaller, end tilfældet er i dag.

Dette synspunkt understøttes af evaluators casestudier, hvor flere lærere peger på hyppigheden og antallet af test, som én af styrkerne ved de andre test på markedet, som de gør brug af og retrækker, jf. kapitel 4.

Lærerne i casestudiet peger derudover på, at eleverne ikke kan kigge efter hinanden under selve testafviklingen. Tilsvarende kan de heller ikke sammenligne resultater efterfølgende, da de ikke har haft de samme spørgsmål. Noget, der i forbindelse med andre test, kan fylde uhensigtsmæssigt meget. Enkelte lærere oplever, at nationale test tages mere alvorligt blandt eleverne. Samtidig opleves det som ekstra motiverende, at testen er it-baseret.

De ulemper, som lærerne nævner, går igen på tværs af skolerne. Et eksempel er, at på trods af, at de fagligt svage elever *ikke* nødvendigvis mister modet undervejs, fordi de kan svare på en del af de stillede spørgsmål, peger flere lærere på, at nogle efterfølgende bliver skuffede, når de får resultatet: "*Men bagefter bliver de [eleverne] alligevel dunket i hovedet med et dårligt resultat. Det kan være svært at forstå, hvorfor de alligevel ligger under middel.*"

For de fagligt stærke elever kan udfordrende spørgsmål på et højt fagligt niveau resultere i, at de bliver ærgerlige og kede af det, da de fra andre testforløb er vant til at kunne svare på alt. Flere lærere fortæller om elever, der er brudt grædende sammen. Det vurderes at være aftagende, efterhånden som eleverne får øget kendskab til den adaptive test og dens opbygning. Det sker på trods af, at lærerne har forklaret om det adaptive princip, herunder at alle elever vil opleve at nå til et fagligt niveau, hvor de ikke kan være med.

At eleverne bliver færdige på forskellige tidspunkter⁹⁵ opleves også som en uhensigtsmæssighed af flere lærere. Det er forskellig praksis fra klasse til klasse, hvorvidt hver enkelt elev må forlade lokalet, når de har gennemført testen, eller de skal vente til, alle er færdige. I de situationer, hvor enkelte elever sidder tilbage, fortæller lærerne, at mange elever bliver frustrerede, fordi de skal svare på så mange spørgsmål. Det kan resultere i, at de sjusker med svarene for hurtigere at blive færdige. I starten var mange af den opfattelse, at det var de fagligt svage, der sad tilbage. Men med tiden har de fleste elever fundet ud af, at det lige så ofte er de fagligt stærke elever. På trods af dette, er oplevelsen af, at det kan være ubehageligt at være den sidste, fortsat aktuel. Dette bekræftes af eleverne, jf. kapitel 5. I forlængelse heraf fortæller lærerne, at de adaptive test kræver et stort forklaringsarbejde over for eleverne, ikke mindst i indskolingen, hvor eleverne ikke har nogen tidligere erfaring. Det er noget, der lægges stor vægt på og prioriteres, som det fremgik ovenfor.

En anden ulempe, som flere lærere oplever, er, at fagligt stærke elever *læses* på et for lavt niveau i det adaptive testforløb. De første tre spørgsmål i elevens test er afgørende for niveauet af de efterfølgende spørgsmål. Lærerne fortæller, at hvis eleven fx er ukoncentreret, sjusker eller af anden grund kommer lidt skævt i gang, så har de set flere eksempler på fagligt stærke elever, der ikke *kommer op* og får testspørgsmål, der afspejler deres reelle faglige niveau senere i testen.

Afslutningsvist skal der på baggrund af litteraturstudiet peges på den ulempe, at eleverne ikke ved det adaptive princip oplever en indre sammenhæng mellem opgaverne, idet testsystemet blander opgaver fra opgavebanken på tværs af de tre profilmråder⁹⁶. Eleverne oplever altså ikke, at de løser et antal opgaver, der ligger i naturlig forlængelse af samme tema, som det er tilfældet for nogle lineære test, eksempelvis folkeskolens afgangsprøve, hvor samme hovedopgave kan indeholde et antal afledte spørgsmål, der skal besvares.

Typen af viden

Det er en antagelse, at den adaptive test kan give et mere præcist billede af elevernes faglige niveau.

I forlængelse af analysen af den pædagogiske anvendelse i kapitel 4 kan det i denne sammenhæng konstateres, at det kun er godt en fjerdedel af lærerne, der vurderer, at det adaptive princip i meget - eller i temmelig høj grad giver et mere præcist billede af elevernes faglige niveau sammenholdt med papirbaserede test. 35 pct. vurderer, at det i nogen grad er tilfældet.

⁹⁵ Antallet af testspørgsmål varierer fra elev til elev. Det er forskelligt, hvor mange spørgsmål den enkelte elev skal igennem, før testsystemet viser grønt (gennemført).

⁹⁶ Defineret i kapitel 7.

Flertallet af lærere oplever altså ikke i særlig høj grad, at testens mulighed for at afspejle elevernes individuelle faglige niveau bidrager til et mere præcist billede. I forlængelse heraf kan en hypotese være, at jo mere præcis en vurdering er/opleves, i jo højere grad vil den efterfølgende kunne understøtte en tilrettelæggelse af en undervisning, der tager udgangspunkt i klassens og den enkelte elevs faglige niveau.

At det adaptive princip ikke i særlig høj grad giver et mere præcist fagligt billede, kan for nogles vedkommende hænge sammen med den tilbagemelding/resultatvisning, som lærerne får i selve testsystemet. Dette element analyseres i næste afsnit.

I forlængelse heraf viser litteraturstudiet, at adaptive test ikke skaber ny viden om elevernes præstationer, som ikke kan etableres via lineære test. Ny viden om elevernes faglige niveau via de nationale test skabes således kun i de tilfælde, hvor lærere ikke tidligere ville have gennemført nogen form for test, og hvor lærerne heller ikke ville have påbegyndt brug af andre test i perioden siden 2010 i det tilfælde, at nationale test ikke havde eksisteret.

Sammenfattende kan det konstateres, at lærerne generelt er tilfredse med det adaptive princip. Størstedelen ser det i nogen eller i høj grad som en styrke, at eleverne testes/udfordres på deres individuelle faglige niveau, og det er en styrke, der vægtes højere end de ulemper, der peges på. Dette understøttes i høj grad i casestudierne. Det er dog evaluators vurdering, at dele af de konkrete ulemper/udfordringer kan løses ved tydelig kommunikation, vejledning og forklaringer til eleverne – noget, som de fleste af lærerne i casestudierne ser ud til at praktisere.

8.3 Tilbagemelding til lærerne

Det fremgår af kapitel 4, at det er et klart politisk mål, at de nationale test skal understøtte en differentieret planlægning af undervisningen. Det kan både være i relation til klassen, grupper af elever (holddannelse) og ikke mindst den enkelte elev⁹⁷.

Ministeriets egen pjece giver inspiration til pædagogisk brug af resultater fra de nationale test⁹⁸. Pjecen giver anvisninger til, hvordan man bruger den resultatvisning, som testsystemet stiller til rådighed, som fx klassens resultater i et søjlediagram, klassens resultat i tabelform samt detaljerede resultater fra den enkelte elev.

Dette afsnit analyserer denne tilbagemelding til lærerne i form af forskellige resultatvisninger, herunder i hvilket omfang de bidrager til, at læreren får et nuanceret billede af elevernes kundskaber og færdigheder.

⁹⁷ Jf. Folkeskolelovens § 18, stk. 2, der fastslår kravet om undervisningsdifferentiering.

⁹⁸ Brug testresultaterne – inspiration til pædagogisk brug af resultater fra de nationale test, Skolestyrelsen, 2011.

Tabel 8.3: Angiv, i hvilken grad du er enig eller uenig i følgende udsagn om visningen af resultaterne af de nationale test. Her menes de oversigter over resultater, som du ser i testsystemet på www.testogprøver.dk, fx søjlediagrammer, tabeller over klassens resultat og den enkelte elevs resultat og testforløb. Angivet i pct.

	Helt uenig	Overvejende uenig	Hverken enig eller uenig	Overvejende enig	Helt enig	I alt (N)
Er generelt brugervenlig	2	8	20	57	13	100 (2162)
Er for detaljeret	20	34	37	8	1	100 (2162)
Giver et godt overblik over klassens faglige niveau i de dele af faget, der testes	3	9	18	57	13	100 (2162)
Giver et klart billede af den enkelte elevs faglige niveau i de dele af faget, der testes	4	15	27	45	10	100 (2162)
Giver et godt grundlag for den faglige opfølgning på resultatet i relation til klassen	4	15	33	42	7	100 (2162)
Giver et godt grundlag for den faglige opfølgning på resultatet i relation til den enkelte elev	4	15	30	43	7	100 (2162)
Styrker generelt min mulighed for at differentiere min undervisning	10	19	37	30	5	100 (2162)

Kilde: Spørgeskemaundersøgelse til lærere, 2013.

Evaluators har i spørgeskemaundersøgelsen til lærerne indledningsvis spurgt til resultatvisningens generelle brugervenlighed. Hele 70 pct. er helt enige eller overvejende enige i, at den generelt er brugervenlig. 20 pct. er hverken enige eller uenige.

Ligeledes er det langt størstedelen, der er helt eller overvejende enige i, at præsentationen af resultaterne i testsystemet giver et godt overblik over *klassens* faglige niveau i de dele af faget, der testes. Denne andel falder en smule, når der bliver spurgt til, hvorvidt resultatvisningen giver et klart billede af *den enkelte elevs* faglige niveau. Det er 55 pct. helt eller overvejende enige i, hvor 26 pct. er hverken enige eller uenige.

Når der spørges ind til lærernes holdning til, hvorvidt resultatvisningen understøtter den faglige *opfølgning* på henholdsvis klassen og i relation til den enkelte elev, er ca. 50 pct. helt eller overvejende enige i, at det er tilfældet. Ca. 20 pct. er helt eller overvejende uenige. Tallet er noget lavere, når der direkte spørges ind til, hvorvidt resultatvisningen styrker lærernes mulighed for at undervisningsdifferentiere. Ca. 35 pct. er helt eller overvejende enig i, at det er tilfældet, hvor ca. 30 pct. er helt eller overvejende uenige. Den resterende del er hverken enige eller uenige.

Dette indikerer, at selvom resultatvisningen umiddelbart er brugervenlig og giver et godt overblik for en stor andel af lærerne, så bliver det mere udfordrende, når resultatet skal omsættes til en differentieret planlægning og gennemførelse af undervisningen. Det stemmer overens med det billede, der tegner sig på baggrund af casestudierne i relation til den pædagogiske anvendelse, jf. kapitel 4.

Evaluators har sammenholdt lærernes holdning til resultatvisningen med lærernes brug af testresultaterne. Resultatet er, at der generelt er en stærk positiv sammenhæng mellem lærernes holdning til visningen af resultaterne af de nationale test og lærernes anvendelse af resultatet af de nationale test. Med andre ord er der en klar tendens til, at de lærere, der har en positiv hold-

ning til testen, i højere grad gør brug af resultaterne i deres pædagogiske arbejde, herunder den fremadrettede planlægning – et resultat, der understreger potentialet i at være i dialog med lærerne om testens formål og muligheder ud fra den betragtning, at hvis flere lærere begynder at vurdere testen som meningsfuld, vil flere sandsynligvis også anvende den.

Ser man på de forskellige resultatvisninger, er der ikke den store forskel på, hvor stor en andel af lærerne, der bruger henholdsvis søjlediagram, klassens testresultat i tabelform samt testresultater for den enkelte elev. Det er mellem 40 og 45 pct. af lærerne, der angiver, at de bruger alle tre resultatvisninger altid eller ofte, og ca. 35 pct. bruger det engang imellem.

I resultatvisningen er det også muligt at se oplysninger om hver enkelt testopgave, som eleverne har svaret på. Den mulighed er der kun en fjerdedel, der altid eller ofte benytter sig af. 40 pct. svarer, at de gør det engang imellem.

Generelt bekræfter spørgeskemaundersøgelsen billedet af, at langt fra alle lærere bruger resultatvisningens muligheder efter hver test. Ca. 40 pct. af lærerne angiver, at de er helt eller overvejende enige i, at de primært bruger de forskellige oversigter, når resultatet for en elev eller klasse ikke er som forventet. 30 pct. er hverken enige eller uenige. Dette resultat understøtter de tilbagemeldinger, som evaluator har fået i casestudierne.

Casestudierne

I casestudierne har evaluator spurgt ind til lærernes anvendelse af resultatvisningen. Generelt er tilbagemeldingen, at langt de fleste har let ved at navigere rundt og kender til de forskellige typer skærbilleder. Dette understøtter tilbagemeldingerne i spørgeskemaundersøgelsen i tabellen ovenfor.

Det er evaluators vurdering, at udfordringen opstår i den mere detaljerede analyse, altså i arbejdet med at oversætte resultatet til pædagogisk praksis, sådan som det også fremgik af kapitel 4.

I interviewene peger flere lærere først og fremmest på, at det er besværligt at få et samlet billede af den enkelte elevs faglige styrker eller svagheder, der er tilstrækkelig detaljeret. I dag giver systemet mulighed for at få sorteret elevernes resultater inden for hvert profilområde, men det er, efter lærernes vurdering, for overordnet. Profilområdet Tal og Algebra rummer mange faglige områder/opgaveemner (fx talkendskab, 10-talssystem, addition/subtraktion, multiplikation/division m.fl.), og en elev vil ikke nødvendigvis have udfordringer inden for alle. Hvis man som lærer ønsker sig et mere detaljeret overblik, er man nødsaget til at gå ind og åbne hver enkelt opgave manuelt.

Nogle af de interviewede lærere udtrykker derfor ønske om, at systemet automatisk genererer en liste over, hvordan eleven/klassen har klaret specifikke faglige områder, sådan at man kan tilrettelægge en undervisning ud fra den enkeltes behov. En lærer fortæller, at hun printer de samlede resultater ud, for efterfølgende med en farveblyant at kunne gruppere resultaterne for hver elev inden for de enkelte faglige emner.

Flere foreslår desuden, at resultatvisningen i testsystemet skal indeholde muligheden for at sammenligne den enkelte elevs resultater over tid, herunder både frivillige og obligatorisk test.

Også litteraturstudiet peger på en ulempe, der er relateret til resultatvisningen. Det kan være en udfordring at forstå resultatet af en adaptiv test. Elevdygtigheden beregnes på en intervalskala. Ved tilbagemelding til læreren vises resultatet på percentilskalaen (1-100). Denne percentilværdi kan efterfølgende indplaceres i én af de fem grupper, der er angivet ved 5-trins-skalaen. Den underliggende intervalskala har den egenskab, at afstanden mellem to punkter er lige stor, mens denne egenskab bortfalder ved transformation til percentilskalaen. Netop denne kobling mellem percentilskala og de fem niveauer opleves af flere lærere og elever som problematisk. Ikke mindst blandt eleverne er der et ønske om, at de får at vide, hvor på percentilskalaen de ligger. De oplever, at intervallerne, der udgør de fem niveauer, er så brede, at det er væsentligt at vide,

om man ligger i bunden eller toppen af niveauet, hvis man skal have et præcist billede af sit faglige niveau⁹⁹.

Af mindre ulemper peger enkelte lærere på, at der ikke er plads til hele klassens resultat på ét skærbillede, når man vil se tabellen for dette. Derfor bliver man nødt til at scrolle frem og tilbage. Andre ærgrer sig over, at skærbilledet ikke er tilpasset smartphones.

Samlet set er det dog først og fremmest den manglende automatisk genererede detaljeringsgrad i forhold til de enkelte elever, som de interviewede lærere peger på.

⁹⁹ Evaluator er opmærksom på, at eleverne som udgangspunkt alene skal have oplyst deres placering på ét af de fem niveauer. Evaluator har dog mødt flere elever, der har kendskab til 100-skalaen, eller som får deres resultat i form af tallet på percentilskalaen.

Boks 8.1: Ekspertvurdering i relation til resultatvisning

Ekspertvurdering

Som en del af den gennemførte ekspertvurdering (se kap. 5) har evaluatoren ligeledes bedt de faglige eksperter vurdere testsystemets resultatvisning.

Tre ud af de fire ekspertgrupper er enige om, at resultatvisningen indeholder nogle brugbare elementer og muligheder. Matematikgruppen er mindre enig i denne betragtning.

Størstedelen af de faglige eksperter mener, at muligheden for at se klassens resultat i et søjlediagram, klassens resultat i tabelform samt detaljerede resultater fra den enkelte elev er brugbar, men det kræver en indsats fra lærerens side. Det vurderes derfor, at der ligger en stor udfordring i at understøtte lærerne i deres arbejde med at omsætte resultaterne til praksis i undervisningen. Fagekasperterne i matematik vurderer, at der fremadrettet bør tænkes i en helt anden og mere præcis tekst, hvis lærerne skal kunne bruge resultatvisningen i praksis, som det uddybes nedenfor.

Alle ekspertgrupper er enige om, at det er relevant at få sat fokus på, hvordan lærernes viden om den pædagogiske anvendelse kan optimeres.

Ekspertgruppen i engelsk er umiddelbart mest positiv. De mener, at resultatvisningen er et godt redskab for lærerne, der principielt giver mulighed for at gå kirurgisk til værks i arbejdet med at analysere elevernes faglige styrker og svagheder, som de formulerer det.¹ Ligesom engelsklærerne i casestudierne henvises der til, at der ikke tidligere har været test tilgængelige, så de nationale test opleves generelt som et løft.

Ekspertgruppen i dansk læsning peger på, at en vigtig funktion i det nuværende system er muligheden for at sortere resultaterne efter profilområder, der kan give et mere samlet overblik og gøre det nemmere at bruge resultatet som udgangspunkt for fx holddannelse. Selvom netop disse funktioner beskrives i Kvalitets- og Tilsynsstyrelsens vejledning, er det indtrykket, at langt fra alle lærere gør brug af denne mulighed.

Ekspertgrupperne i henholdsvis matematik og fysik/kemi kommer med forslag til forskellige funktionaliteter, der som et supplement til resultatvisningen, vil kunne hjælpe lærerne på vej.

Ekspertgruppen i matematik vurderer, at det vil understøtte lærernes anvendelse af resultaterne, hvis testsystemet automatisk udpeger elever, man som udgangspunkt bør være opmærksom på, herunder hvilke specifikke faglige udfordringer/styrker der er tale om.

I dag er det alene muligt at sortere elevernes svar inden for de tre profilområder, men det vurderes af fagekasperterne, ligesom lærerne (se ovenfor), at være for generelt.

I forlængelse heraf peger ekspertgruppen i fysik/kemi på, at de tre profilområder inden for faget ikke er skarpt afgrænset, hvorfor der er opgaveemner (fx syre-base, pH-værdi), der går igen inden for alle tre områder. Gruppen er derfor enig i, at det vil være hensigtsmæssigt med en funktion i resultatvisning, der kan sortere besvarelser på opgaveemneniveau både i forhold til klassen samlet set og i forhold til den enkelte elev.

Afslutningsvis skal det nævnes, at flere af eksperterne på tværs af de fire grupper er enige om, at den samlede vurdering af klassens resultater på tværs af de tre profilområder (resultatvisningen med søjlediagrammer) er mindre hensigtsmæssig. Et gennemsnit af resultatet inden for hvert af de tre profilområder giver ingen mening i relation til den fremadrettede anvendelse. Det er nødvendigt at se på, hvordan klassen klarer sig inden for det enkelte profilområde, for at kunne målrette undervisningen.

8.4 Delkonklusion

Dette kapitel har belyst en række evalueringsspørgsmål under evalueringstemaet: *Testens udformning*. Kapitlet har fokuseret på en analyse af forskellige elementer i relation til den praktiske tilrettelæggelse af testafviklingen på skolerne, visning af resultaterne af testene samt en vurdering af det adaptive princip set ud fra et undervisningsmæssigt/pædagogisk perspektiv.

Samlet set er der tale om en analyse af rammerne for testafviklingen ud fra to forskellige perspektiver. Det ene perspektiv har fokuseret på selve tilrettelæggelsen af testen, herunder *planlægning* og *gennemførelse*, samt fordele og ulemper ved det adaptive princip i testsituationen set ud fra et lærerperspektiv. Det andet perspektiv har fokuseret på *opfølgningen* på testen, dvs. en vurdering af, i hvilken grad hypotesen om, at systemets forskellige resultatvisninger bidrager til, at testen giver læreren et nuanceret billede af elevernes kundskaber og færdigheder, kan bekræftes. Kapitlet kommer således hele vejen rundt om de faser af lærerens undervisning, hvor testene er i anvendelse: planlægning, gennemførelse og opfølgning.

I relation til tilrettelæggelsen af testen på skolerne, såsom booking, klargøring af it samt forberedelse af eleverne mv. kan det overordnet konkluderes, at testafviklingen ud fra et lærer- og ledersynspunkt ser ud til at kunne planlægges og gennemføres på tilfredsstillende vis. De tekniske udfordringer, der har været tidligere, er stort set løst, og de fleste vurderer, at bookingsystemer mv. fungerer godt. Enkelte skoler/kommuner har dog fortsat problemer med manglende hastighed og elever, der bliver *smidt af* midt i en test, men det er en generel problematik, uafhængig af testsystemet.

Hvad angår tidsforbrug, der går til testafviklingen, er dette ikke noget, som hverken lærere eller ledere peger på som en ulempe. Der er dog ingen tvivl om, at der er en klar tidsmæssig gevinst ved at gennemføre it-baserede test, når man sammenholder med papirbaserede test. Det skyldes, at de nationale test automatisk genererer et resultat, hvorfor lærerne sparer rettetid. Dette er dog en funktionalitet, som andre it-baserede test på markedet også indeholder, som det fremgik af kapitel 4.

De nationale test adskiller sig særskilt fra andre test i folkeskolen ved at gøre brug af det adaptive princip, hvor testen løbende tilpasser sig elevens niveau. I forhold til dette princip er konklusionen, at størstedelen af lærerne ser dette som en fordel for eleverne og deres testafvikling. Dette understøttes af litteraturstudiet. Det er dog muligt at pege på en række ulemper, relateret til elevernes oplevelse af testsituationen (fx at eleverne bliver færdige på forskellige tidspunkter, og nogle derfor risikerer at sidde tilbage til sidst), men den styrke, der ligger i, at eleverne hver især udfordres på deres eget faglige niveau, vejer tungest i lærernes vurdering.

Det kan tillige konkluderes, at det kun er godt en fjerdedel af lærerne, der vurderer, at det adaptive princip i meget - eller temmelig høj grad giver et mere præcist billede af elevernes faglige niveau sammenholdt med papirbaserede test.

Som en del af opfølgningen på testen skal den enkelte lærer ind i testsystemets resultatvisning og orientere sig om klassens og de enkelte elevers resultater. Som nævnt ovenfor har det været evaluators hypotese, at denne tilbagemelding til lærerne bidrager til, at lærerne får et nuanceret billede af elevernes kundskaber og færdigheder.

Svaret på hypotesen er ikke entydigt. En betydelig del af lærerne angiver i spørgeundersøgelsen, at resultatvisningen umiddelbart er brugervenlig. De kender systemet, og de kan navigere rundt i de forskellige typer skærbilleder og detaljeringsgrader. Det er dog langt fra alle lærere, der gør brug af denne mulighed, først og fremmest fordi de i højere grad vælger andre evalueringsværktøjer, jf. kapitel 4.

En del af begrundelsen skyldes dog selve resultatvisningen. Her er meldingen fra flere sider, at resultaterne kunne vises på måder, der i endnu højere grad understøtter lærernes arbejde med at implementere resultaterne. Både interviewede lærere på casestudierne og faglige eksperter peger på konkrete funktioner i resultatvisningen, som de mener, vil hjælpe lærerne til at få et endnu mere nuanceret, men også systematiseret billede af klassens faglige niveau.

Ikke mindst er der et ønske om at få et overblik over de konkrete opgaver, eleven har gennemført, uden at skulle åbne hver enkelt testopgave i visningen manuelt. Samtidig er der et ønske om automatisk gruppering af elever, der har specifikke faglige udfordringer. I dag er det muligt at gruppere på profilområdeniveau, men det vurderes af flere lærere til at være for generelt til at kunne planlægge undervisningen efter.

Konklusionen er, at nok er der tale om en resultatvisning, der umiddelbart kan bidrage til et nuanceret billede af kundskaber og færdigheder og tilsvarende er forholdsvis let at navigere rundt i på skærmen, men det kræver fortsat et relativt stort fortolknings- og analysearbejde af den enkelte lærer at skulle omsætte resultatet til praksis. Nye funktionaliteter i visningen og udvikling af analyseværktøjer kan være en del af løsningen. Men som det fremgik af kapitel 4, er der en lang række andre elementer, der spiller ind i forhold til den fremadrettede pædagogiske anvendelse af resultaterne.

9. KONKLUSIONER

I dette afsluttende kapitel sammenfatter evaluatoren de væsentligste **hovedkonklusioner** fra de foregående kapitler. Kapitlet er struktureret i henhold til evalueringens fire temaer samt en tværgående analyse.

Som supplement til hovedkonklusionerne indeholder kapitlet en **perspektivering** med en række opmærksomhedspunkter, der af evaluatoren vurderes relevante at overveje i den videre udvikling af de nationale test.

9.1 Tematiske hovedkonklusioner

Nedenfor sammenfattes hovedkonklusionerne for de fire temaer.

9.1.1 Testenes effekt

Med henblik på at afdække sammenhænge mellem de nationale test, folkeskolens evalueringsskulturer og elevernes faglige niveau har evaluatoren foretaget en række analyser af *effekterne af indførelsen af de nationale test*. Analysernes overordnede resultat er, at de nationale test har haft en positiv indvirkning på elevernes faglige resultater, samt at de har været medvirkende til at styrke skolernes evalueringsskulturer. På baggrund af de tilgængelige data kan der dog ikke på nuværende tidspunkt findes en tydelig sammenhæng mellem skolernes evalueringsskulturer og elevernes faglige resultater.

Analysen af **sammenhængen mellem de nationale test og elevernes faglige resultater** indikerer en positiv effekt af de nationale test i relation til elevernes faglige resultater. Både analyserne på elev-, skole- og landsniveau indikerer således, at de nationale test har en positiv effekt på elevernes faglige niveau. Det har dog kun været muligt at undersøge de generelle effekter af de nationale test – dvs. *om* testene har en effekt på elevernes faglige niveau og ikke, *hvordan* de har en effekt. Resultaterne tyder dog på, at den konstaterede effekt ikke alene skyldes *teaching to the test*, eller at eleverne blot er blevet bedre til at gennemføre testene, men at eleverne faktisk er blevet dygtigere til det fag, de er blevet testet i. Det har dog alene været muligt at se på effekten af nedbrud i dansk, læsning. Når data for 2013 bliver tilgængelige, vil det være muligt at undersøge, om den samme effekt ses i matematik.

Analysen af **sammenhængen mellem de nationale test og skolernes evalueringsskulturer** tyder for det første på, at skolerne overordnet set har styrket deres evalueringsskulturer i de senere år. For det andet er der indikationer på, at de nationale test har medvirket til at styrke skolernes evalueringsskulturer. Styrkelsen af evalueringsskulturen falder til en vis grad tidsmæssigt sammen med indførelsen af de nationale test. Undersøger man de oplevede effekter af de nationale test, er skolelederne samt forvaltningerne forholdsvis positive i deres vurdering af, hvorvidt de nationale test har bidraget til at styrke skolernes evalueringsskulturer. Dog er lærerne mere skeptiske. Evaluatoren vurderer samlet, at de nationale test har haft en understøttende og dagsordenssættende effekt i relation til at understrege et fokus på faglige resultater såvel som bidrag til en større systematik i skolernes evaluering og har derved medvirket til en stærkere evalueringsskulturer. Det er dog samtidig evaluatorens vurdering (som det anføres i lovbemærkningerne), at det ikke er de nationale test alene, der skaber en stærkere evalueringsskulturer. Det er derimod helheden i anvendelsen af nationale test i samspil med en række evalueringsværktøjer samt en stærk ledelse, der understøtter en stærk evalueringsskulturer.

De statistiske analyser af **sammenhængen mellem skolernes evalueringsskulturer og elevernes faglige resultater** viser ikke en overbevisende sammenhæng. Selvom resultaterne ikke kan bekræfte, at der er en statistisk sammenhæng ud fra de tilgængelige data, er dette dog ikke et endegyldigt bevis på fravær af sammenhæng mellem skolernes evalueringsskulturer, og hvordan skolens elever klarer sig. Det er evaluatorens vurdering (ud fra casestudier og eksisterende undersøgelser), at en stærk evalueringsskulturer har en fremmende effekt for at opnå gode faglige resultater. Men en stærk evalueringsskulturer er ikke en tilstrækkelig betingelse for at opnå gode resultater. Evaluatoren vurderer, at der er potentiale for en stærkere sammenhæng, såfremt skolerne fremadrettet styrker deres evalueringsskulturer og benytter de nationale test mere systematisk.

Læs mere om *testenes effekt* i kapitel 3.

9.1.2 Testenes anvendelse

Evalueringsstemaet *testenes anvendelse* omfatter en analyse af en række aktørers anvendelse, oplevelse og vurdering af de nationale test. Det gælder lærere, elever, forældre (som led i skole-hjem-samarbejdet) samt skoleledere og kommuner. Nedenfor sammenfattes de væsentligste konklusioner i relation til anvendelse af de nationale test.

En central del af nærværende evaluering er analysen af, hvorvidt – og i givet fald hvordan – **lærerne anvender de nationale test** til den løbende evaluering af elevernes udbytte af undervisningen, herunder den fremadrettede planlægning af undervisningen. Evalueringen dokumenterer, at langt størstedelen af lærerne gør brug af de nationale test.

Denne overordnede konklusion nuanceres imidlertid ved, at lærerne i overvejende grad bruger testene til skabe overblik og dermed monitorere klassens og elevernes faglige niveau. Derimod er brugen af de nationale test mindre udbredt i forhold til den fremadrettede tilrettelæggelse af undervisningen. Evalueringen viser, at mange lærere (indtil videre) primært anvender de nationale testresultater *summativt* og i langt mindre grad *formativt*. Derudover er der markante forskelle i anvendelsen mellem fagene. Anvendelsen af testresultaterne er således mere udbredt i dansk, læsning, matematik og engelsk, end det er tilfældet i naturfagene. Endelig er det en hovedkonklusion, at testresultaterne typisk ses som et supplement til viden, indhentet fra øvrig evaluering. De nationale test står sjældent alene som redskab i en evalueringskultur.

Som nævnt er det evaluators overordnede indtryk, at det fortsat kun er en forholdsvis lille del af lærerne, der konsekvent og systematisk bruger testresultaterne fremadrettet som grundlag for planlægning af undervisningen. Den manglende formative anvendelse kan forklares ved, at mange lærere bruger andre evalueringsværktøjer. En anden forklaring kan være manglende fokus og rammesætning fra skoleledelsens side. Endelig spiller oplevelsen af manglende relevans ind for en del lærere. Særligt lærere inden for naturfag finder det ikke brugbart med én test relativt sent i folkeskolen (8. klasse), og derudover påvirker en række fagspecifikke forhold graden af test-anvendelse.

Evalueringen undersøger, i hvilket omfang de nationale test har indflydelse på lærernes prioritering af undervisningens indhold. Der refereres ofte til begrebet *teaching to the test*, men evaluator anvender i stedet begrebet *washback* som et mere nuanceret udtryk for, at man kan tale om både negativ og positiv påvirkning af/indflydelse på undervisningen. Evalueringen viser et særdeles differentieret billede af omfanget af washback i relation til brugen af de nationale test. Lærerne på de undersøgte skoler fordeler sig i forskellige grupper, hvor halvdelen af lærerne giver udtryk for, at de slet ikke eller i begrænset omfang målretter deres undervisning i forhold til obligatoriske og frivillige test, hvorimod den anden halvdel gør det i forskellig grad. Hovedkonklusionen i forhold til dette spørgsmål er, at evalueringen tegner et meget varieret billede af testenes påvirkning af undervisningen. En gruppe lærere målretter undervisningen, enten fordi de finder det naturligt (da testene afspejler Fælles Mål), eller fordi de gerne vil have, at eleverne klarer sig godt, da der er fokus på resultaterne i de nationale test. Den anden gruppe omfatter de lærere, som meget bevidst ikke vil lade deres undervisning styre af de nationale test. En vigtig konklusion er i øvrigt, at *teaching to the test* (eller *washback*) er mere udbredt dels i udskolingen, og dels på skoler, hvor ledelsen har fokus på faglige resultater.

Sammenfattende viser analysen af *den pædagogiske anvendelse af de nationale test*, at testene generelt anvendes i den danske folkeskole, men at de i dag primært anvendes summativt til at skabe overblik og til at bekræfte eksisterende viden. En række barrierer eksisterer – såvel generelle som fagspecifikke – i forhold til at styrke særligt den formative anvendelse af de nationale test. Ser man på tværs af fagene, kan evalueringen entydigt konkludere, at dansk, læsning er den test, hvor skolerne/lærerne er længst fremme i arbejdet med at bruge resultaterne fremadrettet. Dette skyldes ikke mindst testens særlige karakter, antallet af test hen over klassetrin, antallet af timer til rådighed i faget samt udviklingen af et supplerende analyseværktøj.

I forhold til **elevernes oplevelse af de nationale test** er evaluators samlede vurdering – baseret på elevernes tilbagemeldinger, lærernes svar samt evaluators observationer – at eleverne

overordnet set er positivt indstillede over for de nationale test. De it-baserede og adaptive test er generelt motiverende og udfordrende for eleverne, men der er også eksempler på, at testene skaber stress og usikkerhed hos visse elever. Forskellig praksis på skolerne vedrørende afslutning af testene, hvor nogle elever bliver færdige før andre og skaber en uhensigtsmæssig afvikling for nogle elever. En hovedkonklusion er endvidere, at eleverne hovedsageligt modtager kollektiv tilbagemelding fra lærerne om resultaterne af de nationale test, mens dialog med den enkelte elev om testresultater kun sker i begrænset omfang. Samtidig peger evalueringen på, at netop den individuelle dialog med eleven om resultater af testene er afgørende for testenes betydning for elevernes bevidsthed om egne styrker og svagheder samt deres motivation for at omsætte denne viden til handling.

Det er evaluators samlede vurdering, at de nationale test i nogen grad styrker **skole-hjem-samarbejdet**. Evalueringen viser, at forældrene generelt set orienteres om tidspunkt for gennemførelse af de nationale test, samt at yderligere information om formål med og praktisk gennemførelse af de nationale test mv. typisk sker især den/de først(e) gang(e), eleverne skal gennemføre en national test. Forældrene modtager typisk skriftlig information om den enkelte elevs resultater gennem forældrebreve, som enten sendes hjem til forældrene eller udleveres ved skole-hjem-samtaler. En hovedkonklusion er, at resultaterne af de nationale test kun i begrænset omfang anvendes til at opstille læringsmål for eleverne (i dialog med hjemmet), ligesom opfølgningen herpå kun i nogen grad skrives ind i elevplanerne. Der synes således at være potentiale for at inddrage resultater af nationale test til opstilling af læringsmål i højere grad, end tilfældet er i dag.

Evalueringen belyser anvendelsen af de nationale test som værktøj for henholdsvis **skolelederes og kommuners kvalitetsarbejde**.

Evalueringen viser, at **skolelederne** generelt gør brug af de nationale test, om end med betydelig variation. De orienterer sig i og analyserer testresultaterne *over tid* primært på skole- og klasseniveau. På de skoler, hvor skolelederne udtrykker høje faglige forventninger til eleverne, er der en tendens til, at skolelederne også orienterer sig i resultater på elevniveau. De nationale test indgår således til en vis grad som en del af skoleledernes kvalitetsarbejde. Evalueringen dokumenterer stor variation blandt skolerne i forhold til, dels om der er klare retningslinjer for brug af de nationale test, dels *hvordan* testresultater anvendes. Nogle skoleledere bruger de nationale test i meget begrænset omfang og fungerer primært som støttende for lærerne i forhold til det praktiske arbejde, men uden at indgå i dialog med lærerne om testresultaterne. Andre skoleledere bruger resultaterne til monitorering og overblik over det faglige niveau, mens en tredje gruppe af skoleledere *handler aktivt* og lægger testresultaterne til grund for tildeling af ressourcer, ændret fagfordeling, sparring med lærere mv. Med udgangspunkt i de eksempler, hvor skoleledelsen meget aktivt rammesætter og anvender de nationale test som en del af skoleledelsen (sammen med øvrig evaluering), vurderer evaluator, at der er et uudnyttet potentiale for de nationale test på en stor andel skoler, der med mere aktivt fokus på anvendelse af de nationale test vil kunne stimulere yderligere positive tiltag til at fremme elevernes faglige niveau. Endelig viser evalueringen, at der blandt de danske skoleledere generelt er en positiv holdning til de nationale test og testenes potentialer. Det er dog de færreste, der finder, at testene er et værktøj til dialog med de kommunale forvaltninger.

Evalueringen dokumenterer, at **kommunerne** generelt stiller krav til skolernes faglige resultater. Kommunerne vurderer i den sammenhæng, at resultaterne af de nationale test er en væsentlig indikator til at vurdere skolernes faglige resultater, men hovedkonklusionen af evalueringen er, at resultater fra de nationale test er "en del af paletten", som det udtrykkes af flere kommuner. Kommunerne interesserer sig således for skolernes faglige resultater (og vurderer også, at resultater er vigtigere end rammebetingelser og pædagogiske processer), men de nationale test er blot ét ud af flere værktøjer til at sammensætte et billede af en skoles faglige niveau. Evalueringen viser endvidere stor variation på tværs af kommuner i forhold til anvendelsen af de nationale test. Men det samlede billede er, at kommunerne overvejende bruger de nationale test som et monitoreringsredskab, der kan skabe et overblik over faglige resultater. Omvendt er testresultater sjældent omdrejningspunktet for kommunernes kvalitetsarbejde, hvor de inddrages i dialog med skolelederne med henblik på fremadrettet pædagogisk handling. Det tyder på, at kommunerne i høj grad uddelegerer ansvaret for at opnå resultater i de nationale test til skolerne, hvil-

ket tilsvarende gælder for opfølgningen herpå. Evalueringen viser, at der er et potentiale for at fremme evalueringskultur og positiv udvikling i elevernes faglige niveau ved en mere aktiv forvaltningsadfærd i dialog med skoleledere om, hvordan lederne fremmer evalueringskultur.

Læs mere om *testenes anvendelse* i kapitel 4, 5 og 6.

9.1.3 Testenes indhold

Evalueringsstemaet *testenes indhold* omfatter en analyse af en række forhold relateret til det faglige indhold af de nationale test, der anvendes i de danske folkeskoler. Dertil kommer en analyse (baseret på litteraturstudier) af fordele og ulemper ved det adaptive princip ud fra et teknisk og økonomisk perspektiv.

Generelt konkluderer evaluatoren, at **valg af testfag** samlet set synes hensigtsmæssigt. Dette er baseret på, at testfagene hver for sig vurderes relevante og testegnede. Imidlertid er der knyttet specifikke svagheder til flere af fagene.

Evalueringen konkluderer, at **valg af profilområder** generelt repræsenterer genkendelige og relevante kernelementer i fagene, og det er en selvstændig pointe, at profilområderne på tværs af fag i høj grad rummer testfagernes testbare områder. Heri ligger også en erkendelse af, at testene langt fra dækker testfagene i deres helhed, hvilket da heller ikke var intentionen. I forlængelse heraf viser evalueringen imidlertid, at der er en række fagspecifikke udfordringer i relation til de nuværende profilområder, herunder særligt i matematik og fysik/kemi samt i begrænset omfang i engelsk. Det samme gælder for testenes evne til via profilområderne at afspejle trinmålene fra Fælles Mål.

En væsentlig hovedkonklusion er endvidere, at der inden for flere testfag er en udfordring i forhold til mangel på svære opgaver i opgavebanken. En optælling af opgaver udført af evaluatoren viser, at der er ganske få opgaver inden for flere profilområder. Det indebærer, at fagligt stærke elever "tømmer" opgaver med højeste sværhedsgrad og/eller møder de samme opgaver, fx i obligatoriske og frivillige test. **Antallet af opgaver** i opgavebanken synes derfor at være en udfordring. Det skal dog understreges, at det ikke har været en del af evalueringens opdrag at foretage en samlet vurdering af størrelsen af den samlede opgavebank.

Endvidere påpeger evalueringen, at det generelt er en udfordring, at opgaveskabelonerne (som er rammen for udvikling af opgaver) ikke afspejler den teknologiske udvikling, der er sket siden 2007. Evalueringen konstaterer tilsvarende, at der er behov for en gennemgang af udvalgte testopgaver i forhold til at sikre hensigtsmæssig layout, grafik samt brug af illustrationer. Behovet for forbedringer er dog mindst udtalt i dansk, læsning.

Endelig viser evalueringen, at der er behov for at forbedre rammerne for opgaveudviklingen, bl.a. i form af systematisk uddannelse af opgaveudviklere samt stærkere tilknytning til forskning.

De nationale test er baseret på **det adaptive princip**, som betyder, at testen løbende tilpasses den enkelte elevs faglige niveau ved, at testen giver eleverne opgaver på baggrund af deres besvarelse af de tidligere opgaver i testen. Evalueringen viser, at de nationale test med det adaptive princip sikrer en hurtigere og mere præcis fastlæggelse af elevernes faglige niveau. De adaptive test giver en rapportering af elevernes niveau, der kan være vanskeligere at forstå end resultatrapportering fra lineære test. Evalueringen viser endvidere, at brug af adaptive test betyder, at udvikling af opgaver til opgavebanken er betydeligt mere ressource- og omkostningskrævende end lineære test. Det gælder i relation til ministeriet og UNI-C samt aflønning af eksperter. Det gælder tilsvarende på skoler, der medvirker i afprøvning af opgaver, der skal analyseres i forhold til, om de lever op til de kvalitetskrav, som det kræver at indgå i opgavebanken for de nationale test.

Læs mere om *testenes indhold* i kapitel 7.

9.1.4 Testenes udformning

Evalueringen har under temaet *testenes udformning* afdækket en række evalueringsspørgsmål vedrørende den praktiske tilrettelæggelse af de nationale test på skolerne, lærernes holdning til det adaptive princip samt tilbagemeldingen til lærerne efter afvikling af testene.

For så vidt angår **den praktiske tilrettelæggelse af de nationale test på skolerne**, er den overordnede konklusion, at tilrettelæggelse og afvikling af de nationale test generelt gennemføres tilfredsstillende. Det er således evaluators vurdering, at tidligere it-relaterede udfordringer med testafvikling ser ud til at være løst. Det specifikke tidsforbrug til testafvikling vurderes tilsvarende ikke at være en udfordring for skolerne, og evaluator konkluderer, at der er en klar tidsmæssig gevinst for lærerne ved at gennemføre de it-baserede nationale test sammenlignet med papirbaserede test. Disse hovedkonklusioner er funderet i tilbagemeldinger fra lærere og skoleledere på et meget stort antal skoler samt i evaluators observationer i forbindelse med casestudier.

De nationale test adskiller sig fra andre test i folkeskolen ved at gøre brug af **det adaptive princip**, hvor testen løbende tilpasser sig elevens niveau. Evaluator vurderer i denne sammenhæng, at størstedelen af lærerne ser det adaptive princip som en fordel for eleverne og deres testafvikling. Det er samtidig muligt at pege på en række ulemper relateret til elevernes oplevelse af test-situationen, men det er evaluators vurdering, at disse ulemper kan afhjælpes ved hjælp af tydeligere vejledning fra lærerens side. Samlet set vurderer evaluator, at de pædagogiske ulemper opvejes af de pædagogiske fordele, som karakteriserer det adaptive princip, herunder særligt at eleverne hver især udfordres på deres eget faglige niveau.

Endelig viser evalueringen, at det ikke er muligt at konkludere entydigt, hvorvidt de forskellige resultatvisninger i testsystemet bidrager til en hensigtsmæssig **tilbagemelding til lærerne**. Mange lærere angiver i spørgeskemaundersøgelsen, at de umiddelbart finder resultatvisningen brugervenlig, men langt fra alle lærere gør brug af systemets mange muligheder, bl.a. fordi de i højere grad vælger at gøre brug af andre evalueringstværværktøjer. Evalueringen konstaterer i øvrigt flere forbedringspunkter i resultatvisningen, der potentielt kan bidrage til den pædagogiske anvendelse af resultaterne. Hovedkonklusionen er således, at den nuværende resultatvisning anvendes, men den samlede vurdering er, at det fortsat kræver et relativt stort fortolknings- og analysearbejde af den enkelte lærer for at få et udbytte af testresultaterne for den enkelte elev, hvilket udgør en barriere for lærerens omsætning af testresultater til praksis, i fremtidig tilrettelæggelse af undervisningen og undervisningsdifferentiering.

Læs mere om *testenes udformning* i kapitel 8.

9.2 Tværgående hovedkonklusion

Nedenfor foretager evaluator en samlet konklusion på tværs af temaer og datakilder.

Det er evaluators samlede konklusion, at de nationale test har en positiv effekt på elevernes faglige niveau. De gennemførte analyser på elev-, skole- og landsniveau indikerer en positiv effekt på elevernes faglige udbytte. Det skal understreges, at det kun har været muligt at undersøge effekten af det midlertidige nedbrud i dansk, læsning. Samtidig vurderer evaluator, at evalueringsskulturen er styrket i de danske folkeskoler i de senere år. Indførelsen af de nationale test har i nogen grad været en medvirkende årsag til denne udvikling – primært som understøttende for en generel udvikling hen imod stærkere dokumentation og systematik af undervisningspraksis og fagligt udbytte. En hovedkonklusion er dog samtidig, at det ikke er de nationale test alene, der har fremmet en stærkere evalueringsskulturen. Det er snarere mangfoldigheden af nationale og lokale initiativer inden for evaluering, der *i samspil* har bidraget til denne udvikling. De nationale test har således været en del af en *skoleudvikling*, der i højere grad end tidligere, er karakteriseret ved systematik, dokumentation af resultater og (i nogen grad) fastsættelse af mål.

Evalueringen viser, at selv om der kan konstateres positive sammenhænge og effekter af indførelsen af de nationale test, er der en række sammenhænge i evalueringens *forandringsteori*, der ikke har vist sig at holde empirisk. Med andre ord dokumenterer evalueringen, at en række af de forventede resultater, der indgår i forandringsteorien, ikke realiseres i praksis. I forhold til *den pædagogiske anvendelse* er en hovedkonklusion, at de nationale test i dag i langt overvejende

grad anvendes summativt og *ikke* formativt forstået som fremadrettet anvendelse af testresultater til planlægning af undervisningen. Tilsvarende er antagelserne om styrket skole-hjem-samarbejde og bedre dialog med den enkelte elev om fremadrettede læringsmål som følge af de nationale test udfordret i praksis. Evalueringen viser endvidere, at de nationale test tillægges betydelig vægt af såvel skoleledere som kommuner i deres kvalitetsarbejde. Det generelle billede er imidlertid, at skolelederne som hovedregel anvender testresultater til at skabe overblik over det faglige niveau, men sjældnere som omdrejningspunkt for dialog med såvel lærere som forvaltning om den videre skoleudvikling. Evalueringen viser videre, at der på visse områder er forskellige opfattelser blandt særligt lærere og skoleledere af dels betydningen af de nationale test og dels, hvordan de respektive aktører arbejder med test og evaluering. Disse forskelle kan have betydning for forudsætningerne for at udvikle en *fælles* evalueringskultur på de enkelte skoler.

Ovenstående betyder ikke, at de forventede forandringer som følge af indførelse af de nationale test ikke finder sted. Evalueringen har dokumenteret stor variation i måden, hvorpå man anvender de nationale test i kommuner, på skoler og af den enkelte lærer. Vi har afdækket flere eksempler på skoler, hvor forandringsteorien "*holder i praksis*". Det vil sige skoler med en skoleledelse med tydelig fokus på faglige resultater, og hvor de nationale test spiller en betydelig rolle for både skoleledere og for lærerne i deres pædagogiske praksis.

Dokumentationen af positive effekter af de nationale test på den ene side og betydelige mangler i anvendelsen af de nationale test på den anden side (som forudsat i lovgivningen) får samtidig evaluatoren til at konkludere, at der synes at være et potentiale for stærkere effektskabelse fremadrettet. Dette underbygges med en række af gode eksempler, jf. ovenfor.

Samtidig vurderer evaluatoren, at den dokumenterede effekt inden for dansk, læsning bl.a. kan tilskrives, at lærerne i dansk har haft nogle andre muligheder for at anvende de nationale test i læsning hensigtsmæssigt (og i mange tilfælde også formativt). Det gælder dels karakteristika ved testen i dansk, læsning (hyppigheden, et velafgrænset fokus på læsefærdigheder, kvaliteten af opgaverne), men det spiller også ind, at læseområdet har haft fokus i mange år i den danske folkeskole. Der er fagdidaktiske miljøer på skolerne, lærerne understøttes ofte af læsevejledere, og der er supplerende evalueringsværktøjer til rådighed. Med andre ord har *forudsætningerne* for at bruge de nationale test som pædagogisk værktøj i dansk, læsning været mere gunstige end i andre testfag. På den baggrund vurderer evaluatoren, forenklet fremstillet, at forandringsteorien "*holder*" for så vidt angår dansk, læsning. Det er derfor sandsynligt, at effekten på elevernes færdigheder i nogen grad kan tilskrives den pædagogiske indsats, som lærerne har været ansvarlige for.

Det skal afslutningsvis understreges, at data til nærværende evaluering er indsamlet over en periode fra ultimo 2012 og til og med foråret 2013. Tilsvarende har der været anvendt registerdata over en relativt kort tidsperiode. På denne baggrund konstaterer evaluatoren, at de nationale test har været implementeret over få år, hvorfor de mere langsigtede effekter i forandringsteorien givetvis endnu ikke kan registreres. En fortsat gennemførelse af de nationale test med stærkt fokus på implementering i henhold til intentionerne vil sandsynligvis medvirke til, at de nationale test i de kommende år bliver mere virkningsfulde. Tilsvarende vil det blive muligt at indhente data over en længere tidsperiode, hvorfor effektvurderinger kan blive mere robuste og nuancerede.

9.3 Perspektivering

Dette afsluttende afsnit indeholder en perspektivering af evalueringens hovedkonklusioner. Formålet med perspektiveringen er at pege på en række opmærksomhedspunkter, som evaluatoren finder væsentlige i forhold til den videre udvikling af de nationale test. Opmærksomhedspunkterne er funderet i de data, som evaluatoren har indhentet som led i evalueringen.

De nationale test som pædagogisk redskab

Evalueringen dokumenterer, at de nationale test generelt bruges af lærerne i den danske folkeskole. Testene bruges – som det var intentionen – som et pædagogisk værktøj, om end primært til at skabe overblik over klassens og elevernes faglige niveau og i mindre grad til den fremadrettede tilrettelæggelse af undervisningen.

I det videre arbejde med de nationale test synes der således at være et potentiale i forhold til at understøtte *den formative anvendelse* af de nationale test. Evaluator ønsker at pege på følgende opmærksomhedspunkter, der potentielt kan bidrage til at styrke den pædagogiske anvendelse af de nationale test.

Styrket ledelse: Evalueringen viser, at skoleledelsen kan spille en væsentlig rolle, hvis man ønsker at understøtte brugen af de nationale test som et pædagogisk redskab. De skoler, der er kommet længst i arbejdet med at udvikle en evalueringskultur, har en ledelse, der klart og tydeligt har rammesat brugen af de nationale test. Det er entydigt for lærerne på disse skoler, at testene er et evalueringsværktøj, der tillægges værdi. Ledelsen har her fokus på opfølgning på resultatet i form af løbende dialog med medarbejderne. Denne dialog danner grundlag for fremadrettede pædagogiske og organisatoriske beslutninger for både lærere og ledelse.

På dette grundlag synes der at ligge et stort potentiale i at få alle skoleledere til i endnu højere grad at arbejde systematisk og dialogisk med evaluering og evalueringsresultater (herunder nationale test), således at ledelsesmæssige beslutninger kan ske på et oplyst grundlag.

Styrket vejledning, bl.a. fra ressourcepersoner: Flere lærere i evalueringen peger på, at de fortsat ser en udfordring i at oversætte resultaterne af de nationale test til pædagogisk praksis. Det er med andre ord vanskeligt for mange lærere at gå fra den summative til den formative anvendelse.

Et af målene med den kommende reform af folkeskolen er en præcisering og forenkling af Fælles Mål med henblik på udarbejdelse af læringsmål, der sætter elevernes læringsudbytte tydeligere i centrum¹⁰⁰. Tydelige mål for elevernes læring skal bidrage til at øge det faglige niveau. Resultaterne fra de nationale test kan i den sammenhæng være en væsentlig indikator på elevens faglige niveau inden for dele af Fælles Mål.

På baggrund af evalueringen er der tydelige indikationer på, at ressourcepersoner/faglige vejledere kan bidrage til at understøtte de refleksioner og dialoger, der ligger i arbejdet med at formulere tydelige (lærings)mål, at dokumentere fagligt udbytte og ikke mindst at forholde sig til pædagogisk praksis. Dvs. på baggrund af testresultatet at hjælpe læreren med at se den enkelte elev og de faglige styrker/udfordringer han/hun har, samt vejlede til at finde frem til den rette pædagogiske indsats/undervisning/undervisningsmaterialer.

I den sammenhæng synes det oplagt at trække på de erfaringer, som der er gjort særligt på læseområdet i kommuner og på skoler over hele landet. Dog er der også gode erfaringer fra andre testfag. Et væsentligt opmærksomhedspunkt er således, at lærerne har brug for vejledning til at omsætte testresultater (og resultater fra anden evaluering) til pædagogisk praksis. Herigennem styrkes mulighederne for videndeling tilsvarende.

Antallet af test: På baggrund af evalueringen kan det konstateres, at antallet af test inden for hvert fag spiller en rolle for lærernes (mulighed for) pædagogisk anvendelse af testen. Dette fremgår tydeligst på læseområdet, hvor der gennemføres flest nationale test, og hvor man generelt er længst fremme i forhold til at arbejde formativt med testresultaterne.

Som begrundelse for fravalg af de nationale test som pædagogisk værktøj peger flere lærere netop på, at testen gennemføres for sjældent til at give mening som et (pædagogisk) redskab, der kan danne grundlag for henholdsvis løbende monitorering og fremadrettet anvendelse. De fo-

¹⁰⁰ Aftale mellem regeringen (Socialdemokraterne, Radikale Venstre og Socialistisk Folkeparti), Venstre og Dansk Folkeparti om et fagligt løft af folkeskolen (7. juni 2013).

retrækker at teste oftere, så de løbende og kontinuerligt kan følge elevernes faglige udvikling eller mangel på samme og efterfølgende kan tilrettelægge en undervisning på baggrund heraf.

Dertil kommer, at visse af testene i dag er placeret på (for nogle lærere) u hensigtsmæssige klassetrin, som fx matematik i 6., hvor flere lærere afgiver klassen og derfor ikke har tid til og mulighed for at anvende resultatet som grundlag for den fremadrettede planlægning. Et andet eksempel er test inden for flere naturfag i 8. klasse, hvor der er meget kort tid til opfølgning, inden eleverne ophører med at have faget/forlader folkeskolen.

Fra skoleåret 2014/2015 vil eleverne i henhold til folkeskolereformen skulle undervises i engelsk fra 1. klasse. Der vil derfor gå syv år (seks år, hvis de frivillige test tælles med), før lærerne har de nationale testresultater til rådighed som pædagogisk værktøj. Denne lærergruppe er i forvejen meget positive over for testen, og det må formodes, at der vil være en stor interesse for – og anvendelse af – en national test tidligere i skoleforløbet.

På baggrund af evalueringens resultater samt de kommende ændringer af folkeskolen vurderes det relevant at overveje antallet af test inden for de forskellige fag, herunder placering på klassetrin, med henblik på at understøtte den pædagogiske anvendelse af testresultaterne.

Bedre resultatvisninger: Hvis brugen af de nationale test som pædagogisk redskab skal styrkes, er det værd at overveje, hvordan mulighederne for resultatvisninger kan styrkes. Evalueringen viser, at det er ikke nok i forhold til at sikre pædagogisk anvendelse, at lærerne er i stand til at navigere rundt på testsystemets hjemmeside og forstå resultaterne umiddelbart. Der synes at være et behov for forholdsvis let ikke kun at udpege fx elever med faglige udfordringer, herunder også hvilke faglige udfordringer der er tale om. Mange lærere oplever, at det er omstændeligt at få et præcist billede af den enkelte elevs faglige profil, der efterfølgende vil kunne danne grundlag for fx holddannelse eller individuelle undervisningstiltag. Det kræver reelt, at læreren går ind i systemet og åbner hver enkelt testopgave. Derudover efterlyses der en mulighed for at kunne gruppere eleverne inden for de faglige temaer, hvor de måtte have udfordringer/styrker.

Der synes således at være et potentiale i forhold til at videreudvikle de eksisterende funktionaliteter i resultatvisningen, der i højere grad kan understøtte den fremadrettede pædagogiske anvendelse.

I forlængelse heraf skal det nævnes, at de nationale test vil blive teknisk omlagt i forbindelse med den kommende reform af folkeskolen, så de kan anvendes af lærerne som et pædagogisk redskab til at vurdere elevernes kompetenceniveau og progression i forhold til på forhånd fastlagte faglige kriterier (i dansk og matematik).

De nationale test som styringsværktøj

Evalueringen viser, at såvel skoleledere som kommuner gør brug af resultaterne af de nationale test som styringsværktøj, men at der samtidig er meget stor variation i graden af anvendelse. Det er på den baggrund en hovedkonklusion, at brugen af de nationale test som styringsværktøj kan styrkes betydeligt, så testene i højere grad indgår i skolelederes og kommuners kvalitetsarbejde.

Det fremadrettede arbejde med at styrke de nationale test som styringsværktøj kan bl.a. bygge videre på de gode eksempler, der allerede findes. Evalueringen har dokumenteret flere eksempler på kommuner og skoler, hvor de nationale test indgår som en væsentlig del af arbejdet med at formulere faglige mål, løbende og systematisk dialog samt konkrete handlinger baseret på testresultater.

Evaluatoren konstaterer, at det som led i folkeskolereformen er en ambition dels at bruge de nationale test som styringsværktøj i relation til *nationale mål* for folkeskolen, og dels at omlægge de nationale test til såkaldte *kriteriebaserede test*, jf. ovenfor. Sidstnævnte kan *umiddelbart* skabe bedre forudsætninger for at bruge de nationale test som led i skolelederes og kommuners kvalitetsarbejde, da det i princippet vil være lettere at følge op på formulerede målsætninger (nationalt, kommunalt og på forskellige niveauer på den enkelte skole).

Der er imidlertid også flere potentielle ulemper, der bør imødegås ved en sådan omlægning. En bekymring kan være overdreven fokus på den styringsmæssige dimension, der kan føre til øget snyd med testene. En anden bekymring kan være, at lærerne satser på at løfte de elever, der ligger tæt på de nationalt eller lokalt fastsatte kriterier for fx "en god læser" – og dermed undlader at bruge testresultater til at fokusere på den enkelte elevs styrker og svagheder og tilrettelægge en differentieret undervisning.

Med andre ord er det et væsentligt opmærksomhedspunkt, at overdreven brug af testresultater i en styringsoptik kan komme i modstrid med den oprindelige intention relateret til at bruge testene som et pædagogisk redskab. Flere amerikanske forskningsstudier peger på u hensigtsmæssige konsekvenser af brugen af test¹⁰¹.

Et yderligere opmærksomhedspunkt knytter sig til kravet i folkeskoleloven om, at testresultater er fortrolige. Kravet om fortrolighed medfører, at der i nogle tilfælde er usikkerhed om, hvordan testresultater kan anvendes. Det gælder dog særligt som styringsværktøj, når testresultater aggregeres og skal anvendes af skoleledelsen og de kommunale forvaltninger. Konsekvensen er i nogle tilfælde, at testresultater *ikke* eller *i begrænset omfang* anvendes (fx i relation til dels skolebestyrelse, dels kommunalbestyrelse) med henvisning til fortrolighedsprincippet. Fremadrettet synes der at være behov for tydeligere retningslinjer for og vejledning til, hvordan nationale testresultater må og kan anvendes proaktivt. Dette skal bidrage til, at kvalitetsudvikling af skoler og det kommunale skolevæsen sker i en dialog baseret på dokumentation og ikke, som det har været fremhævet, ud fra "*principper om lukkethed*".

Brug de gode eksempler

Evaluators har ovenfor fremhævet potentialet for styrket brug af de nationale test – både pædagogisk og som styrings- og dialogværktøj. Grundlaget for dette er de gode eksempler, som evaluator har mødt i forbindelse med analysearbejdet. De gode eksempler findes derude i de kommunale forvaltninger, på skolerne og i klasseværelserne. Det er derfor relevant at overveje, at de gode eksempler *i hele styringskæden* spredes og mangfoldiggøres. Mange forvaltninger, ledere og lærere har fortsat ikke set mulighederne i de nationale test. Der ligger et stort formidlings- og vejledningsarbejde i at få disse aktører til at se potentialerne.

Brugen af det adaptive princip

Evalueringen peger på, at brugen af det adaptive princip i de nationale test har en række fordele. Dette gælder dels pædagogisk såsom hurtigere og mere præcis fastlæggelse af elevernes faglige niveau, og dels en bedre oplevelse for eleverne for såvel fagligt svagere som fagligt stærkere elever.

Evalueringen peger dog også på, at formidlingen af testresultater er en udfordring, fordi den ikke er intuitiv, enkel og let forståelig for elever og forældre. Der kan derfor være et perspektiv i at vurdere mulighederne for at understøtte lærernes samarbejde med elever og forældre med en lettere resultatvisning, der i højere grad kan understøtte dialogen.

Endelig viser evalueringen, at brugen af det adaptive princip i de nationale test kræver betydelig ressourceanvendelse, idet alle opgaver, der udvikles eller ændres bare det mindste, skal gennemgå en meget ressourcekrævende afprøvning og Rasch-test. Denne proces tager ifølge evaluators oplysninger ca. et år. Med udgangspunkt i dette opmærksomhedspunkt kan der eventuelt være perspektiv i at benytte den tilgang, der er fremhævet i det internationale litteraturstudie som en mulighed: At afprøve opgaver som en del af selve testafviklingen, hvor eleverne stilles opgaver under afprøvning blandt selve testopgaverne, uden at opgaverne under afprøvning medregnes i elevernes testresultat.

Udvikling af kvaliteten af opgaver fra de nationale test

Evalueringen viser, at valg af henholdsvis testfag og profilområder generelt har været hensigtsmæssige. Imidlertid peger evalueringen på, at der er flere udfordringer knyttet til testopgavernes

¹⁰¹ Left Behind by Design: Proficiency Counts and Test-Based Accountability. Review of Economics and Statistics 92, 263-283, Neal, D. and D. W. Schanzenbach, 2010, Teaching to the rating: School accountability and the distribution of student achievement. Journal of Public Economics 92: 1394:1415, Reback, R., 2008.

kvalitet samt opgavebankens størrelse. Evaluator er i den forbindelse opmærksom på, at der løbende arbejdes på at udvikle kvaliteten af opgavebanken, herunder at der er tale om en ressourcetrækkende proces, når opgaver skal ændres, jf. ovenfor.

I forhold til den fremtidige udvikling af opgavernes kvalitet er det et generelt opmærksomhedspunkt, at *løbende kvalitetsudvikling/-sikring* af opgavebanken skal prioriteres. Høj kvalitet af de opgaver, der anvendes i de nationale test, er en forudsætning for, at testene opfattes som valide, genkendelige og ikke mindst anvendelige af alle aktører, hvad enten testresultater skal bruges i pædagogiske eller styringsmæssige sammenhænge.

Et særligt opmærksomhedspunkt, som vurderes at kunne bidrage til kvaliteten af opgavebanken, er en øget professionalisering af arbejdet med at udvikle opgaver/items. Det kunne være via (efter)uddannelse af opgaveudviklere samt tættere kobling til den nyeste forskning.

Fokus på elevernes læring

Evalueringen viser, at eleverne overordnet set er positive over for de nationale test. En hovedudfordring er imidlertid, at eleverne typisk modtager kollektiv tilbagemelding fra lærerne om testresultater. Den *individuelle* tilbagemelding til eleverne om styrker og svagheder samt overvejelser om fremadrettede forbedringspotentialer er således mangelfuld.

Der synes på baggrund af evalueringen at være perspektiver i fremadrettet at fokusere på, hvordan de nationale test kan bidrage til at formulere læringsmål ikke blot for klassen, men også for og med den enkelte elev. Med folkeskolereformen in mente kan de nationale test i højere grad bidrage til at sætte elevernes læring i centrum – både som grundlag til at formulere fremadrettede mål og som redskab til at dokumentere udviklingen i læringsudbytte. En tydeligere rammesætning fra skoleledelsens side samt videndeling om/vejledning til lærere vedrørende arbejdet med at omsætte testresultater til læringsmål synes at være relevante opmærksomhedspunkter i den fremadrettede udvikling af de nationale test.