
Teknologiforståelse og adaptive test i skolen

– læreres viden om og anvendelse af de nationale test

Jonas Dreyøe, Ph.d.-stipendiat, Aalborg Universitet, jmd@hum.aau.dk
Andreas Lindenskov Tamborg, Post doc, Ph.d., Københavns Universitet, andreas_tamborg@ind.ku.dk

Sanne Lisborg, Ph.d.-stipendiat, Aalborg Universitet, sali@hum.aau.dk
Jonathan Hermansen, adjunkt, Ph.d., Københavns Professionshøjskole, johe@kp.dk

Resume

Denne artikel omhandler sammenhængen mellem læreres holdninger til og viden om de nationale test og deres anvendelse af testdata. Vi undersøger en tese om, at lærernes viden om resultaterne fra de nationale test er betinget af deres forståelse af den bagvedliggende teknologi udtrykt gennem deres holdning til og anvendelse af testen. Artiklen er baseret på en spørgeskemaundersøgelse af Bundsgaard & Puck (2016), der omhandler lærernes brug af resultaterne fra de nationale test, deres holdninger til testen og deres faktuelle viden om, hvordan den kan anvendes. Bundsgaard & Puck (2016) har lagt deres indsamlede data offentligt tilgængeligt, og det rummer muligheden for at nuancere og udvide deres konklusioner. Spørgeskemaundersøgelsen indeholder besvarelser fra 768 danske folkeskolelærere. På baggrund af datasættet har vi udvalgt en række spørgsmål vedrørende læreres anvendelse af, opfattelse af, viden om og fortolkning af de nationale testresultater. Datasættet giver mulighed for at sammenholde lærernes viden om, anvendelse af og holdninger til de nationale test, hvilket ikke tidligere er blevet undersøgt.

I vores artikel konkluderer vi, at der er sammenhæng mellem lærernes viden om udlægningen af resultaterne fra de nationale test og deres anvendelse af testresultaterne. Lærerne, der udviser større viden om fortolkningen af testresultaterne, er mere tilbøjelige til at undersøge elevernes svar samt sammenholde resultaterne med tidligere år. Omvendt finder vi ikke nogen sammenhæng mellem lærernes holdning til de nationale test og deres viden om udlægningen af testresultaterne. Analysen præsenterer ny viden om sammenhængen mellem viden, forståelse og anvendelse af de nationale test blandt danske lærere på baggrund af regressionsanalyser, der kontrollerer for lærernes erfaring med testen, anciennitet i lærergerningen og undervisningsfag. I diskussionen fremhæves implikationerne ved at introducere et teknologisk komplekst redskab, som de nationale test, som evalueringsværktøj og didaktisk udviklingsredskab i folkeskolen. Artiklen bidrager med at identificere nogle af de mulige konsekvenser, der er ved at anvende de nationale test som udgangspunkt for lærernes formative evalueringspraksis.



Nøgleord: nationale test, spørgeskemaundersøgelse, adaptiv algoritme, data literacy, teknologiforståelse

Abstract

This paper aims to understand the connection between teachers' use of, attitude towards and knowledge about the national test, as well as understanding the implications these connections have for the utilization of the test in Danish schools. We conclude that teachers knowledge of the test results is associated with the teachers' utilization of these results. On the other hand, we do not find any association between teacher attitudes toward the national test and their knowledge of how the results should be interpreted. We lastly discuss the implications that arise by introducing a technologically advanced tool in the Danish school system. We also contribute by identifying some of the possible consequences that emerge by using the national test as a point of departure for teachers' evaluation practices.

Keywords: national test, survey, adaptive algorithm, data literacy, technological literacy

Indledning

Tests af forskellig art er efterhånden blevet en fast bestanddel af de fleste undervisningspraksisser i de vestlige lande (Biesta, 2011). Tests har fået større betydning for ikke bare evaluering, men også tilrettelæggelse og gennemførelse af uddannelse (Kelly, Andreasen, Kousholt, McNess & Ydesen, 2017). Overordnet set er der en række fordele ved denne tendens, såsom at det bliver muligt at få overblik over, hvordan danske skoler klarer sig i forhold til fælles nationale mål for landets skoler, samt at diagnosticere særlige udviklingsområder for elever. Da tests ofte er årligt tilbagevendende, er de brugbare til at måle på effekten af nationale uddannelsesinitiativer (Agrawal, 2004). Samtidig kritiseres testkulturen også for at være forbundet med en lang række alvorlige uhensigtsmæssigheder. Fx påpeger Biesta (2011), at tests bidrager til at underlægge skoler en konkurrencestatslogik, hvor pædagogiske værdier og idealer om dannelse reduceres til statistiske komponenter i kvantitative undersøgelser. Biesta (2011) fremhæver også, at det at score højt i tests ofte ender med at blive et altoverskyggende mål i sig selv, uden der stilles spørgsmålstejn ved, om testene reelt måler det, vi finder værdifuldt. Fænomenet "teaching to the test" er også en hyppig fremført kritik af tests, især fordi testresultater ofte bliver offentliggjort, hvilket kan have stor betydning for både den konkrete skole og den enkelte lærer (Biesta, 2011). Tidligere studier (Jones, Jones & Hargrove 2003; Kelly, Andreasen, Kousholt, McNess & Ydesen, 2017) har fx vist, at policydrevne testkulturer kan lede

lærere til at tilrettelægge undervisning ud fra, hvad de tror, der forventes af dem, og ikke på baggrund af en vurdering af elevernes behov. Dette kan underminere læreres videns- og erfaringsgrundlag og potentielt sænke kvaliteten af undervisningen.

De danske nationale test, som er genstand for denne artikels undersøgelse, har et erklæret mål om ikke blot at kunne generere grundlag for sammenligning af, hvordan danske skoler præsterer, men også at bidrage til at udvikle evalueringskulturen på danske skoler (Undervisningsministeriet, 2019b). Desuden er det et centralt mål for de nationale test, at data skal bruges af lærerne til at øge elevernes læringsudbytte. Anvendelse af data til at udvikle undervisning er en central bevægelse i både dansk og international skolepolitik, der har sine rødder i forskningsfelterne *school effectiveness* og ledelse for læring (Robinson, 2015). En gennemgående tanke i disse felter er, at i takt med, at lærere får erfaring med, hvordan de kan bruge data, vil de blive bedre og mere effektive til at evaluere og udvikle deres undervisningspraksis (Datnow & Park, 2017). Dette fremgår af den seneste skolereform, hvor mål- og effektstyring står stærkt, og hvor resultaterne fra de nationale test bliver et væsentligt værktøj til at udvikle undervisningen (Bjerg & Stau-næs, 2014). I dette felt beskrives lærernes kompetencer til at analysere og bruge data under betegnelsen 'data literacy', der defineres som "the ability to understand and use data effectively to inform decisions ... composed of a specific skill set and knowledge base that enables educators to transform data into information and ultimately into actionable knowledge" (Mandin-ach & Gummer, 2013, s. 30). Der er en bred enighed i forskningslitteraturen om, at der er et behov for at udvikle lærernes data literacy (Klausen, 2018; Dantnow & Park, 2017), men også om, at det er vanskeligt. Tidligere studier har eksempelvis påpeget flere barrierer for at opbygge data literacy i uddannelsessektoren såsom en hyppig lav grad af data self-efficacy hos lærerne (German, 2014) og en manglende grundlæggende viden om data (DeLuca & Bellara, 2013; Piro & Hutchinson, 2014).

Antagelsen, der er udgangspunktet for denne artikel, er, at der er brug for en tilsvarende "nationale test literacy" blandt lærere for at kunne anvende resultaterne af testene på pædagogisk og didaktisk meningsfulde måder. Behovet for såkaldt data literacy er tydelig i den aktuelle uenighed blandt pædagogiske forskere og statistikere om, hvordan og i hvilket omfang resultaterne af nationale test kan og bør anvendes til pædagogiske formål¹.

1 <https://politiken.dk/debat/kroniken/art7168552/Drop-kritikken-af-de-nationale-tests>



Denne pointe er yderligere underbygget af Krejsler (2017), der har vist, at tanken om at basere eller informere pædagogisk arbejde på testresultater er kompleks grundet flere uklarheder om, hvorvidt der er tale om lokal/global og intern/ekstern evidens, samt forholdet mellem evidensbaseret og informering. Det er derfor selvsagt ikke nogen let opgave for landets lærere at skulle navigere i, hvordan og i hvilket omfang de nationale test kan eller bør give anledning til ændrede undervisningspraksisser. Baseret på tilgængelige data fra et tidligere studie (Bundsgaard & Puck, 2016) undersøger denne artikel relationen mellem læreres viden om nationale test og deres anvendelse af resultaterne af testen i deres undervisning. Omdrejningspunktet for denne artikel er således spørgsmålet: *Hvad er sammenhængen mellem læreres forståelse af, holdninger til og anvendelse af testens resultater, og hvilke implikationer medfører disse forhold for brugen af test i skolen?*

Indledningsvis præsenterer vi baggrunden for de nationale test og viser, at hidtidige studier af lærernes anvendelse af og viden om testen når frem til divergerende konklusioner. Vi introducerer det todelte formål med de nationale test: at de både er tiltænkt som et summativt såvel som et formativt evalueringsredskab. Dernæst beskriver vi artiklens empiriske grundlag, der beror på spørgeskemadata indsamlet på baggrund af besvarelser fra danske lærere. Herunder uddyber vi også begrænsningerne ved de anvendte data, og vi begrundet, hvorfor det er relevant at inddrage holdninger, viden og anvendelse i en samlet analyse. I analysen gennemgår vi resultaterne af den anvendte regressionsmodel, hvor vi bl.a. finder en sammenhæng mellem lærernes viden om de nationale test og deres baggrundskarakteristika samt lærernes anvendelse af de nationale test. Afslutningsvis konkluderer vi, at de nationale test øjensynligt har betydning for lærernes praksis, og at der er en sammenhæng mellem lærernes viden om udlægning af testresultaterne og deres anvendelse. Desuden fremhæver vi, at det kan være vanskeligt at forene de to formål, som de nationale test har som ambition.

Baggrunden for de nationale test

Nationale test har siden 2010 været obligatoriske i den danske folkeskole. Indførelsen af nationale test var et resultat af en evalueringsrapport fra OECD (2004), der viste, at skolernes dokumentation af elevernes læring var utilstrækkelig. Elever i den danske folkeskole oplever at bliver testet minimum ti gange i løbet af deres grundskoleforløb. De nationale test er digitale, selvrettende og adaptive, dvs. at de tilpasser sig elevens faglige niveau

(Undervisningsministeriet, 2015). Eleverne bliver testet i tre profilområder i det pågældende fag. I matematik er det eksempelvis tal og algebra, statistik og sandsynlighed samt geometri. Disse profilområder bliver testet skiftevis, indtil usikkerheden for elevens estimerede dygtighed i et af de tre profilområder kommer under den fastsatte grænse. Så snart et af disse profilområder kommer under denne grænse, vil testen koncentrere sig om de resterende områder. Når eleven er under grænsen i alle tre områder, vil lærerens dashboard "lyse grønt" som en angivelse af, at testen med 95 procents sikkerhed har ramt elevens dygtighed. Det adaptive element i testene er centralt for at forstå, hvordan teknologien bag testen fungerer, og hermed hvordan den kan bruges. Testenes adaptive karakter omfatter, at de undervejs tilpasser sig den enkelte elevs dygtighed igennem anvendelse af Rasch-modellen. Testen starter altid med at give eleverne en opgave på middelniveau, hvorefter elevens besvarelse af denne opgave får betydning for sværhedsgraden af den næste opgave. Hvis eleven svarer rigtigt, stilles en sværere opgave. Hvis eleven svarer forkert, stilles en nemmere opgave. Efter tre besvarelser forekommer den første beregning af elevens estimerede dygtighed. Denne bliver efterfølgende beregnet løbende efter hver opgave. Elevens estimerede dygtighed vil være baseret på alle elevens afgivne svar. De spørgsmål, eleverne præsenteres for, vil testen så vidt muligt forsøge at placere så tæt på elevens estimerede dygtighed. Eleverne kan ikke tildeles opgaver i testen, de allerede har besvaret, og hvis elever undlader at besvare opgaver, bliver de håndteret som fejlbesvarelser (Undervisningsministeriet, 2015).

Lærernes brug af nationale test

I den senere tid har danske læreres brug af nationale test fået tiltagende opmærksomhed i både medier og forskning. I det følgende fremhæver vi relevante pointer fra denne litteratur. Schou, Nørgaard, Kvols, Raalskov, & Andersen (2015) har undersøgt, hvilken indflydelse de nationale test har på undervisningen. Et af fundene er, at lærerne mener, at deres egne evalueringsmetoder giver en bedre forståelse af elevernes faglighed end de nationale test. På trods heraf oplever flere, at de må tilrettelægge deres undervisning på baggrund af testen, så de sikrer sig, at eleverne klarer sig godt, det der kaldes 'teaching to the test' (Schou et al., 2015). Et studie har imidlertid identificeret, at danske lærere sammenlignet med engelske lærere i højere grad planlægger og gennemfører undervisning uafhængigt af tests, men ud fra vurderinger af deres elevers behov (Kelly, Andreasen, Kousholt,



McNess, & Ydesen, 2017). Dette studie identificerer således en relativ resiliens blandt danske lærere imod krav forbundet til at bruge testen som udviklings- og evalueringssværktøj. Holm (2015) har i et andet forskningsprojekt specifikt undersøgt, hvordan lærere forholder sig til formidlingsopgaven af testresultaterne samt brugen af resultaterne til at tilrettelægge undervisningen. Han pointerer, at formidlingskravet har skabt pædagogiske dilemmaer for lærerne. I undersøgelsen ses der flere eksempler på "at lærerne ikke tillægger de nationale test validitet, fordi resultatet ikke er i overensstemmelse med lærernes vurdering af elevernes dygtighedsniveau" (Holm, 2015, s. 62). Dette betyder, at der opstår situationer, hvor lærere vælger ikke at formidle testresultaterne til forældre og elever, da de vurderer, at testresultater ikke giver et retvisende billede af elevens niveau. Undersøgelsen finder også, at testresultaterne af flere lærere anses som komplekse og vanskelige at anvende som udgangspunkt for pædagogiske og didaktiske refleksioner i forhold til den enkelte elev, hvilket står i modstrid med intentionen bag de nationale test (Holm, 2015).

I en evaluering foretaget af Rambøll (2013) tegnes der et mere positivt billede af brugen af de nationale test. I evalueringen lyder konklusionen, at de nationale test har været medvirkende til at styrke evalueringskulturen på skolerne, samt at de har en gavnlig effekt på elevernes faglige niveau. Det fremhæves imidlertid som en væsentlig udfordring for lærerne at anvende testresultaterne til at igangsætte pædagogiske og didaktiske indsatser. Bundsgaard & Puck foretog i 2016 en større statistisk undersøgelse af, hvordan lærere anvender de nationale test samt deres viden herom og holdninger til testen (Bundsgaard & Puck, 2016). I lighed med Schou et al. (2015) konkluderer de, at flere af lærerne målretter deres undervisning til testene, inden de gennemføres, dvs. en tendens til at praktisere *teaching to the test*. Bundsgaard og Puck (2016) konkluderer også, at lærerne samlet set ikke opfatter de nationale test som et brugbart redskab til at udvikle deres undervisningspraksis, og de finder, at lærerne ikke besidder den fornødne viden om testen og dens resultater til meningsfuldt at kunne anvende og fortolke testresultaterne. Denne rapport tegner således et langt mere negativt billede af lærernes brug af og opfattelse af de nationale test end Rambøll-rapporten.

Senest har Bundsgaard & Kreiner (2019) publiceret en undersøgelse af de måletekniske aspekter af de nationale test. Her påviser de, at der i flere tilfælde var store forskelle på de sværhedsgrader, som de nationale test benytter, når der udvælges opgaver og beregnes resultater for den enkelte elev. Konsekvensen er, at beregningen af den enkelte elevs testresultat kan være

fejlbehæftet. På den baggrund konkluderer Bundsgaard & Kreiner (2019), at der er så store vanskeligheder med testens måling, at det ikke er forsvarligt at bruge resultaterne som grundlag for evaluering, hverken lokalt på skolerne eller i politisk sammenhæng.

Evaluering og nationale test

I diskussionen om lærernes brug af de nationale test er forståelsen af evaluering central. Forskningsfeltet inden for evaluering er omfattende både nationalt og internationalt, og her vil vi kun berøre de typer af evaluering, der er væsentlige for at forstå centrale aspekter ved de nationale test. Overordnet skelnes der mellem summative og formative formål med at benytte tests. Den summative evalueringstilgang omfatter, at resultaterne af evaluering bliver anvendt til kontrol og dokumentation fx med henblik på sanktion og belønning (Andersen, Wahlgren, & Wandal, 2017). Harlen (2005) betegner dette "summarizing learning", dvs. en form for opsummering af læringsudbyttet. Den formative anvendelse omhandler ønsket om at afdække, hvad den enkelte elev kan, for at kunne forbedre elevens læring og undervisningen generelt. Den formative evaluering indeholder anvisninger af, hvordan afstanden mellem elevens læring og de opstillede standarder kan minimeres (Taras, 2005), hvilket Harlen betegner som "helping for learning" (Harlen, 2005). Distinktionen mellem disse to typer af evaluering ligger hovedsageligt i, hvordan resultaterne af evalueringen anvendes, og ikke i forhold til, hvordan evalueringer er designet. Dog er der forskel på, hvordan evalueringer skal designes, afhængigt af, om de primært er tiltænkt at tjene et summativt eller formativt formål.

Målet med de nationale test er todelt. For det første er det at kunne evaluere på tværs af de danske folkeskoler, hvorfor kommune og skoleledelse får oplyst de enkelte skolers resultater målt i forhold til landsresultatet. Derudover bliver det samlede resultat af de nationale test offentliggjort på landsplan, så den faglige udvikling kan følges politisk jf. de nationale mål for udvikling i folkeskolen (Undervisningsministeriet, 2019a). I denne sammenhæng er der således tale om et summativt formål, hvor elevernes, skolerne og kommunernes præstationer kan sammenlignes over tid. For det andet er testresultaterne tiltænkt at skulle anvendes formativt af den enkelte lærer i evalueringssøjemed: "Testene anvendes sammen med andre elementer som led i den løbende evaluering fremadrettet til brug i vejledningen af den enkelte elev, den videre planlægning af undervisningen og underretning af forældrene med henblik på at tilrettelægge en undervisning og et



forældresamarbejde, der understøtter eleven bedst muligt” (Undervisningsministeriet, 2019a, s. 1). Det er således obligatorisk, at lærerne skal anvende testresultaterne formativt både i samtale med eleven og forældrene (Holm, 2015). Det understreges dog, at testen ikke kan stå alene, men skal bruges som et redskab blandt andre i den samlede vurdering af eleven og klassen. Her nævnes fx observation, samtaler, afleveringer og den løbende evaluering af undervisningen (Undervisningsministeriet, 2019b). Testen er således både udviklet med et summativt og et formativt sigte.

I det følgende vil vi redegøre for vores tilgang til indsamling og bearbejdning af data for at besvare artiklens forskningsspørgsmål.

Metode og data

Denne artikels datagrundlag stammer fra en spørgeskemaundersøgelse, som Bundsgaard & Puck (2016) anvendte som empirisk grundlag for deres rapport om nationale test. Spørgeskemaundersøgelsen er gennemført i første halvår af 2016 og består af to separate spørgeskemaer rettet mod henholdsvis lærere og skoleledere. I vores analyse benytter vi imidlertid kun besvarelser fra lærerne, da denne artikel alene omhandler lærernes viden, holdninger og brug. Det anvendte stikprøvedesign er en kombination af en klyngeudvælgelse og stratificeret udvælgelse, da der ikke foreligger et centralt register, der kunne anvendes til at foretage en simpel udvælgelse af potentielle respondenter (for en udførlig beskrivelse af indsamling og tilrettelæggelse se Puck, 2016). Puck angiver, at stikprøven for lærerdatasættet er repræsentativt på skoleniveau (Puck, 2016, s. 12). De anvendte stratificeringskriterier er skolestørrelse, kommune og region. Såfremt de udvalgte stratificeringskriterier har betydning for det, man ønsker at undersøge, har denne tilgang den fordel, at stikprøven bliver repræsentativ for hele populationen (Groves et al., 2009).

Svarprocenten for spørgeskemaundersøgelsen blandt lærerne er knap 43 procent (Bundsgaard & Puck, 2016). I alt omfatter datasættet for lærerne 768 besvarelser. Denne artikel beror dog alene på besvarelser fra respondenter, der har besvaret alle spørgsmål, der indgår i analysen, hvorfor nettodatasættet består af besvarelser fra i alt 720 lærere.

Spørgeskemaets design

Som tidligere angivet fokuserer vi i denne artikel på sammenhængen mellem lærernes viden om og brug af de nationale test. Denne sammenhæng undersøger vi i relation til lærernes viden om udlægningen af resultaterne af de

ationale test. For at belyse disse forhold har vi dermed behov for data om 1) lærernes viden om de nationale test, 2) lærernes brug af de nationale test samt 3) lærernes karakteristika. I det følgende vil vi beskrive, hvordan lærernes besvarelser af spørgeskemaerne giver os indsigt i disse tre forhold. Det skal dog bemærkes, at den primære variabel i vores analyser er lærernes viden om de nationale test målt som antallet af korrekte svar i en række af vidensspørgsmål om de nationale test.

Måling af lærernes viden om de nationale test

Spørgeskemaet måler lærernes viden om de nationale test på baggrund af en række spørgsmål om de nationale test, som lærerne skal forholde sig til. I spørgeskemalitteraturen betegnes denne type spørgsmål ofte som 'kunds-kabs spørgsmål' (Olsen, 2006), fordi der findes ét korrekt svar på dem. I det følgende vil vi imidlertid omtale dem som 'vidensspørgsmål', da det er denne term, Bundsgaard & Puck (2016) selv anvender.

Vidensspørgsmålene i spørgeskemaet er opdelt i to. I den første del skulle respondenterne forholde sig til et spørgsmål om, hvordan resultaterne fortolkes. Her var angivet fem svarmuligheder, hvor den ene angav, at respondenteren ikke kendte svaret. Kun den ene af de øvrige fire svarmuligheder er korrekt. I den anden del af vidensspørgsmålene skulle lærerne angive, hvorvidt fem udsagn enten er korrekte, forkerte, eller om lærerne ikke kendte svaret. Den anden del af vidensspørgsmålene rummer fem svar fra hver lærer. Vi har på baggrund af besvarelserne opsummeret antallet af korrekte svar, så den enkelte respondent har kunnet få op til seks korrekte svar, hvis de har valgt det korrekte udsagn i første del samt svaret korrekt på de efterfølgende fem udsagn.

Begrænsninger ved måling af lærernes viden om nationale test i spørgeskemaet

I vores arbejde med målingen af lærernes viden om de nationale test er vi blevet opmærksomme på en række mulige begrænsninger i designet af spørgeskemaet, som kan have haft indflydelse på lærernes besvarelser. Disse vil vi her kort redegøre for.

I Bundsgaard & Pucks (2016) spørgeskema er det værd at bemærke, at det eneste af udsagnene, der er korrekt, refereres til en usikkerhed ved fortolkningen af resultaterne, hvorimod de øvrige udsagn udtrykker en vished om fortolkningen. Dette lyder: "Der er en vis usikkerhed på resultatet, så to resultater der er tæt på hinanden kan godt være udtryk for at eleverne har



samme dygtighed" (s. 52). Dette kan være en uhensigtsmæssig tilrettelæggelse af spørgsmålet, da det muligvis kunne lede mere kritisk indstillede lærere til at svare korrekt på spørgsmålet, da de vil være mere villige til at betvivle testresultaternes udsagnskraft (se Holm, 2015). Det er dermed sandsynligt, at det eksplicite mål for spørgsmålet, lærernes viden om udlægningen af testresultater, kunne blive påvirket af deres holdning. Eksempelvis lyder et af de øvrige udsagn: "En elev der scorer 50 er altid cirka dobbelt så dygtig som en elev der scorer 25" (Bundsgaard & Puck, 2016, s. 51), hvor det korrekte svar er, at udsagnet er forkert.

Herudover fremgår det af brødteksten forud for vidensspørgsmålene, at respondenterne opfordres til "ikke at gætte. Svar hellere 'Jeg kender ikke svaret.'" (Ibid.) Ordlyden afspejler en vanskelig balancegang, hvor man på den ene side ønsker at få informerede svar, men omvendt kan afholde respondenter fra at svare. Det er en fordel, at respondenterne har mulighed for at angive, at de ikke kender svaret (Sturgis, Allum, & Smith, 2007), men den meget eksplicite betoning i spørgsmålsordlyden kan have haft en indvirkning på respondenternes svar (Sanchez & Morchio, 1992). Dette kan muligvis have øget andelen, der har svaret, at de ikke kender svaret.

Det kan også påpeges, at der ikke er enighed om, hvorvidt der findes entydige korrekte/forkerte svar på de vidensspørgsmål, der indgår i Bundsgaard & Pucks spørgeskema. Wahlgren, Andersen, & Wandall (2017) påpeger, at udlægningen af spørgsmålet: "Man kan meningsfuldt sammenligne resultater for to årgange for den samme elev (i matematik og dansk læsning)" er sandt, hvorimod forfatterne af rapporten udlægger dette som falsk. De nationale test er udarbejdet på baggrund af Rasch-modellen, som gør det muligt at placere alle målingerne på samme intervallskala på tværs af klassetrin. Det betyder ifølge Wahlgren, Andersen, & Wandall (2017), at det er muligt at måle progressionen på individniveau på tværs af klassetrin.

Undersøgelse af lærernes holdninger og anvendelse

Vi har valgt at inddrage spørgsmål om lærernes holdninger til de nationale test, da tidligere studier viser, at dette kan have betydning for deres udlægning og anvendelse af dem (Holm, 2015). For så vidt angår lærernes holdning til nationale test har vi genskabt resultaterne fra Puck (2016), som inddrager 21 spørgsmål, som respondenterne har kunnet erklære sig enige eller uenige i på en sekspunktsskala. De 21 spørgsmål er præsenteret i to batterier i spørgeskemaet, hvor der spørges ind til en række forhold, såsom om de er enige i udsagnet, at de nationale test "Betyder meget for forvaltningen i min kom-

mune" og "Giver et indblik i den enkelte elevs niveau". Vi har i lighed med Puck (2016) valgt at opsummere svarene på de 21 spørgsmål i færre variable på baggrund af en faktoranalyse.

Faktoranalysen er udført på samme måde som Puck (2016, s. 31-55), hvilket indebærer, at beregningen er sket som en *principal component*-faktoranalyse med Kaiser-normalisering af *varimax rotation*. Disse beregninger af spørgsmålsbatterierne tager afsæt i, at der ligger en eller flere bagvedliggende faktorer, der forklarer lærernes holdning til nationale test og brugen af testresultaterne. Vi har dog valgt kun at inddrage tre af de fire dimensioner, som Puck udleder af faktoranalysen. Dette sker af to hensyn. For det første er det kun de tre første dimensioner, der har eigenverdi på over 1, hvilket er den gængse grænseværdi for udvælgelse af antallet af dimensioner (Kline, 1994). For det andet er det vores vurdering, at de tre faktorer er lettere at fortolke, da overlappet mellem dimensionerne er mindre. Vi har valgt at anvende de samme betegnelser til hver af de tre faktorer som Puck (2016), da de grundlæggende måler det samme. Dvs. at faktorer gennemgående bliver forklaret af de samme items. Den første faktor kalder vi "Anvendeligt redskab", der dækker over, at denne faktor primært måler de positive egenskaber, som respondenterne tillægger nationale test. Tilsvarende betegner vi den anden faktor "Betydningsfuldt for aktører", da denne siger noget om, hvorvidt respondenterne oplever, at centrale aktører (såsom elever, forældre og kolleger) tillægger nationale test stor betydning. Den sidste faktor benævner vi "Uheldig påvirkning", da denne bl.a. omfatter, at lærerne mener, at testene har flyttet faget i en uheldig retning, og de er uenige i, at testen er meningsfuld for eleverne. Denne faktor rummer også, at respondenterne mener, at forvaltningen tillægger nationale test stor betydning.

Undersøgelsen af lærernes anvendelse af testen er baseret på deres besvarelse af, hvilke måder de har valgt at anvende resultaterne. Alle syv spørgsmål vedrørende lærernes brug af de nationale test er præsenteret i spørgeskemaet som et samlet batteri, hvor respondenterne er blevet præsenteret for en ensartet sekspunktsskala, der omhandler omfanget af anvendelsen af testresultaterne i arbejdet med eleverne. Svarskalaens yderpunkter udgøres af hhv. kategorien "Slet ikke" og "For alle elever". Det er således antallet af elever, som lærerne bruger testresultaterne i forhold til, der er sigtet for spørgsmålet om anvendelse.

De syv spørgsmål om anvendelse af de nationale testresultater er ligeledes opsummeret i færre variable ved hjælp af en faktoranalyse. Denne analyse er ligeledes foretaget på samme vis som Puck (2016). Med afsæt i



eigenværdi på mindst 1 som grænse har vi udledt to faktorer af analysen. Vi har valgt de samme betegnelser for de to faktorer som Puck (2016): "Følger op på resultater" og "Undersøgende tiltag til resultater". Fælles for de to faktorer er, at de begge er handlingsorienterede, da lærerne i begge tilfælde *gør noget* på baggrund af elevernes resultater i nationale test. Faktoren "Følger op på resultater" dækker over en praksis, hvor lærerne undersøger elevernes konkrete svar, gør en særlig indsats på baggrund af resultatet og giver eleverne flere opgaver inden for de områder, eleverne er blevet testet i.

"Undersøgende tiltag til resultater" angiver, hvorvidt lærerne anvender testens resultat, blandt andet til at sammenligne med elevernes tidligere resultater og til at teste eleverne yderligere for at sammenligne dette med resultaterne i de nationale test. Dette mener vi opsummeres bedst af, at lærerne er orienteret mod at anvende de nationale test som et diagnostisk redskab, der kan afdække, hvorvidt eleverne har lært det, de skal, eller en mere præcis diagnose af, hvilke vanskeligheder en elev kan have.

Undersøgelse af lærernes karakteristika

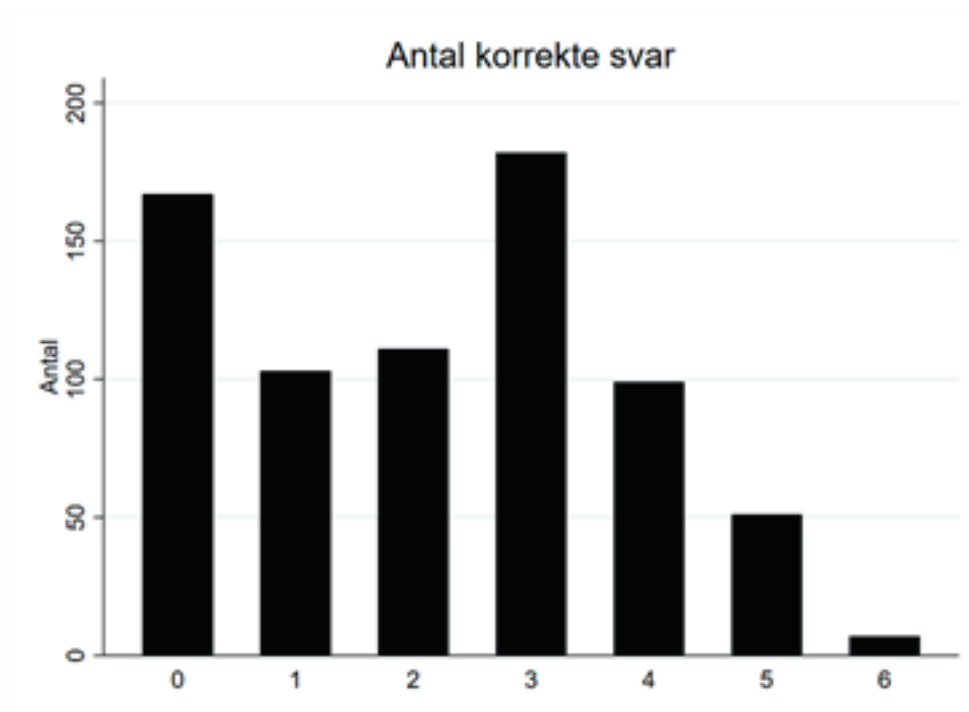
Det, vi i denne undersøgelse omtaler som lærernes karakteristika, dækker over lærernes undervisningsfag, anciennitet samt deres erfaring med at afvikle nationale test med klasser.

Undervisningsfag dækker over, hvilket fag den pågældende lærer primært underviser i. For overblikkets skyld og for at tage højde for, at nogle af fagene kun dækker over et beskedent antal besvarelser, har vi inddelt fagene i tre overordnede kategorier: humanistiske fag, naturfag og praktiske fag. Denne inddeling sker på baggrund af den overordnede kategorisering, som Undervisningsministeriet læringsportal (EMU) anvender (Undervisningsministeriet, 2018).

Anciennitet afspejler antallet af år, som de adspurgte lærere har undervist. I spørgeskemaet udarbejdet af Bundsgaard & Puck (2016) er respondenterne blevet spurgt til, hvor mange år de har undervist, hvor de er blevet præsenteret for en skala med femårsintervaller (og et enkelt seksårsinterval) som svarkategorier. Spørgsmålet indeholder i alt ti svarkategorier. Testerfaring er målt som antallet af gange, som den pågældende lærer har haft en klasse til nationale test. Denne variabel antager værdierne 0-10 samt tillægskategorien "flere end 10". Spørgsmålet indeholder altså i alt 12 svarkategorier. For overblikkets skyld vælger vi i den følgende analyse at behandle de to variable som intervallskalerede variable, da kategorierne afspejler et kontinuum, selv om afstandene mellem kategorierne ikke er eksakte.

Statistisk analyse og resultater

Denne artikels bidrag består i at undersøge danske læreres viden om udlægning af nationale testresultater set i lyset af deres holdninger til og anvendelse af selvsamme tests. For at undersøge det enkelte forholds betydning set i lyset af de øvrige er det nødvendigt at foretage en multivariat analyse. I en sådan analyse er det muligt at tage højde for, at eksempelvis lærernes erfaring med nationale test kan have betydning for sammenhængen mellem holdninger og viden om testresultaterne. Regressionsanalyser er en udbredt metode til at foretage denne type af multivariate analyser, og det er denne tilgang, vi har valgt at benytte. Den afhængige variabel i denne analyse er antallet af korrekte svar. Denne variabel har egenskab af at være en ikke-negativ tællevariabel, der kan antage værdierne 0 til 6. Derfor er det ikke oplagt at anvende en traditionel lineær regressionsmodel til at estimere sammenhængene. Figuren nedenfor viser fordelingen af korrekte svar blandt respondenterne i analysen:



Figur 1. Fordelingen af antal korrekte svar

Som Figur 1 illustrerer, er der flest, der har hhv. nul og tre korrekte svar, hvorimod kun en mindre andel af lærerne har svaret korrekt på alle seks vidensspørgsmål. En Poisson-regressionsmodel er formentlig den mest anvendte

type af model, når man har en afhængig tællevariabel. Vi har imidlertid valgt en negativ binomial regressionsmodel til at estimere sammenhængen mellem antallet af korrekte svar og de øvrige forklarende forhold. I forhold til en Poissonmodel har en negativ binomial regressionsmodel den fordel, at den ikke antager, at middelværdi og varians for den afhængige variabel skal være den samme (Long & Freese, 2014). Vi har testet denne antagelse ved hjælp af en likelihood ratio test ($p=0,026$), hvilket understøtter brugen af en negativ binomial regressionsmodel frem for en Poissonmodel. Regressionsanalysen er foretaget med brug af robuste standardfejl (Long & Freese, 2014, s. 512).

I tillæg til variabelen vedrørende antallet af korrekte svar indeholder analysen tre forklarende variable om holdninger til de nationale test og to variable om brugen af testresultaterne. De fem variable er dem, der er udledt på baggrund af faktoranalyserne. Hertil kommer de tre forhold om henholdsvis undervisningsfag, anciennitet og erfaring med de nationale test. Tabellen nedenfor indeholder en deskriptiv opgørelse af alle de anvendte variable i vores analyse:

	n	Gennemsnit	Std. Afvigelse	Min.	Maks.
Antal korrekte svar	720	2,17	1,63	0,00	6,00
Anvendeligt redskab	720	0,00	1,00	-1,93	3,49
For betydningsfuldt for aktører	720	-0,01	1,01	-2,50	2,93
Uheldig påvirkning	720	0,01	1,00	-3,64	2,21
"Kollektiv brug"	720	-0,01	1,00	-1,20	3,34
"Individuel brug"	720	0,01	0,99	-3,75	1,65
Naturfag	720	0,28	0,45	0,00	1,00
Humanistisk fag	720	0,68	0,47	0,00	1,00
Praktisk fag	720	0,04	0,19	0,00	1,00
Anciennitet	720	4,04	2,18	1,00	9,00
Testerfaring	720	5,82	3,10	0,00	11,00

Tabel 1. Beskrivende statistik

Som tabellen viser, indgår der 720 respondenter i analysen, idet personer, der ikke har besvaret samtlige spørgsmål, som analysen inkluderer, sorteres fra. Det fremgår af tabellen, at respondenterne i gennemsnit har 2,17 korrekte svar.

Herudover kan man se fordelingen af respondenterne i forhold til undervisningsfag: naturfag (ca. 28 procent), humanistiske fag (ca. 68 procent) og praktisk fag (ca. 4 procent). I gennemsnit har respondenterne undervist i ca. 16-20 år svarende til den fjerde kategori for anciennitet. Ingen har angivet at have undervist over 46 år, der er den sidste og tiende svarkategori for anciennitet. I gennemsnit har underviserne haft knap seks klasser til nationale test.

	Koefficient (Standardfejl)
Anvendeligt redskab	-0,001 (-0,033)
For betydningsfuldt for aktører	-0,026 (0,028)
Uheldig påvirkning	0,032 (0,030)
Følger op på resultater	-0,011 (0,033)
Undersøgende tiltag til resultater	0,093 (0,032)**
Humanistisk fag (ref. naturfag)	-0,174 (0,065)**
Praktisk fag (ref. naturfag)	0,067 (0,127)
Anciennitet	0,027 (0,014)*
Testerfaring	0,038 (0,009)***
Konstant	0,541 (0,093)***

n=720, *p<0,05, **p<0,01, ***p<0,001, LR-test af samlet modelfit p<0,001, pseudo r² = 0,02

Tabel 2. Negativ binomial regressionsmodel af antal korrekte svar

Regressionsanalysen af ovenstående model viser, at der er en del af de forklarende variable, der er statistisk signifikante. Nedenfor redegør vi for disse signifikante forklarende variable.

For det første viser Tabel 2, at respondenterne, der anvender testen som et diagnostisk redskab ("Undersøgende tiltag til resultater"), gennemsnitligt har flere korrekte svar end den øvrige population. Dette resultat er ikke umiddelbart overraskende, da det understøtter, at jo mere man som lærer forstår, hvordan de nationale test måler, og hvordan man kan sammenligne resultaterne, jo mere bruger man dem som lærer.

For det andet viser tabellen, at lærere, der underviser i humanistiske fag, gennemsnitligt har færre korrekte svar end de, der underviser i naturfag. Dette resultat er mere overraskende end det første resultat, da det viser, at lærere, der underviser i de humanistiske fag, har mindre viden om, hvordan de nationale test måler, end de naturfaglige lærere. Dette kan dog hænge



sammen med den uddannelsesmæssige baggrund, naturfaglige lærere har, da de nationale test er en kompleks matematisk teknologi.

For det tredje viser tabellen, at der er statistisk signifikant sammenhæng mellem anciennitet og antallet af korrekte svar. Således har lærerne med flere års erfaring som lærere gennemsnitligt betragtet flere korrekte svar end deres lærerkollegaer med færre års erfaring. Dette resultat kunne indikere, at kendskabet til den korrekte udlægning af de nationale testresultater kunne hænge sammen med andre former viden og kompetencer, der opnås med erfaring som lærer.

For det fjerde viser tabellen, at antallet af korrekte svar og lærernes testerfaring er positivt korreleret, således at respondenterne, der har haft flere klasser til nationale test, gennemsnitligt har flere korrekte svar. Der er ikke umiddelbart nogen god forklaring i dataen på dette fænomen, men en sandsynlig forklaring er, at jo oftere lærere bruger de nationale test, jo mere sætter de sig ind i dem, deres udsigelseskraft, og hvordan resultater kan anvendes.

For det femte viser tabellen, at der ikke er statistisk signifikant sammenhæng mellem lærernes holdninger til de nationale test og antallet af korrekte svar. Dette resultat er umiddelbart et interessant resultat, da det viser, at læreres viden om de nationale test ikke er påvirket af deres holdninger til testen. Det betyder, at man altså ikke kan få lærere til at vide mere om testen ved at få dem til at holde mere af den.

Eftersom holdninger og anvendelse kun har begrænset betydning for lærernes viden om udlægningen af testresultaterne, er det interessant at forsøge at sætte nogle tal på, hvor stor betydning de øvrige lærerkaraktistika har for det forventede antal korrekte svar på vidensspørgsmålene. Tabellen nedenunder indeholder betydning opgjort i procent:

	Procent
Humanistisk fag (ref. naturfag)	-16,0**
Anciennitet	2,8*
Testerfaring	3,8***

n=720, *p<0,05, **p<0,01, ***p<0,001,

Tabel 3. Procentuel ændring i forventet antal korrekte svar

Tabel 3 viser, at lærere, der underviser i humanistiske fag i skolen, har et forventet antal af korrekte svar, der er 16 procent lavere end naturfagslærere. Tilsvarende viser tabellen, at for hvert anciennitetstrin (ca. 5 år), en lærer har undervist, øges det forventede antal af korrekte svar med 2,8 procent. Afslutningsvis viser tabellen, at for hver gang en lærer har haft en klasse til nationale test, øges det forventede antal af korrekte svar med 3,8 procent. Alle tallene i tabellen er beregnet således, at de øvrige variable i analysen holdes konstant.

Sammenfattende om resultaterne kan det siges, at de peger på flere vigtige elementer, når man skal undersøge, hvordan lærernes viden om nationale test hænger sammen med deres brug af og holdning til testen. Først og fremmest viser analysen, at en større viden om fortolkningen af testresultater hænger sammen med en øget anvendelse af den praksis, som vi betegner "Undersøgende tiltag til resultater". Dette indikerer, at der kan være en positiv sammenhæng mellem viden og anvendelse. Analysen illustrerer også, at erfaring som lærer og erfaring med testen hænger sammen med en større viden om udlægningen af resultaterne fra de nationale test. Resultaterne fortæller dog ikke noget entydigt om, hvorvidt en øget testfrekvens ville give danske lærere mere viden om den pågældende testform. Ydermere viser analysen, at lærerne i de humanistiske fag svarer korrekt på færre af vidensspørgsmålene, når deres resultater sammenlignes med lærerne i de naturfaglige fag. Afslutningsvis er det værd at hæfte sig ved, at der tilsyneladende ikke umiddelbart er nogen sammenhæng mellem lærernes holdning til testen og deres viden om udlægningen af de nationale testresultater.

Diskussion og konklusion

Analysens resultater giver anledning til flere spørgsmål og implikationer. Et af de bemærkelsesværdige resultater af de kvantitative analyser beskrevet i denne artikel er, at selvom der er forskelle på, hvordan lærerne anvender resultaterne af testene, indeholder begge faktorer i vores analyser (hhv. "Undersøgende tiltag på resultater" og "Følger op på resultater"), at lærerne gør *noget* med eller i deres undervisning på baggrund af testresultaterne uanset omfanget af deres viden om dem. Det betyder, at en gruppe af de adspurgte lærere ændrer deres undervisning på baggrund af informationer, hvis beskaffenhed de reelt ikke kender eller forstår. I det følgende afsnit vil vi fokusere på at beskrive nogle mulige forklaringer på dette fænomen ved hjælp af Biestas (2011) analyser af aktuelle tendenser om at måle og sam-



menlige elevers præstationer. Vi har valgt at inddrage Biestas (2011) pointer her, fordi de kan hjælpe os til at forstå mulige motivationer og årsager til, at lærere ændrer deres praksisser uanset deres viden om testresultaternes udsigelseskraft. Vi vil afslutte diskussionen med at give bud på, hvordan disse fænomener kan imødegås og håndteres.

Set i en bredere uddannelsespolitisk kontekst kan de danske nationale test betragtes som en del af en større trend rettet mod at måle og sammenligne elevers præstationer i uddannelsessystemet. Eksempler på denne trend ses i både TIMMS, PIRLS og PISA, der er flittigt anvendt verden over. Ifølge Biesta (2011) er størstedelen af disse test sat i verden for at måle og sammenligne lande med hinanden med henblik på at få overblik over, hvem der klarer sig bedst. Formålet med de danske nationale test er imidlertid *også* at medvirke til at skabe evalueringskulturer på landets skoler, og det beskrives, at de nationale test ikke skal stå alene (Undervisningsministeriet, 2019a). Ifølge Biesta (2011) er disse test og målinger underlagt et rationale om, at skolers præstationer skal være transparente, så det er lettere for forældre at vælge den bedste skole, hvilket kan skabe et konkurrencemæssigt grundlag for, at skoler og lærere gør, hvad der står i deres magt for at forbedre undervisningskvaliteten.

Biesta fremhæver, at det øgede fokus på tests har medført, at lærerene kan afregnes. Han anvender begrebet "accountable" for kvaliteten af deres undervisning, snarere end at de er ansvarlige ("responsible") for undervisningens kvalitet. For Biesta er forskellen, at førstnævnte er forbundet med at aflægge rapport og regnskaber, der kan underkastes revision, hvorimod sidstnævnte især omhandler at stå til ansvar for ens handlinger. Ifølge Biesta er *accountability*-begrebet gledet fra den finansielle sektor ind i den offentlige sektor, hvor det fungerer som ledelsesprincip. Her peger Biesta på, at en *accountability*-kultur ofte er præget af, at "praksisser må tilpasse sig revisionsprocessens principper" (Biesta, 2011, s. 63). Set i dette lys underlægges læreres praksis revision gennem de nationale test, hvor elevernes præstationer (dvs. kvaliteten af lærernes undervisning) gøres til genstand for revision. Hvis nogle elever klarer sig mindre godt, er det i denne optik oplagt, at den pågældende lærer må genoverveje tilrettelæggelsen af sin undervisning. De nationale test kan i den forstand forstås som en revision af læreres arbejde, hvor det er let at forestille sig, at lærere vil være tilbøjelige til (eller føle sig forpligtet til) at ændre deres undervisning, hvis blot få elever klarer sig mindre godt. I en sådan situation vil lærerne både stå til regnskab over for skoleledelse, forvaltning og forældre, hvilket kan tænkes at skabe et pres

for at demonstrere og dokumentere, at de ikke blot er fortsat som vanligt. Sagt med andre ord kan testresultater gøre det vanskeligt for lærere at fortsætte ufortrødent med at undervise, som de har gjort hidtil, desto mindre alle de testede elever scorer ganske højt i testen. Da det imidlertid ikke ligger lige for at afgøre testresultaternes gyldighed, samt i hvilken udstrækning og ikke mindst hvordan de kan informere pædagogiske tiltag, placeres man let som lærer i et dilemma mellem en forpligtigelse til at gøre *noget* og at gøre noget *velbegrundet*.

Der er naturligvis flere veje ud af dette dilemma. Forudsat at de nationale test vedbliver med at eksistere, er en af flere mulige udveje, at lærere bliver udstyret med tilstrækkelig viden om testene til at kunne foretage meningsfulde pædagogiske ændringer. Vi vil her beskæftige os med en basal forudsætning for dette, nemlig at tilnærme os en bestemmelse af, hvad lærere bør vide for at kunne anvende de nationale tests resultater på fornuftige måder, velvidende, at det er vanskeligt at give et udtømmende svar på dette spørgsmål. Først og fremmest vurderer vi, at det er vigtigt at vide, at de nationale test er udviklet med to parallelle formål. De skal på den ene side udgøre grundlaget for sammenligningen af skoler og på den anden side samtidigt understøtte aktive evalueringskulturer på skolerne. Det første af disse to formål har et tydeligt summativt sigte, da en sammenligning forudsætter et mål for, hvordan den enkelte skole klarer sig. Det andet formål må imidlertid formodes ikke blot at være rettet mod at sikre summative evalueringskulturer, men derimod evalueringskulturer, der fremmer læring, dvs. formative kulturer. Det kan dog i praksis være vanskeligt at forene de to intentioner. Test-items i en summativ evaluering har til formål så præcist som muligt at *måle* læring og ikke *skabe* læring. Dette sigte kan føre til kuriøse spørgsmål, som eksempelvis orddelingsspørgsmål i de nationale test i dansk, der tidligere har været genstand for diskussion. At stille 2. klasser til opgave at foretage en orddeling af *vennervenerationerne* kan givetvis give indblik i elevens ordforråd og læseniveau, fordi egenskaber i elevers læsefærdigheder isoleres, men det er vanskeligere at se, hvordan en sådan opgave er meningsfuld i undervisningsøjemed. Den summative evaluering kan potentielt have en læringsmæssig slagside, da *teaching to the test* samt *backwash*-effekten kan forekomme (Yeap, 2010). Det medfører, at lærerne kan føle sig nødsaget til at undervise i de elementer, som eleverne skal testes i, selvom de vurderer, at det ikke indeholder et stort læringspotentiale. Gode summative test egner sig godt til at sammenligne skoler, mens de derimod ikke nødvendigvis er velegnede til at understøtte formative formål (William & Black, 1996). Det



modsatte gør sig gældende med formative evalueringer, der er velegnede til at støtte læreres pædagogiske og didaktiske arbejde, mens de ikke genererer resultater, der kan skabe et meningsfuldt sammenligningsgrundlag. Dette peger dermed på, at Undervisningsministeriets to ønsker om at tilgodese summative og formative formål i de samme tests er modstridende, eller i hvert fald som minimum kræver grundige overvejelser over, hvordan testen designes, så der opnås balance mellem at generere sammenligningsgrundlag uden at kompromittere lærernes muligheder for at anvende testene med pædagogiske formål. Dette ville imidlertid forudsætte, at lærerne besidder en grundlæggende forståelse af, hvordan resultaterne fra de nationale test skal udlægges. Denne ambition ville indebære et omfattende kendskab til statistiske metoder og adaptive testformer, hvilket hidtil ikke har været et særskilt prioriteret mål fra politisk hold.

Referencer

- Andersen, M., Wahlgren, B., & Wandall, J. (2017). *Evaluering af læring, undervisning og uddannelse*. København: Hans Reitzels Forlag.
- Agrawal, M. (2004). Curricular reform in schools: the importance of evaluation. *Journal of Curriculum Studies* 36(3), 361-379.
- Biesta, G.J.J. (2011). *God uddannelse i målingens tidsalder: Etik, politik, demokrati*. Århus: Klim.
- Bjerg, H., & Staunæs, D. (2014). *Læringscentreret skoleledelse – tænketeknologier til forskningsinformeret skoleledelse*. Frederikshavn: Dafolo.
- Bundsgaard, J., & Kreiner, S. (2019). *Undersøgelse af De Nationale Tests måleegenskaber*. DPU, Aarhus Universitet.
- Bundsgaard, J., & Puck, M.R. (2016). *Nationale test – danske lærere og skolelederes brug, holdninger og viden*. DPU, Aarhus Universitet.
- Datnow, A., & Park, V. (2017). *Datainformeret ledelse*. Frederikshavn: Dafolo.
- DeLuca, C., & Bellara, A. (2013). The current state of assessment education: Aligning policy, standards, and teacher education curriculum. *Journal of Teacher Education* 64, 356–372. <http://dx.doi.org/10.1177/0022487113488144>
- German, J. (2014). *Teachers' perceptions of self-efficacy: The impact of teacher value-added*. Ashland, Ohio: Ashland University. Doktorafhandling.
- Groves, R.M., Fowler Jr., F.J., Couper, M.P., Lepkowski, J.M., Singer, E., & Tourangeau, R. (2009). *Survey methodology*. Hoboken, New Jersey: John Wiley & Sons.
- Harlen, W. (2005). Teachers' summative practices and assessment for learning – tensions and synergies, *The Curriculum Journal* 16(2), 207-223. Lokaliseret den 16. oktober 2019 på: DOI: 10.1080/09585170500136093
- Holm, L. (2015). Læreropfattelser og praksis i relation til de nationale test i dansk/læsning. *CEPRA-Striben* 18, 58-63. Lokaliseret den 16. oktober 2019 på: <https://doi.org/10.17896/UCN.cepra.n18.146>
- Jones, M.G., Jones, B.D., & Hargrove, T.Y. (2003). *The unintended consequences of high-stakes testing*. Lanham, Maryland: Rowman & Littlefield Publishers.
- Kelly, P., Andreasen, K.E., Kousholt, K., McNess, E., & Ydesen, C. (2018). Education governance and standardised tests in Denmark and England. *Journal of Education Policy* 33(6), 739-758.

- Klausen, M.S. (2018). *Skoleledelse med data – i praksis*. Frederiksberg: Samfundslitteratur.
- Kline, P. (1994). *An easy guide to factor analysis*. London: Routledge.
- Krejsler, J.B. (2017). Capturing the 'Evidence' and 'What Works' Agenda in Education: A Truth Regime and the Art of Manoeuvring Floating Signifiers. I: Eryaman, M.Y., & Schneider, B. (red.). *Evidence and Public Good in Educational Policy, Research and Practice*, p. 21-41. Cham: Springer.
- Long, J.S., & Freese, J. (2014) *Regression models for categorical dependent variables using Stata*. College Station, Texas: Stata Press.
- Mandinach, E.B., & Gummer, E.S. (2013). A Systemic View of Implementing Data Literacy in Educator Preparation. *Educational Researcher*, 42(1), 30-37. <https://doi.org/10.3102/0013189X12459803>
- OECD, (2004). OECD-rapport om grundskolen i Danmark. Uddannelsesstyrelsens temahæfteserie nr. 5 – 2004.
- Olsen, H. (2006). *Guide til gode spørgeskemaer* (Vol. 1). København: Socialforskningsinstituttet.
- Piro, J., & Hutchinson, C. (2014). Using a Data Chat to teach instructional interventions: Student perceptions of data literacy in an assessment course. *The New Educator* 10, 95-111. Lokaliseret den 16. oktober 2019 på: <http://dx.doi.org/10.1080/1547688X.2014.898479>
- Puck, M.R. (2016). *Baggrundsrapport om indsamling af spørgeskemadata til undersøgelse af lærernes og ledernes viden om, holdning til, og brug af de nationale test*. UC Viden.
- Robinson, V. (2015). *Elevcentreret skoleledelse*. Frederikshavn: Dafolo.
- Rambøll (2013). *Evaluering af de nationale test i folkeskolen*. Rapport.
- Sanchez, M.E., & Morchio, G. (1992). Probing "dont know" answers: Effects on survey estimates and variable relationships. *Public Opinion Quarterly* 56(4), 454-474.
- Schou, L.R., Nørgaard, L., Kvols, A. M., Raalskov, A.M.H., & Andersen, M. (2015). Når nationale test tager livtag med folkeskolen. *CEPRA-Striben* 18, 38-45. Lokaliseret den 16. oktober 2019 på: <https://doi.org/10.17896/UCN.cepra.n18.144>
- Sturgis, P., Allum, N., & Smith, P. (2007). An experiment on the measurement of political knowledge in surveys. *Public Opinion Quarterly* 72(1), 90-102.
- Taras, M. (2005). Assessment – summative and formative – some theoretical reflections. *British Journal of Educational Studies* 53(4), 466-478. Lokaliseret den 16. Oktober 2019 på: DOI: 10.1111/j.1467-8527.2005.00307.x
- Undervisningsministeriet (2015). *Den adaptive algoritme i de nationale test*. Lokaliseret den 29. maj 2019 på: <https://uvm.dk/-/media/filer/uvm/udd/folke/pdf17/jan/170110-den-adaptive-algoritme-i-de-nationale-test.pdf>
- Undervisningsministeriet (2018). *Lærere og pædagogisk personale*. Lokaliseret den 29. maj 2019 på: <https://arkiv.emu.dk/omraade/gsk-1%C3%A6rer/>
- Undervisningsministeriet (2019a). *Notat om de nationale test*. Lokaliseret den 29. maj 2019 på: <https://www.uvm.dk/-/media/filer/uvm/aktuelt/pdf-19/190412-notat-om-de-nationale-test.pdf?la=da>
- Undervisningsministeriet (2019b). *Om de nationale test*. Lokaliseret den 29. maj 2019 på: <https://uvm.dk/folkeskolen/elevplaner-nationale-test-trivselsmaaling-og-sprogproever/nationale-test/om-de-nationale-test>
- Wiliam, D., & Black, P. (1996). Meanings and Consequences: A Basis for Distinguishing Formative and Summative Functions of Assessment? *British Educational Research Journal* 22(5), 537-548.
- Yeap, B.H. (2010). Improving mathematical thinking through assessment. *Journal of Science and Mathematics Education in Southeast Asia* 33(2), 187-197.