

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

Dokumenttype
Bilag til Evaluering af de nationale test i folkeskolen

Dato
September 2013

BILAG 1: METODEBILAG



METODEBILAG

BILAG 1: METODEBESKRIVELSE

Nærværende bilag indeholder en detaljeret beskrivelse af metoden anvendt i evalueringen af de nationale test. Som supplement hertil har evaluator udarbejdet en detaljeret redegørelse for registeranalyserne anvendt til effektanalyserne, som fremgår af bilag 3.

Evalueringen af de nationale test baserer sig på en række aktiviteter og dokumentation, der er gennemført/indsamlet i perioden november 2012 til juni 2013:

- Desk research
- Interessentinterview
- Eksplorative casestudier
- Registerdata
- Spørgeskemaundersøgelser (blandt skoleforvaltninger, skoleledere og lærere)
- Analyse af testdata
- Casestudier
- Ekspertvurdering 1 og 2.

Kommissoriet fra Kvalitets- og Tilsynsstyrelsen indeholder en række evalueringsspørgsmål, som skal belyses i denne evaluering. I kommissoriet er disse markeret med bogstaverne A til P. Med henblik på at skabe overblik over, hvilke datakilder der belyser hvilke evalueringsspørgsmål, har evaluator udarbejdet nedenstående tabel. Det skal bemærkes, at interessentinterview og eksplorative casestudier ikke fremgår af tabellen, da disse ikke udgør selvstændige datakilder, men derimod har bidraget til at kvalificere den øvrige dataindsamling. Dog beskrives formål med og tilgang til disse aktiviteter nedenfor i afsnittet.

Som det fremgår af rapporten, har vi i analysen lagt vægt på at datatriangulere, da dette bidrager til at forstærke og/eller nuancere analyser og konklusioner. Nedenstående tabel viser, at der i høj grad har været mulighed for datatriangulering, eftersom de fleste evalueringsspørgsmål belyses gennem flere datakilder.

Tabel 1: Oversigt over evalueringsspørgsmål og datakilder

Tema	Evalueringsspørgsmål	Datakilder				
		Desk research	Registerdata (fx testdata)	Spørgeskemaundersøgelser	Casestudier	Ekspertvurdering 1 og 2
Tema 1: Testenes effekt	A: Vurdering af, om elevernes faglige niveau er styrket efter indførelsen af de nationale test		✓			
	B: Vurdering af, om skolernes evalueringskultur er styrket efter indførelsen af de nationale test	✓	✓	✓	✓	
Tema 2: Testenes indhold	E: Testenes udvalg af fag, faglige emner og relation til fagernes trinmål i Fælles Mål	✓	✓			✓
	I: Vurdering af testenes udformning i forhold til de forskellige fag	✓	✓			✓
	F: Testopgavernes faglige kvalitet og antallet af opgaver	✓	✓			✓
	H: Vurdering af fordele og ulemper ved det adaptive princip (teknisk/økonomisk)	✓				✓
	C: Testene som bidrag til at give læreren et nuanceret billede af elevernes kundskaber og færdigheder			✓	✓	
Tema 3: Testenes udformning	G: Den praktiske tilrettelæggelse på skolerne, herunder tids- og ressourceforbrug			✓	✓	
	H: Vurdering af fordele og ulemper ved det adaptive princip (pædagogisk)	✓		✓	✓	
	J: Tilbagemelding til lærerne i form af forskellige resultatvisninger og deres bidrag til, at læreren får et nuanceret billede af elevernes kundskaber og færdigheder			✓	✓	✓
Tema 4: Testenes anvendelse	L: Lærernes anvendelse af testresultaterne til at vurdere og følge elevernes faglige styrker, svagheder og potentialer			✓	✓	
	M: Lærernes anvendelse af testresultaterne, som bidrag til et bedre grundlag for planlægningen af undervisningen			✓	✓	
	D: De nationale tests indflydelse på lærernes prioritering af undervisningens indhold			✓	✓	
	O: Testresultaternes værdi for forældrene i skole-hjem-samarbejdet			✓	✓	
	P: Elevernes oplevelse af de nationale test			✓	✓	
	N: Anvendelse af testene som værktøj for kommuners og skolelederes kvalitetsarbejde	✓		✓	✓	
	K: Testenes bidrag til målbare arbejder og bedre opgaveløsning			✓	✓	

Analysestrategi til analyse af effekt og sammenhænge

Evalueringen af de nationale test undersøger årsagssammenhængen 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.

Overstående 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 afskiller sig mht. deltagelse i test. Ligeledes analyserer vi, om der er forskelle på resultater hos elever, der har været udsat for testnedbrud 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 indført (Calmar Andersen 2004). 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.

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

Evalueringsdesignet 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 en teoribaseret evalueringstilgang med definition af en forandringsteori for årsags-virknings-kæder mellem input-aktivitet-output-outcome-impact 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. Evaluator har således udviklet en forandringsteori med to spor; et spor med fokus på lærerne og den pædagogiske praksis og et spor med fokus på ledelse. Med afsæt heri har evaluator endvidere udviklet en række hypoteser, der løbende testes i evalueringens analyser. Forandringsteorier såvel som hypoteser er præsenteret i nærværende bilag.

Model for evalueringskultur

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

Modellen består af fire niveauer, som udgør et kontinuum fra niveau 1 (der udtrykker overholdelse af kravet om afholdelse af obligatoriske test, men hvor resultaterne ikke anvendes i overensstemmelse med lovens intentioner) til niveau 4 (der udtrykker fremadrettet brug af testene som et pædagogisk redskab, der kan understøtte høj grad af undervisningsdifferentiering).

Modellen afspejler først og fremmest Kvalitets- og Tilsynsstyrelsens egen definition af evalueringskultur samt Peter Dahler Larsens definition af evalueringskultur/evalueringskapacitet, der supplerer denne definition med nogle ekstra elementer.

Den gode evalueringskultur medvirker ligeledes til kontinuerlig pædagogisk udvikling og forandring. Derfor har evaluator i arbejdet med modellen inddraget undersøgelser/reviews, der peger på de elementer, der understøtter dette i en undervisningssammenhæng. Først og fremmest Creating Effective Teaching and Learning Environments. First Results from TALIS. OECD (2009), School leadership and student outcome: Identifying what works and why. Auckland: University of Auckland, Robinson, V et al (2009), Ledelse, læring og trivsel i folkeskolen. SFI. Det Nationale Forskningscenter for Velfærd, Andersen, S.C. et al. (2011) samt Review on Evaluation and Assessment in Education: Denmark 2011, OECD.

Modellen anvendes især som et *analytisk værktøj* til at kategorisere skoler i forhold til, hvordan de arbejder med evaluering, ligesom modellen har dannet grundlag for udvikling af spørgeskemaer og interviewguides.

Forandringsteorier og hypoteser

Evalueringsens analytiske afsæt er en overordnet **forandringsteori** for den intervention, der er genstand for evalueringen, dvs. en forandringsteori for indførelse af de nationale test i folkeskolen. Forandringsteorien står centralt i evaluators evalueringsdesign (virkningsevaluering uddybet med *contribution analysis*).

Med udgangspunkt i såvel forandringsteori som modellen for evalueringskultur (beskrives nedenfor) har evaluator formuleret en række **hypoteser** om sammenhænge i forandringsteorien. Brug af hypoteser til at skærpe de forhold, der skal testes empirisk, er særdeles udbredt inden for forskning og nyere evalueringstilgange, herunder virkningsevaluering. Hypoteserne repræsenterer de forhold, som evaluator finder afgørende at teste som led i evalueringen og med et klart afsæt i kommissoriets evalueringsspørgsmål. Tabellen nedenfor præsenterer evalueringens hypoteser, ligesom det fremgår, på hvilken baggrund hypotesen er dannet.

Tabel 2: Oversigt over hypoteser

Hypoteser	Baggrund for hypoteser
Overordnede effekthypoteser	
1. Anvendelse af nationale test styrker elevernes faglige niveau	Lovgrundlag og kommissorium
2. Anvendelse af nationale test styrker folkeskolernes evalueringskultur	Lovgrundlag og kommissorium
3. En stærk evalueringskultur i folkeskolen styrker elevernes faglige niveau	Udledt af ovenstående
Skoleledelse	
4. Formel lederuddannelse hos skoleledelsen betyder stærkere evalueringskultur på skolen	Inspireret af SFI 2011
5. Skoleledelsens formulering af klare retningslinjer og målsætninger for anvendelsen af de nationale test har positiv betydning på lærernes pædagogiske anvendelse af de nationale test	Rambøll 2011 OECD AKF 2010
6. 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)	Inspireret af SFI 2011

Hypoteser	Baggrund for hypoteser
7. 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)	TALIS OECD Robinson et al. 2009
Lærerne	
8. 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)	Lovgrundlag og kommissorium
9. Resultatet af de nationale test målretter den enkelte lærers planlægning af undervisning i relation til klassen og den enkelte elev	Lovgrundlag og kommissorium
10. Resultatet af de nationale test styrker den enkelte lærers differentiering af undervisning	Lovgrundlag og kommissorium
11. Inddragelse af resultater fra de nationale test styrker og kvalificerer dialogen i lærerteams om udvikling af undervisning	Vejledning udgivet af KTST
12. Inddragelse af ressourcepersoner styrker fortolkningen af resultatet af de nationale test og målretter lærerens opfølgning i relation til den enkelte elev/klassen	Inspireret af EVA 2008 Vejledning udgivet af KTST
Eleven	
13. Lærerens tilbagemelding og dialog med eleven om resultatet af nationale test styrker elevernes bevidsthed om egne læreprocesser og dermed motivation for læring	Lovgrundlag og kommissorium Interessentinterviews
Skole-hjem-samarbejdet	
14. Nationale test styrker skole-hjem-samarbejdet	Lovgrundlag og kommissorium Interessentinterviews
15. Dialog med hjemmet om resultatet af de nationale test fremmer forældrenes mulighed for at understøtte deres barns læring	Interessentinterviews
Forvaltningen	
16. Skolernes og forvaltningens aktive samarbejde/dialog om skolens testresultater fører til en målretning af skolernes arbejde (og dermed forbedrede faglige resultater)	Lovgrundlag og kommissorium
17. Et klart kommunalt fokus på faglige resultater har en positiv indflydelse på skolernes faglige niveau	Inspireret af AKF 2010

Dataindsamlingsaktiviteter

Dataindsamlingsaktiviteterne og kvaliteten af de tilvejebragte data beskrives i det følgende for de enkelte datakilder i evalueringen.

Desk research, interessentinterview og eksplorative casestudier blev gennemført som led i evalueringens opstartsfasen og beskrives kort herunder, mens beskrivelsen af de øvrige dataindsamlingsaktiviteter beskrives mere udførligt.

Desk research

Et væsentligt afsæt for evalueringen er, at evaluatoren er informeret om og funderet i den viden, som eksisterer på området. Det gælder lovgrundlag, relevante undersøgelser og forskning samt den dokumentation, som Kvalitets- og Tilsynsstyrelsen besidder efter en årrække med forberedelse og gennemførelse af projektet om de nationale test.

I forhold til sidstnævnte har evaluatoren afholdt et møde med styrelsens nøglemedarbejdere og er blevet opdateret på den viden, som styrelsen ligger inde med. Derudover har evaluatoren gennemført egen research af relevant viden, som er indgået i arbejdet med hypoteser, forandringsteorier mv.

Interessentinterviews

Evaluators har i opstartsfasen gennemført interviews med en række interessenter på folkeskoleområdet. Følgende organisationer blev interviewet:

- Danmarks Lærerforening v. afdelingschef Jesper Støier
- Skolelederforeningen v. formand Anders Balle
- Kommunernes Landsforening v. chefkonsulent Anette Aunbirk
- Skole og Forældre v. nu afdækket formand Benedikte Ask Skotte
- Danske Skoleelever v. nu afdækket formand Vera Rosenbeck.

Interviewene er gennemført med henblik på at give interessenterne mulighed for at give input til evalueringen. Temaerne for interviewene var følgende: Interessenternes involvering i projektet, status for anvendelse af test; styrker og svagheder ved brug af de nationale test (herunder målopfyldelsen); anvendelse og nytte af de nationale test; opmærksomhedspunkter for evalueringen, samt bud på rivaliserende/alternative forklaringer på effekt.

Interviewene har bidraget til at kvalificere udarbejdelsen af en model for evalueringskultur og at skærpe fokus i den kommende spørgeskemaundersøgelse blandt skoleledere. De vil på samme måde indgå og understøtte udarbejdelse af henholdsvis spørgeskemaundersøgelse blandt forvaltninger og spørgeskemaundersøgelse blandt lærere.

To eksplorative casestudier

Evaluators har ligeledes i opstartsfasen gennemført eksplorative casestudier på to skoler.

Formålet har været at sikre evaluators en fokuseret forståelsesramme i dialog med praktikere. Det vil sige at få et indblik i skoleledere og læreres aktuelle forståelse af evalueringspraksis, begreberne *evalueringskultur* og *styrket faglighed* samt deres kobling til de nationale test. Desuden blev der spurgt ind til daglig praksis og konkrete erfaringer i relation til testene.

Evaluators har endvidere anvendt de eksplorative casestudier til at spørge ind til, hvilke alternative forklaringer der kan være på eleveres fremgang inden for testfagene. Også disse interviews har bidraget til at kvalificere og skærpe det videre arbejde med ikke mindst skoleleder- og lærerspørgeskemaer samt alternative forklaringer.

Samlet set har deskresearch, interessentinterview og eksplorative casestudier bidraget til at tilvejebringe viden, der skærper analysestrategien og kvalificerer dataindsamlingen.

Registeranalyser

Registeranalyserne har til formål at undersøge effekten af de nationale test.

Som udgangspunkt er det meget vanskeligt statistisk at undersøge effekten af de nationale test. 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 forhold. Udfordringen er, at der ikke har noget at sammenligne med. Alle eleverne i testsystemet har taget testene, og vi ved derfor ikke, hvordan de ville have klaret sig, hvis de ikke havde taget en test.

For at imødegå disse vanskeligheder med at undersøge effekten af de nationale test i sig selv, har evaluators til denne evaluering anvendt to metoder. Til effekter på elev- og skoleniveau benytter vi os af et sammenbrud i it-systemet i 2010, der gjorde, at ikke alle elever blev testet det år. Til effekter på landsplan undersøger vi forskelle mellem årgange lige før og lige efter testenes indførelse. Datagrundlag og gennemførte analyser beskrives nærmere i bilag 3.

Spørgeskemaundersøgelser

Der er gennemført spørgeskemaundersøgelser blandt hhv. forvaltninger, skoleledere og lærere.

Det primære formål med spørgeskemaundersøgelserne blandt kommunale forvaltninger var at opnå et overblik over kommunale politikker, strategier og vurderinger af evalueringsspørgsmål om karakteren af evalueringskultur på kommunens skoler.

Det primære formål med spørgeskemaundersøgelserne blandt hhv. skoleledere og lærere var at opnå et overblik over skolelederens og læreres erfaringer med test og vurderinger af evalueringspraksis. I hvilken grad og hvordan de nationale test skaber en evalueringskultur og bidrager til refleksion over elevernes faglige styrker, svagheder og potentialer, og en efterfølgende planlægning og tilrettelæggelse af undervisningen, der søger at gavne de enkelte elevers udvikling mest muligt.

Således har spørgeskemaundersøgelserne bidraget til at identificere og forklare eventuelle sammenhænge mellem indførelsen af de nationale test og hhv. evalueringskultur og elevernes faglige niveau. For at kunne identificere og forklare disse sammenhænge har spørgeskemaundersøgelserne belyst en række temaer, som er uddybet nedenfor.

Temaer i spørgeskemaundersøgelse til forvaltningerne

Det overordnede formål med spørgeskemaundersøgelsen til forvaltningerne var at afdække, hvordan de nationale test anvendes som værktøj i kommunernes kvalitetsarbejde. Samtidig var det vigtigt at undersøge, i hvor høj grad kommunen har et klart fagligt fokus på skolernes faglige resultater, og om de formulerer resultatmål i relation til dette. Intentionerne i lovgrundlaget (lovforslaget om de nationale test mv.) har været et væsentligt grundlag for at formulere spørgsmål til spørgeskemaet.

På den baggrund blev spørgeskemaet struktureret efter følgende temaer:

1. Kommunale indsatsområder og fokus på kvalitet
 - a. Kommunale initiativer på skoleområdet
 - b. Har kommunen opstillet mål for skolers faglige resultater, trivsel e.l.?
2. Dialog mellem forvaltning og skoler
 - a. Hvilken form antager dialogen mellem skoler og forvaltning (skolelederkonference, tilsyn, rådgivning fra pædagogiske konsulenter, ad hoc kontakt)
 - b. Hvor hyppig er kontakten mellem skoler og forvaltning
 - c. Hvilke områder har forvaltningen kontakt med skolerne om (skolens økonomi, faglig udvikling af skolen, pædagogisk udvikling af skolen, elever med særlige behov, socialt udsatte elever, personalejuridiske problemstillinger, kommunens skolepolitiske målsætning, kvalitetsudviklingsredskaber)
3. Kvalitetsrapporter
 - a. Hvilke parametre indgår i kvalitetsrapporterne, og hvordan anvendes kvalitetsrapporterne som styrings- og dialogværktøj over for skolerne?
 - b. Kommunernes anvendelse af den nationale præstationsprofil og resultaterne fra de nationale test
 - c. Hvordan inddrages skolerne i udarbejdelsen af rapporterne?
 - d. Hvordan følger kommunerne op på (utilfredsstillende) resultater i kvalitetsrapporterne?

Temaer i spørgeskemaundersøgelse til skolelederne

Det overordnede formål med spørgeskemaundersøgelsen til skolelederne var at afdække skolens overordnede evalueringskultur, herunder hvilke redskaber skolen bruger til evaluering, og særligt hvordan resultaterne fra de nationale test anvendes i skoleledelsens pædagogiske ledelse.

Spørgsmålene er designet med henblik på at kunne vurdere, om der er sket en udvikling i skolens evalueringskultur efter de nationale test blev indført. Derudover ønskes der et overordnet indblik i, hvordan de nationale test anvendes blandt lærergruppen (efter skoleledelsens opfattelse).

På den baggrund blev spørgeskemaet struktureret efter følgende temaer:

1. Kontaktoplysninger på lærere
2. Baggrundsoplysninger
3. Organisering
4. Skoleledelse

5. Evaluering og evalueringskultur
6. Anvendelse og holdninger til nationale test.

Temaer i spørgeskemaundersøgelsen til lærerne

Det overordnede formål med spørgeskemaundersøgelsen til lærerne er at undersøge, hvordan de danske folkeskolelærere *i praksis* anvender de nationale test. Det vil sige, om testresultaterne bruges til at vurdere elevers faglige styrker og svagheder og derefter anvendes resultaterne som grundlag for en bedre fremadrettet planlægning af undervisningen?

På den baggrund blev spørgeskemaet struktureret efter følgende temaer:

1. Baggrundsoplysninger på lærerne
 - a. Alder, køn, uddannelse, erfaring etc.
2. Implementeringsmiljø/organisering af arbejdet
 - a. Prioritering/vurdering af opgaver
 - b. Vurdering af egne kompetencer
 - c. Vurdering af ledelsens involvering i drøftelser af metode og undervisning
3. Evaluering generelt
 - a. Anvendelse af forskellige evalueringsredskaber/metoder (forandring i anvendelse over tid)
 - b. Anvendelse af evalueringsresultater
4. De nationale test
 - a. De overordnede rammer for arbejdet med nationale test
 - b. Lærerens anvendelse af nationale test til at vurdere og følge elevernes faglige styrker, svagheder og potentialer og som bidrag til et bedre grundlag for planlægningen af undervisningen
 - c. Lærerens syn på nationale test (formål, anvendelighed, tilbagemelding til lærerne i form af forskellige resultatvisninger, bidrag til at læreren får et nuanceret billede af elevernes kundskaber og færdigheder)
 - d. De nationale tests indflydelse på lærernes prioritering af undervisningens indhold
 - e. Lærerens brug af resultaterne i skole-hjem-samarbejdet
 - f. Skoleledelsens dialog med læreren om testresultaterne (opfølgning, sparring).

Spørgeskemaundersøgelserne bidrager således til at belyse alle temaer i evalueringen (jf. tabel 1 i nærværende bilag). Indholdsmæssigt har nogle dele af spørgeskemaet været det samme for alle tre respondentgrupper, mens andre dele af skemaet har varieret på tværs af de tre respondentgrupper.

Population og fremgangsmåde

Dataindsamlingen foregik i perioden december 2012 til februar 2013, hvor spørgeskemaundersøgelsen til skolelederne og til forvaltningen blev udsendt først, hvorefter spørgeskemaundersøgelsen til lærerne blev udsendt. Spørgeskemaundersøgelsen er gennemført som en totalundersøgelse blandt skoleledere på samtlige af landets folkeskoler og blandt alle 98 (skole-)forvaltninger, hvor skolecheferne udgør respondentgruppen.

Spørgeskemaundersøgelserne involverer derfor følgende tre populationer:

1. Samtlige kommunale (skole-)forvaltninger i Danmark
2. Samtlige skoleledere i Danmark
3. Omkring 4.000 lærere, der blev udvalgt af deres ledere på baggrund af evaluators opstillede kriterier.

Spørgeskemaundersøgelsen til skolelederne blev sendt til skolens overordnede e-mail og adresseret til skolelederen. Det fremgik af følgen mailen, at respondenterne skulle være repræsentanter for den samlede skoleledelse og have indsigt i skolens arbejde med de nationale test. Det behøver derfor ikke være skolelederen selv, der besvarede spørgeskemaet.

Spørgeskemaundersøgelsen til lærerne blev sendt til de lærere, som skolelederne havde udvalgt ud fra følgende kriterier (opstillet af evaluators):

- En lærer for hvert af de nedenstående fem fag, der sidste skoleår underviste i:
 - Dansk, læsning i 2. klasse
 - Matematik i 3. klasse
 - Dansk i 4. klasse
 - Enten matematik eller dansk i 6. klasse
 - Engelsk i 7. klasse.
- To lærere, der sidste år underviste i ét af følgende fire fag på 8. klassetrin:
 - Geografi, biologi, dansk, læsning eller fysik/kemi.

Inden udsendelse af spørgeskemaerne blev de pilottestet på 3-5 respondenter, der repræsenterede henholdsvis forvaltning, skoleledere og lærere.

Svarprocenter

Der er høje svarprocenter i alle spørgeskemaundersøgelserne. Population og svarprocenter fremgår af tabellen nedenfor.

Tabel 3: Svarprocenter for spørgeskemaundersøgelserne

	Population (antal)	Respondenter (antal)	Svarprocent
Forvaltning	98	78	80 pct.
Skoleledere	1.209	728	60,2 pct.
Lærere	3.658*	2.162	59 pct.

Note: *Bruttostikprøve, hvor der oprindeligt var 3868, men 210 faldt ud af stikprøven pga. returnerede e-mails.

Datakvalitet

Spørgeskemaundersøgelsen er som nævnt gennemført med høje svarprocenter for forvaltning, skoleledere og lærere. Mellem 59 og 80 pct. for hver af de tre respondentgrupper, der 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.

Datafremstilling

Data fra spørgeskemaundersøgelserne er primært præsenteret i tabeller, hvor de simple frekvenser er vist. For alle spørgsmål i spørgeskemaundersøgelsen til lærerne gælder det, at der er kontrolleret for sammenhænge mellem hhv. klassetrin og testfag. De steder, hvor der fremgik en relevant sammenhæng, vil det være bemærket og/eller illustreret ved en tabel. For de andre to spørgeskemaundersøgelser er der ikke konsekvent kontrolleret for en sammenhæng, men der er kontrolleret for udvalgte spørgsmål, som ud fra en teoretisk eller empirisk begrundet forventning kunne have en sammenhæng. Derudover er der konstrueret indeks (fx over lærernes holdning til de nationale test) til at teste for bredere sammenhænge, hvor det var relevant. Samtlige indeks har et Cronbachs Alpha på 0,8 eller derover, og korrelationerne mellem hver item er på 0,3 eller derover (Pearsons r)¹.

Ekspertvurdering 1 og 2

Ekspertvurdering betegner en metodisk tilgang, hvor evaluators inddrager relevante eksperter i analyserne af udvalgte evalueringsspørgsmål i evalueringen. Til nærværende evaluering er der gennemført to ekspertvurderinger. Ekspertvurdering 1 har fokus på styrker og svagheder ved fag, faglig kvalitet i opgaverne mv., mens ekspertvurdering 2 har fokus på fordele og ulemper ved det adaptive princip. Tilgangen i de to ekspertvurderinger er lidt forskellig. Dette beskrives nedenfor.

¹ Testet ved item-item analyse og item-scale analyse.

Ekspertvurdering 1

Ekspertvurdering 1 har til formål at vurdere testenes udvalg af fag, faglige emner og relation til fagenes trinmål i Fælles Mål samt testopgavernes faglige kvalitet og antallet af opgaver.

Evaluators har udvalgt følgende fire ud af de seks testfag til ekspertvurderingen; dansk, læsning, matematik, fysik/kemi og engelsk. Dansk, læsning, matematik og fysik/kemi er blandt andet udvalgt, fordi de indgik i reviewpanelets rapport fra 2007, og det derfor kunne styrke analysen i nærværende evaluering at kunne inddrage og sammenholde analysen med observationer og konklusioner herfra. Engelsk er udvalgt som repræsentant for sprogfagene.

Ekspertvurdering 1 er gennemført vha. interview med nøglepersoner og seminarer med faglige eksperter. Interview med nøglepersoner er gennemført som personlige interview og har bidraget med baggrundsviden om de nationale test, ligesom de har bidraget til at kvalificere de efterfølgende seminarer med eksperter.

Der er gennemført et heldagsseminar for hvert af de fire udvalgte testfag. Med afsæt i en af evaluators udviklet skabelon har de faglige eksperter drøftet en række temaer vedrørende indholdet i de nationale test. De faglige eksperter har bl.a. drøftet styrker og svagheder ved fagenes profilområder, opgaveemner og trinmål. Endvidere er kvaliteten af opgaver i opgavebanken blevet drøftet bl.a. i forhold til spørgsmålsformuleringer, distraktorer, illustrationer, layout mv.

Til brug for de faglige seminarer har evaluator udvalgt en række opgaver fra opgavebanken, som fungerer som afsæt for de faglige eksperters drøftelser. I udvælgelsen af opgaver er det prioriteret, at samtlige opgavetyper og opgaveemner er repræsenteret. Der er endvidere opgaver repræsenteret fra samtlige fem sværhedsgrader. Herudover har kriteriet været, at de udvalgte opgaver i et vist omfang skulle afspejle den reelle fordeling af opgavetyper, -niveauer og sværhedsgrader. Det er prioriteret, at der er eksempler på opgavetyper med samme opgaveemne og sværhedsgrad, således eksperterne kan vurdere, hvorvidt sværhedsgraden reelt er den samme. Det er ligeledes prioriteret, at der er eksempler på opgavetyper med samme opgaveemne men med forskellig sværhedsgrad, således det er muligt at vurdere, om der er passende forskel på sværhedsgraderne. Det skal dog bemærkes, at der i matematik og fysik/kemi er mange opgavetyper set i forhold til dansk, læsning og engelsk. For at sikre, at alle disse opgavetyper blev repræsenteret mere end en gang, er der i mindre omfang skelet til den reelle opgavefordeling.

Evaluators har faciliteret seminarer med eksperter og har for hver af de fire udvalgte testfag udarbejdet et internt fagnotat, der sammenfatter de faglige eksperters vurderinger af ovenstående områder. Analyser af de nationale tests indhold baserer sig bl.a. på tværgående analyse af disse fagnotater.

Ekspertvurdering 2

Ekspertvurdering 2 har til formål at belyse fordele og ulemper ved det adaptive princip, der benyttes i de nationale test. Der foretages vurderinger af tekniske, økonomiske og pædagogiske fordele og ulemper.

Ekspertvurdering 2 er baseret på et internationalt litteraturstudie og analyse udarbejdet til nærværende evaluering af professor Peter Allerup, Aarhus Universitet. Herudover benyttes indsamlede data fra spørgeskemaundersøgelser og casestudier. Ekspertvurderingen afrapporteres i selvstændigt notat vedlagt som bilag 4 til evalueringsrapporten. Hovedpointer er inddraget i evalueringsrapporten.

Datakvalitet

Som det fremgår ovenfor forudsætter gennemførelse af ekspertvurdering 1 og 2 indgående teknisk og faglig indsigt. Evaluator har derfor gennemført disse i samarbejde med førende eksperter på pågældende områder. Dette bidrager til at sikre kvaliteten af disse analyser. Ydermere er begge ekspertvurderinger gennemført på baggrund af på forhånd udarbejdede skabeloner, der bidrager til at sikre systematik og gennemsigtighed. Samlet set er det således evaluators vurdering, at datakvaliteten af ekspertvurdering 1 og 2 er god.

Casestudier

Casestudierne har som formål at supplere de bredere og kvantitativt funderede dataindsamlingsaktiviteter (spørgeskemaundersøgelserne og effektanalyserne). Casestudierne udgør den kvalitative forskningsmetode, der skal bidrage til en dybdegående forståelse af den kompleksitet, som kendetegner anvendelsen af nationale test og evalueringskultur. Casestudierne har bidraget med viden til at belyse den pædagogiske anvendelse af de nationale test, samt skolernes evalueringskultur.

Der er i alt gennemført 10 casestudier for at få en bredde på tværs af skolerne. Oprindeligt var der planlagt 12 casestudier, men grundet lockouten af lærerne i perioden fra den 2. til 26. april 2013 blev samtlige planlagte casestudier aflyst og skulle genbookes. Flere skoler faldt fra som følge af lockouten, hvorefter nye skoler skulle rekrutteres. Det lykkedes os at rekruttere nye skoler, men i sidste øjeblik sprang enkelte skoler fra grundet tidsmæssigt pres som følge af lockouten.

Fremgangsmåde

Casestudierne er gennemført over to hele dage på de udvalgte skoler. Over de to dage var der arrangeret fokusgruppinterviews og interviews med elever, lærere, skoleleder, forvaltning og forældre til børn, der havde gennemført de nationale test². Der blev gennemført to interviews med skolelederen, et i starten af besøget og et som afslutning på besøget. Dette var for at kunne forholde dem og få en uddybning af de umiddelbare konklusioner fra casestudiet. Derudover var der i en eller flere klasser klasserumsobservationer af testafviklingen. Tilrettelæggelsen af casestudierne tog hensyn til, at interviews og observationer skulle foregå, når det var hensigtsmæssigt og muligt på den enkelte skole.

Grunden til, at casestudierne strækker sig over to dage, er for at sikre bredde i respondenterne (fx på elever og læreres holdninger, testfag og klassetrin). Samtidig var det vigtigt at evaluator kunne nå at få en fornemmelse af og se indikatorer på, hvordan skolerne anvender de nationale test, primært gennem forskellige klasserumsobservationer. Derudover var det vigtigt for at kunne vurdere data at få forskellige kilder inden for hvert casestudie til at belyse anvendelse af de nationale test og skolens evalueringskultur.

Forud for casestudierne blev der udarbejdet spørgeguides målrettet den enkelte informanttype og interviewform. Der blev ligeledes udarbejdet observationsark til brug for klasserumsobservationerne. Observationsarket indeholdt skalerede afkrydsningsmuligheder i forhold til vurdering af testafviklingen, lærerens rolle heri og elevernes oplevelse af de nationale test. Derudover var der i observationsarket mulighed for at notere kommentarer og refleksioner til vurdering af testafviklingen.

På baggrund af interviews og observationer er der udarbejdet caserapporter for hvert af de 10 casestudier (til intern brug i evalueringsteamet). I caserapporterne er skolens organisering, skoleledelse, evalueringspraksis, anvendelse af de nationale test og informanternes vurdering af effekter af de nationale test beskrevet.

Udvælgelse af forsøg

De 10 casestudier er foretaget på skoler, der er udvalgt, så der sikres dybdegående kvalitative data. For at øge validiteten af data er skolerne udvalgt med henblik på at få spredning i forhold til geografisk placering, kommune, skolestørrelse, evalueringskultur og præstation i de nationale test. Dette for at sikre, at der ikke fremkommer et skævvredet billede af anvendelsen af de nationale test. En oversigt over de 10 casestudier vises nedenfor.

² En skole var for lille til et todages besøg, hvorfor vi valgte at afholde samtlige aktiviteter over en dag. Her indgår samtlige lærer i respondentgruppen.

Tablet 4: Oversigt over casestudierne

Casestudie*	Kommune	Region	Skolestørrelse
Skole 1	Aalborg Kommune	Nordjylland	Lille
Skole 2	Randers Kommune	Midtjylland	Mellem
Skole 3	Horsens Kommune	Midtjylland	Stor
Skole 4	Favrskov Kommune	Midtjylland	Mellem
Skole 5	Holstebro Kommune	Midtjylland	Lille
Skole 6	Næstved Kommune	Sjælland	Stor
Skole 7	Slagelse Kommune	Sjælland	Stor
Skole 8	Køge Kommune	Sjælland	Mellem
Skole 9	Nordfyns Kommune	Syddanmark	Stor
Skole 10	Faaborg-Midtfyn Kommune	Syddanmark	Lille

Note: *Skolernes navne er udeladt af hensyn til anonymitet.

Skolerne er blevet rangeret i forhold til evalueringskultur og præstation i de nationale test for at kunne sikre variation mellem skolerne. Skolerne er rangeret i forhold til evalueringskultur på baggrund af skolelederbesvarelserne, hvorfra et indeks for evalueringskultur blev opstillet på baggrund af skolens anvendelse af de nationale test. Skolernes rangering i forhold til præstation i de nationale test bygger på præstation i dansk, læsning i 8. klasse, da der var en klar sammenhæng mellem præstation og evalueringskultur ved denne test.

Datakvalitet

Kvaliteten af casestudier vurderes typisk ud fra to kriterier. Kriterierne er dækningsgraden (er genstandsfeltet for casen veldefineret og afgrænset, vurderes data at være tilstrækkeligt og dækkende, og er der allokeret nok tid og ressourcer til at gennemføre casestudiet?) og muligheden for alternative perspektiver (har casestudiet involveret en tilsigtet afdækning af rivaliserende perspektiver?).

Der er gennemført 10 casestudier af to dages varighed. De 10 cases er udvalgt, så der sikres dybdegående kvalitativ viden fra forskellige typer af skoler. Der er sikret spredning i relation til de af evaluator 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. På denne baggrund vurderer evaluator, at kvaliteten af data tilvejebragt gennem de gennemførte casestudier er særdeles god.

Forandringsteorier

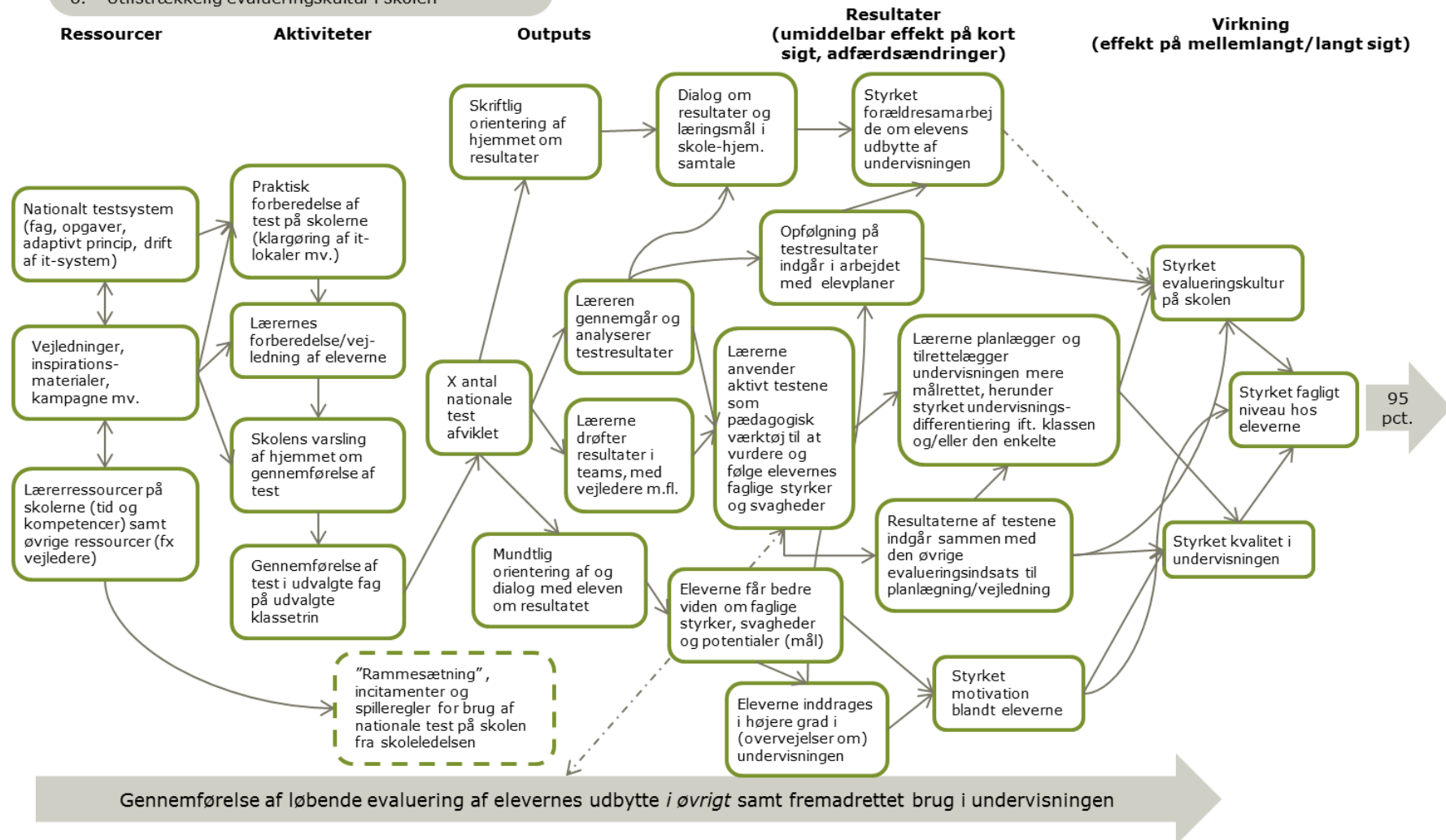
På de følgende to sider præsenteres forandringsteorier, evaluator har udviklet.

- Forandringsteori for indførelse af nationale test i folkeskolen: Lærersporet
- Forandringsteori for indførelse af nationale test i folkeskolen: Ledersporet

Udfordringer/problemer:

1. Elevernes faglige resultater er utilfredsstillende
2. For få gennemfører en ungdomsuddannelse
3. Den løbende evaluering fylder for lidt
4. Utilstrækkelig undervisningsdifferentiering
5. Hjemmet ved for lidt om elevernes faglige niveau
6. Utilstrækkelig evalueringskultur i skolen

Forandringsteori for indførelse af nationale test i folkeskolen ("lærersporet")



Udfordringer/problemer:

1. Elevernes faglige resultater er utilfredsstillende
2. For få gennemfører en ungdomsuddannelse
3. Den løbende evaluering fylder for lidt
4. Utilstrækkelig undervisningsdifferentiering
5. Hjemmet ved for lidt om elevernes faglige niveau
6. Utilstrækkelig evalueringskultur i skolen

Forandringsteori for indførelse af nationale test i folkeskolen ("ledersporet")

