WISEflow
Europe/Copenhagen(CEST)
30 May 2017

UCC Professionshøjskolen

Bachelor LU CLE13o-05BA2

Prædefineret information

Startdato:	11-05-2017 09:00	**Termin:**	jun 2017
Slutdato:	02-06-2017 12:00	**Bedømmelsesform:**	Dansk 7-trinsskala
Eksamensform:	Produkt, fremlægning + forsvar		
SIS-kode:	280717||0617||CLE13o-05BA2||73999||B||PRO||BA||7TRIN		
Intern bedømmer:	Pierre O. Johnsen		
Intern bedømmer:	Annette Zachariassen		

Deltager

Navn:	Sabrina Ferguen
Kandidatnr.:	171675|430734|jun 2017|50345|4273|
UCC-id:	z130030@ucc.dk

Information fra deltager

Titel *:	Forståelser og misforståelser - om de nationale test
Engelsk titel *:	Understandings and misunderstandings - about the national tests
Tro og love erklæring *:	Ja

Gruppe

Gruppenavn:	Sabrina og Dina
Gruppenummer:	2
Øvrige medlemmer:	Dina le Maire Wandall

Forståelser og misforståelser

– om de nationale test

Indholdsfortegnelse

Indledning **3**

Problemformulering *4*

Hypoteser *4*

Empiri, metode og læsevejledning **5**

Evaluering **8**

Formativ/summativ *8*

Bagom DNT **10**

Det adaptive princip *10*

Normbaserede, kriteriebaserede og ipsative referencer af resultater *11*

Er resultaterne retvisende? *14*

Statistisk usikkerhed 14

Analyse af elevers testforløb 14

Formålet med DNT *15*

Sådan anvendes testene i praksis **16**

Anvendelse af testresultater *16*

Formativ anvendelse af testresultaterne 17

Fortolkning af resultater 20

Introduktion til testen *23*

Kvalificerede gæt i DNT 23

Forventningsafstemning i instruktioner 25

Kritik af DNT **27**

Lærere, elevers og forskeres kritik *27*

Vores kritik af DNT *34*

Diskussion **35**

Konklusion og handleperspektiv **42**

Litteraturliste **45**

Indledning

Evaluering i den danske folkeskole er og har i løbet af det sidste årti været et *hot topic* i den offentlige debat, hvor bølgerne til tider er gået højt. Især er de nationale test (DNT) blevet debatteret heftigt. DNT er et emne, vi som kommende folkeskolelærere ikke kan undgå at forholde os til, i og med anvendelse af dem er obligatorisk og dette tilsyneladende ikke går i retning af at ændre sig i den nærmeste fremtid. Derfor vil vores opgave belyse DNT som evalueringsredskab og anvendelsen samt det iboende fortolkningsarbejde af testene med henblik på at afdække, hvilke potentialer og udfordringer, de bærer med sig i det pædagogiske virke som lærere i den danske folkeskole.

"De nationale test er et værktøj til den løbende evaluering i folkeskolen. Testene er et pædagogisk redskab, som kan bidrage til den videre tilrettelæggelse af undervisningen". Sådan lyder en kort introduktion til DNT inde på Danmarks Læringsportal, EMU udgivet af Undervisningsministeriet (UVM). Det fremgår af bemærkningerne til loven at testen hovedsageligt er tiltænkte formative formål, men det er vores indtryk, at der er uoverensstemmelse med hvordan, de oprindeligt blev tænkt, og hvordan de bliver anvendt af lærerne og de mange andre interessenter i den danske folkeskole.

Diskussionen om DNT dukker med korte mellemrum op i vores hverdag på forskellige faconer og i diverse medier; nyhedsartikler fra hhv. Politiken, Berlingske og Folkeskolen.dk for bare at nævne få. På trods af den igangværende diskussion ser det ikke ud til, at der er ret mange, der er fortrolige med testens mekanismer og formål og lærerne mangler kendskab til, hvordan testen afvikles, fortolkes og anvendes hensigtsmæssigt i en undervisningssammenhæng. Vores indtryk er, at artiklerne om emnet i langt de fleste tilfælde nærmere er et udtryk for enkeltpersoners holdninger frem for faktabaserede tekster, der kan forklare os om testens bagvedliggende teknikaliteter og hvorfor testene ikke anvendes til de oprindeligt tilsigtede formål. Vi undrer os over den udbredte kritik af testen som evalueringsredskab, da test i sig selv hverken er nyt eller revolutionært inden for folkeskolens verden; Er det den politiske kontekst, den adaptive virkemåde, eller hvad er det ved netop DNT, der er anderledes end tidligere? Og præcis hvordan kommer uoverensstemmelsen mellem testens nuværende anvendelse i praksis og dens tiltænkte formål til udtryk? Disse er nogle af de spørgsmål, vi har sat os for at undersøge for bedre at forstå, hvordan vi kan og bør anvende testen mest hensigtsmæssigt med målet om at finde

frem til dens muligheder og begrænsninger. Vi beskæftiger vi os primært med testene i matematik og læsning/dansk ift. at tale om DNT, da der er hhv. tre og fire obligatoriske nationale test i disse fag i løbet af hver folkeskoleelevs skolegang (i modsætning til både engelsk og de naturvidenskabelige fag, som kun har én obligatorisk test). Det, at der måles flere gange over tid, giver anledning til at tale om progression, som er et nøglebegreb i den nye reform. Derudover er det de eneste af testene, som giver læreren en kriteriescore og en progressionsgraf - hvad dette indebærer, vil blive uddybet senere i opgaven.

Vi er af den overbevisning, at det er ved at undersøge emnet ud fra en nysgerrig tilgang, hvor egne holdninger sættes i baggrunden, og både kritikere og fortalere så vidt muligt imødekommes, vi kan opnå størst mulig indsigt i emnet. Dog vil vi, udover at redegøre for, fortolke og diskutere de kritikpunkter, vi finder mest essentielle, hvor egne holdninger kommer til udtryk. Vi vil forsøge at stille kritikpunkterne overfor UVMs hensigt med at implementere DNT i første omgang med henblik på at finde frem til løsningsorienterede handleperspektiver. Vi vil bl.a. undersøge, hvorvidt de pædagogiske gevinster ved anvendelsen af testen forekommer lige så tydeligt for dem, der skal bruge testene i praksis, som for dem, der har udviklet dem.

Inden længe kommer vi til at stå i et klasserum, hvor vi er underlagt nogle politisk satte rammer, og vi er derfor interesserede i at finde ud af, hvordan vi kan handle mest hensigtsmæssigt inden for disse, hvilket har ledt os frem til nedenstående problemformulering og hypoteser.

Problemformulering

Hvilke pædagogiske udfordringer og potentialer ligger der i arbejdet med de nationale test, og hvordan kan læreren anvende dem formativt i praksis?

Hypoteser

Til at anskueliggøre debatten om hvorvidt DNT er et hensigtsmæssigt og gavnligt evalueringsværktøj eller ej, har vi valgt at belyse vores problemformulering ved en række opstillede hypoteser. Disse tager udgangspunkt i vores forhåndsviden om og indtryk af testen samt de holdninger, vi er stødt på i forbindelse med indsamling af empiri. Inspireret af Karl Poppers kritiske rationalisme (Brandt, 2009, s. 391-402), har vi forsøgt at opstille

Lærerprofession.dk - et site for lærerpraksis og professionsudvikling

hypoteser således, at de er falsificerbare og kan be- eller afkræftes gennem empirisk undersøgelse.

1. *Om læreres holdninger*
    a. De fleste lærere er imod DNT
    b. Lærerne ser testene som et udtryk for mistillid fra det politisk niveau (Ministeriet og kommunerne)
    c. Lærerne har manglende kendskab til, hvordan resultaterne hhv. fortolkes og anvendes hensigtsmæssigt
    d. Evalueringen kan fungere som en støtte for læreren i forbindelse med at løsrive sig fra sine narrativer om eleverne (Hattie 2014, Løw 2009).
2. *Om elevers tilgang*
    a. Eleverne bryder sig ikke om at blive testet.
    b. Det er især de dygtige elever, som har negative oplevelser med adaptive tests
    c. Elevernes oplevelse afhænger af lærerens introduktion til testen - den er også afgørende for deres endelige resultater.
3. *Sammenhæng i test og undervisning*
    a. Testene bliver ikke anvendt formativt, som ministeriet beskriver er formålet.

Empiri, metode og læsevejledning

Vores problemformulering knytter sig til et mikroperspektiv (hvordan kan vi, som lærere, anvende DNT mest hensigtsmæssigt i klasserummet), men for at besvare dette må vi zoome ud til et makroperspektiv (hvordan DNT spiller ind på samfundsniveau). DNT åbner for en mangefacetteret diskussion med flere perspektiver. Det er især de landspolitiske perspektiver (makroniveauet), der præger debatten, men vi vil også fokusere på de praktiske faktorer, da de spiller en afgørende rolle for vores handlemuligheder ift. at anvende testen som et pædagogisk redskab. Vi tager udgangspunkt i et makroperspektiv, hvorefter vi i handlingsperspektiv zoomer tilbage til klasserummet og vores handlemuligheder som lærere i den danske folkeskole.

For at belyse diskussionen om DNT fra et lærerperspektiv, har vi undersøgt 340 læreres holdning til og viden om DNT. Måden, vi undersøger dette på, er gennem et

spørgeskema. Vi har udsendt invitation til et stort antal skoler og modtaget tilbagemeldinger fra lærere fra forskellige folkeskoler og kommuner i Danmark. Heraf er der 213 respondenter, der har besvaret alle spørgsmål. Vi spørger ind til lærernes personlige opfattelser af testen samt deres anvendelse og fortolkning af testresultater. I forbindelse hermed har vi stillet nogle "tjekspørgsmål" (som kan besvares rigtigt eller forkert) for at få indblik i, hvor fortrolige lærerne er med testens mekanismer. Afslutningsvis er der et kommentarfelt for at indhente respondenternes holdninger til DNT og/eller andre kommentarer.

Emnet DNT knytter sig som nævnt til både et mikro- og makroperspektiv, hvorfor det er hensigtsmæssigt at benytte både kvalitativ og kvantitativ tilgang til etablering af egen empiri. Vi har derfor:

- Dels behandlet vores undersøgelse kvantitativt for at få indblik i en større sammenhæng, hvor vi anskuer besvarelserne ud fra et kritisk rationalistisk perspektiv (vi forventer ikke, at komme frem til en endegyldig sandhed, men måske kan vi nærme os et realistisk billede af, hvordan verden ser ud). Pointer fra undersøgelsen vil bl.a. blive illustreret gennem grafer og diagrammer.
- Dels kombineret med en hermeneutisk tilgang, hvor vi fortolker de holdninger, der gemmer sig bag tal og statistikker i et forsøg på at forstå, hvorfor respondenterne har besvaret spørgeskemaet som de har. Herunder behandler vi kommentarerne, vi har modtaget, via en hermeneutisk tilgang, hvor vi anvender fortolkning som metode.

For at supplere de to tilgange har vi på Nærum Skole lavet en systematiseret observation af en lærers instruktion til en national test i 4. klasse. Observationen er systematiseret på baggrund af UVMs og NordicMetrics' vejledninger til gennemførelse af test (2013b). Observationen behandles udelukkende via en hermeneutisk tilgang, da formålet er, at forstå lærerens instruktioner og via fortolkning sammenligne disse med vejledningerne. I forbindelse med observationen benyttede vi muligheden for at få et indblik i nogle af elevernes oplevelser før og efter testen. Her blev 10 elever udvalgt til at besvare et spørgeskema både før og efter testen. Fokus i spørgeskemaet var elevernes personlige oplevelse af DNT, da dette har været et tilbagevendende emne i debatten. Besvarelserne fra eleverne behandles ud fra en kvantitativ metode, som endvidere kombineres med fortolkning.

Vi har også overvejet andre undersøgelsesdesign: fx at lave kvalitative semi-strukturerede interviews med eleverne, som sidenhen kunne behandles ud fra en hermeneutisk tilgang, da denne er oplagt til forstå enkeltpersoners meninger og holdninger, i modsætning til kvantitative spørgeskemaer. Her kunne vi have anvendt spørgeskemabesvarelserne som belæg for at udvælge nogle elever med forskellige holdninger til interviews. Eller at skaffe flere respondenter (elever, der har gennemført DNT) til vores undersøgelse og således fået et større og mere retvisende indblik i de danske folkeskoleelevers holdninger til test.

For at bruge kræfterne bedst muligt har vi i stedet valgt at trække på andres empiri, fx undersøgelser lavet i forbindelse med Jeppe Bundsgaard og Morten R. Pucks rapport, *Nationale test: Danske lærere og skolelederes brug, holdninger og viden* (2016), som er repræsentativ for danske folkeskolelærere (s. 37), som vi bruger komparativt ift. vores egen undersøgelse. Derudover har vi fået adgang til upublicerede data fra en undersøgelse, der blev lavet af firmaet NordicMetrics i 2014 i forbindelse med DR-programmet *Folkeskolen Forfra*. Det er en undersøgelse af alle 140 elever fra 6.-8. klasse på en nordjysk folkeskole, hvor elevernes oplevelser og holdninger om test med fokus på DNT bliver belyst.

Vores opgave er bygget op af følgende dele:

1. Redegørelse for centrale begreber; *evaluering* og herunder forskellen på begreberne *summativ* og *formativ*
2. Beskrivelse af DNT og mekanismerne, den består af. Her belyses bagvedliggende tekniske funktioner og kompleksiteten i måleredskabet.
3. Belysning af, hvordan testen anvendes i praksis pt. med fokus på at sætte dette i relation til UVMs formål med udgangspunkt i hhv. vores egen spørgeskemaundersøgelse og Bundsgaard og Pucks rapport.
4. Redegørelse og fortolkning af kritik af DNT, som løbende vil blive diskuteret
5. Egne kritikpunkter af testen og testsystemet.
6. Diskussion, hvor testen anskues ud fra hhv. et ideologisk perspektiv og praktisk perspektiv, hvor vi inddrager bl.a. Gert Biesta og John Hattie (2009, 2014).
7. Konklusion og handleperspektiv

Sabrina Ferguen - et site for lærerpraksisser professionsudvikling

Evaluering

Undervisning kan beskrives som en proces, der kræver forudgående planlægning og sideløbende/efterfølgende evaluering. Ifølge folkeskoleloven skal der, som led i undervisningen, løbende foretages evaluering (Folkeskoleloven §13.2), og hermed skabes grundlag for opfølgning på elevens faglige udvikling/progression (Folkeskoleloven §18 "*Undervisningens tilrettelæggelse, herunder valg af undervisnings- og arbejdsformer, metoder, undervisningsmidler og stofudvælgelse, skal i alle fag leve op til folkeskolens formål, mål for fag samt emner og varieres, så den svarer til den enkelte elevs behov og forudsætninger.*"). Evalueringen skal således danne grundlag for både vejledning af den enkelte elev og lærerens planlægning og tilrettelæggelse af undervisningen. Endvidere skal der ske underretning af forældrene om elevens standpunkt samt udbytte af undervisningen (Folkeskoleloven §13 "*Eleverne og forældrene, jf. § 54, skal regelmæssigt underrettes om lærernes og eventuelt skolelederens syn på elevernes udbytte af skolegangen"*). Begrebet evaluering optræder første gang i folkeskoleloven i 1993, men at evaluere er hverken nyt eller revolutionerende i sig selv. Begrebet evaluering har været i konstant udvikling gennem tiden og har dermed også fået nye betydninger, end ordet havde i sin oprindelige form. Oprindeligt kommer det af fransk og betyder *værdifastsættelse* og blev hovedsageligt anvendt i handelskontekster som fx ift. valuta, bytteforhold og møntfod (Andersen m.fl. 2017, s. 14). I 1932 introducerede den amerikanske uddannelsesforsker Ralph Tyler (1902-1994) begrebet evaluering med helt ny og uddannelsesmæssig betydning (Andersen m.fl. 2017, s. 15). Siden da bredte det sig til andre sektorer og i Danmark optræder det i en pædagogisk kontekst fra slutningen af 60'erne. Der findes mange forskellige evalueringsmetoder, og udover det findes der forskellige former for anvendelse af disse. Det er ikke til at sige, hvilke der fungerer bedst, da det er kontekstafhængigt.

Formativ/summativ

Omkring 1960'erne blev det tydeligt for bl.a. den engelske filosof Michael Scriven, at man er nødt til at skelne mellem to hovedformer for evaluering, hvorefter han introducerede begreberne summativ og formativ evaluering (Andersen m.fl. 2017, s. 101). Al evaluering er retrospektivt, men kan anvendes på forskellige måder, og den essentielle forskel på de to

former for evalueringer beror i, hvornår i et forløb man udfører evalueringen. Den summative evaluering har til formål at *opsummere* værdien af fx et undervisningsforløb, der er blevet gennemført og sker således ved afslutningen af forløbet. Et klassisk eksempel herpå er folkeskolens afgangsprøve, som fungerer afrundende for elevens folkeskolegang. Den formative evaluering, derimod, forekommer før og/eller undervejs i et forløb med hensigten at skabe grundlag for at *forme* udviklingen af det, der evalueres. En metafor for de to forskellige former for evalueringer kan være, når en kunde kommer ind på en restaurant og smager på dagens suppe, hvorefter han fælder en dom - dette er en summativ evaluering. Når kokken derimod smager på suppen i køkkenet, inden den serveres, er det en formativ evaluering, der sker med henblik på at gøre suppen bedre (er den saltet rigtigt? skal der flere krydderier i?). Ved dette forstår vi, at den formative evaluering foretages med henblik på fremadrettede handlinger og skaber således grundlag for planlægning af undervisning i det, der evalueres.

Ofte knyttes forskellige dataindsamlings- og evalueringsmetoder an til de to hovedformer for anvendelse, fx anvendes en dataindsamlingsmetode som logbøger eller porteføljer hyppigt i forbindelse med en formativ kontekst, hvor dataindsamlingsmetoder som test eller fremlæggelser (fx som belæg for at give årskarakterer) kan forbindes med en summativ evaluering.

Mens prøver og eksaminer (fx overhøringer og vurdering af skriftlige- eller praktiske præstationer) altid har været anvendt i folkeskolen og var fx beskrevet i loven af 1814, er test historisk set af nyere dato (Andersen m.fl. 2017). Test kan defineres som besvarelser af en serie af afgrænsede spørgsmål, hvor det relativt entydigt kan afgøres om svaret er rigtigt eller forkert, og hvor bedømmelsen sker på grundlag af optælling af korrekte svar (Andersen m.fl. 2017). Principperne bag bedømmelse på grundlag af test er først og fremmest udviklet i engelsktalende lande, og her anvendes test mest summativt (Wandall, 2013). I forlængelse heraf kan man argumentere for, at DNT, som er en adaptiv test tiltænkt formativ evaluering forekommer kulturfremmed i Danmark. De bagvedliggende mekanismer, herunder testens adaptive princip, uddybes i følgende afsnit.

Sabrina Ferguen - et site for lærerpraksis over professionsuddeling

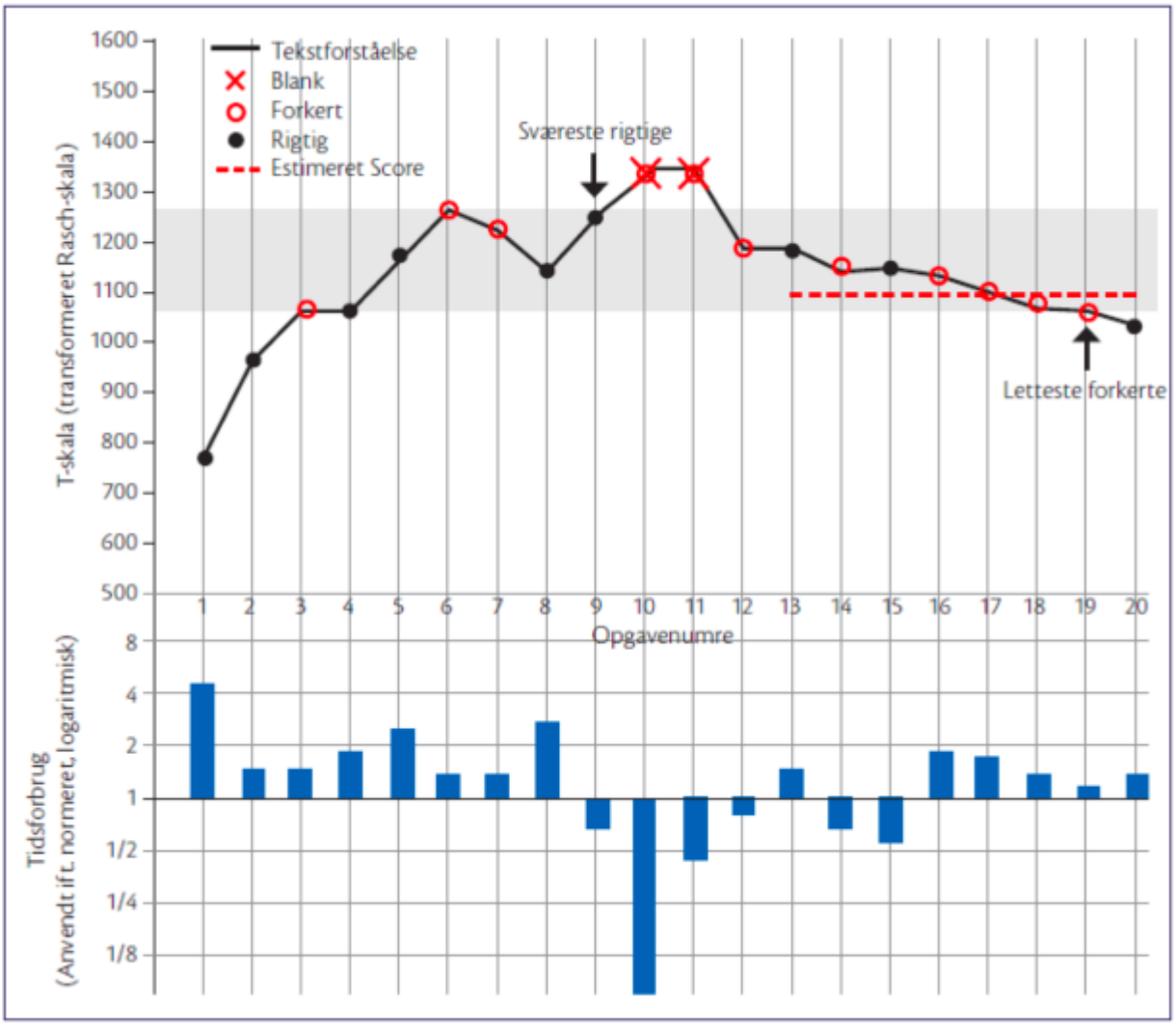
Bagom DNT

I dette afsnit beskrives noget af teknikken bag DNT. DNT opstod baggrund af: 1) middelmådige PISA-resultater i 2000 og 2003, 2) påpegning af mangel på evalueringskultur af OECD og EVA i 2004 og 3) regeringsbeslutning i 2005 om etablering af nationale test (OECD-rapport om grundskolen i Danmark, 2004). I artiklen *National tests in Denmark–CAT as a pedagogic tool* (Wandall, 2011), fremgår en beskrivelse af DNT som IT-baserede, selvscorende, adaptive test, som blev implementeret i fuld skala i 2010, hvor det blev obligatorisk at gennemføre testene i en række fag på bestemte klassetrin. Danske folkeskoleelever skal således gennemføre mindst 10 af sådanne test i løbet af deres skoletid i fagene læsning/dansk (2., 4., 6. og 8. klassetrin), matematik (3., 6. og 8. klassetrin), engelsk (7. klassetrin) og de naturvidenskabelige fag (fysik/kemi, biologi og geografi på 8. klassetrin). Testene skal gennemføres i løbet af perioden februar til og med april. Derudover er der for hver test stillet to frivillige testforsøg til rådighed pr. elev hvilket gør det muligt at fx teste årligt i matematik fra 2.-9. klasse og læsning fra 1.-9. klasse.

Det adaptive princip

At testen er adaptiv betyder, at testens forløb tilpasses elevens svarmønster: Når en elev svarer rigtigt på et spørgsmål, er det næste sværere og omvendt, hvis de svarer forkert. For at kunne udvælge opgaverne er det nødvendigt at kende deres sværhed – alle opgaver er derfor blevet afprøvet på ca. 700 elever, og kun de opgaver, der passer til den såkaldte Rasch-model, kan anvendes (Wandall, 2011). Rasch-modellen er udviklet i begyndelsen af 1950'erne af den danske matematiker og statistiker, Georg Rasch og bygger på princippet om, at der er sammenhæng mellem elevers dygtighed og opgavers sværhedsgrad; en elev med en given dygtighed har 50% sandsynlighed for at svare rigtigt på en opgave, der har en sværhedsgrad svarende til elevens niveau. Hvis elevens dygtighed overstiger opgavens sværhedsgrad, er sandsynligheden for at eleven svarer korrekt større og omvendt (Andersen m.fl. 2017, s. 49-51). Forudsætningen for, at man ved testen kan bestemme elevens dygtighed, er, at eleven har gjort sit bedste for at svare rigtigt – dvs. besvaret opgaverne i overensstemmelse med sin dygtighed. Forudsat de gør det, vil det være muligt at beregne elevens niveau gennem testen. I et ideelt testforløb vil halvdelen af opgaverne være

besvaret korrekt og den anden halvdel forkert - på den måde er det muligt at fastlægge elevens øvre og nedre faglige barre inden for det, der testes i. Dette er illustreret i figur 1.

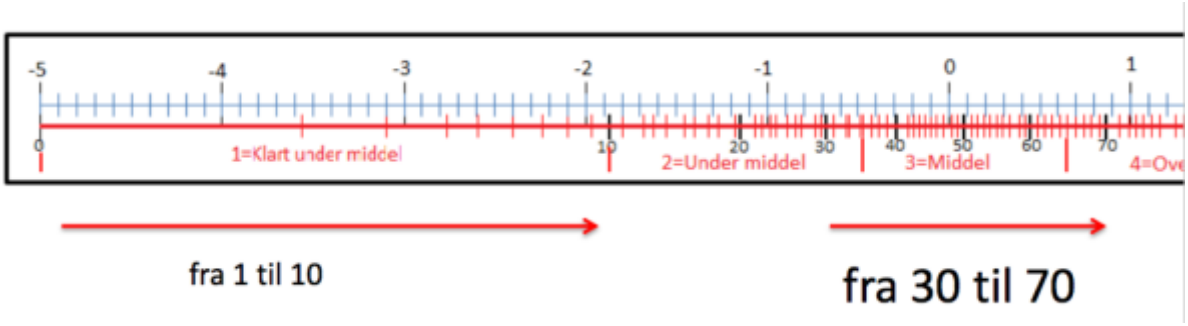


Figur 1 - Det adaptive testforløb. (Andersen m.fl. 2017, s. 68, figur 14)

Normbaserede, kriteriebaserede og ipsative referencer af resultater

Når testen er gennemført, kan læreren se resultaterne på forskellig vis; de kan enten få vist resultaterne på en kriteriebaseret skala eller som normbaserede resultater. De kriteriebaserede resultater holder elevens besvarelser oppe imod kriterier, fastsat af UVM, for at score hhv. *ikke tilstrækkelig*, *mangelfuld*, *jævn, god, rigtig god* og *fremragende.* Den normbaserede reference placerer eleven på en skala ift. andre elever på klassetrinnet (baseret på en percentilskala), der fortæller, hvor han eller hun ligger ift. landets elever på tilsvarende klassetrin i 2010, da skalaen blev normeret. Det vil altså sige, at hvis en elev scorer 58 på den normbaserede skala, er der 42% af landets elever, der har scoret bedre

end ham eller hende (i 2010). Når man måler i percentiler, skal man være opmærksom på, at fortolkningsarbejdet bliver kompliceret, i og med et fremskridt på ét point ikke altid er lige meget værd ift. målt læring (faglige tilvækst), da percentiler bygger på en rangordning af eleverne, hvor de fleste ligger omkring midten. Det kan sammenlignes med et maratonløb, hvor der er 100 deltagere. Her vil der typisk være stor spredning blandt de bagerste og de forreste løbere, mens størstedelen "klumper sammen" i midten. Hvis en af de bagerste eller forreste løbere vil nå længere op i placeringen (metafor for rangorden), må de hæve hastigheden (metafor for dygtighed) betydeligt i modsætning til løberne i midten, der kun behøver at hæve hastigheden en lille smule for at overhale hele midterfeltet. I metaforen svarer løbernes placering til elevernes percentiler og i stedet for løbehastighed taler vi om faglig udvikling (progression). I praksis betyder det, at en elev, der rykker sig fra 1-10 reelt set har flyttet sig næsten dobbelt så meget, som en elev, der rykker sig fra 30 - 70.

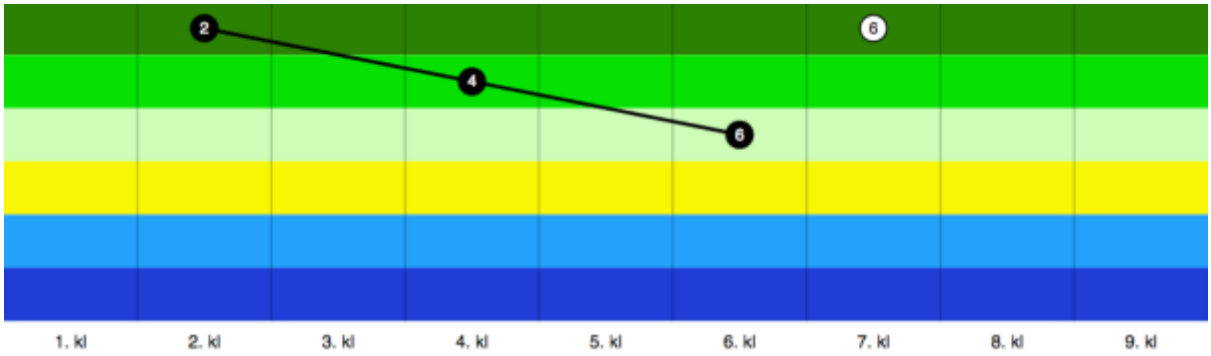


Figur 2 - Screenshot fra oplæg v. Jakob Wandall på Tårnborgkonferencen 2015 (Drachmann, 2015)

Ovenstående figur illustrerer førnævnte pointe om forholdene i den normbaserede skala, hvor tallene over linjerne illustrerer logits fra Rasch-modellen, som altså er det, der skal svare til elevens dygtighed, som måles på en intervalskala, hvor hvert point er lige meget værd. Tallene under linjerne er det, der viser den tilsvarende percentil, som måles på en ordinalskala, hvor hvert point ikke er lige meget værd alt efter hvor på skalaen, man er. De røde, lodrette streger viser hver percentil. Herunder er udviklingen fra hhv. 1-10-percentilen og 30-70-percentilen vist ved længden af de røde pile under skalaen. Når elevens resultater sammenlignes med resultater fra tidligere præstationer af samme elev som i figur 2, kaldes det en ipsativ reference (Dudley & Swaffield, 2008, s. 105).

I testsystemet findes der eksempler på alle tre typer af referencer (normbaserede, kritieriebaserede og ipsative referencer), men der, hvor læreren kan finde den ipsative reference, er den kombineret med elevens kriteriescore. Dette er, hvad der i testsystemet

kaldes visning af *"progression",* hvor man kan følge historikken i elevens kriteriescore. Et eksempel på en progressionsvisning fra testsystemet findes i nedenstående figur 3. De farvede bånd i figuren repræsenterer de førnævnte kategorier fra *ikke tilstrækkelig* til *fremragende*. De sorte prikker viser kriteriescoren fra de obligatoriske tests, hvor den hvide prik repræsenterer scoren fra en frivillig test. Eleven har i de obligatoriske test scoret *fremragende* i 2. klasse, *rigtig god* i 4. og *god* i 6. klasse. Det, at grafen er nedadgående betyder ikke nødvendigvis, at eleven er blevet dårligere til at læse. At scoren for 2. klassetesten er kategoriseret som *fremragende* skal ses ift., at eleven går i 2. klasse. Det kan meget vel være et lavere læseniveau end en elev, der scorer *rigtig god* i 4. klasse. Det samme gør sig gældende for 4. og 6. klassetestene.



Figur 3 - En 7. klasse elev fra Tove Ditlevsens skole – anonyme elevdata fra vores praktikforøb (Kilde: screenshot fra testogprøver.dk*)*

Samme pointe gør sig gældende for de normbaserede resultater, som ligeledes er bestemt efter klassetrinnet. En analyse[1] har fx vist, at en elev, der scorer 45 i en 4. klassetest i sprogforståelse, og som senere scorer 34 i 6. klassetesten, er blevet betydeligt dygtigere til at læse. Pointen er, at dygtighed fx i læsning med den såkaldte Rasch-model kan opgøres helt uafhængigt af elevens alder og klassetrin, og netop Raschmodellen er grundlag for DNT (Wandall, 2011). Også på andre områder adskiller DNT sig fra andre landes testsystemer, fx har læreren adgang til at gennemse sine elevers besvarelser af opgaverne, når testen er taget, og kan på den måde få konkretiseret de områder, hvor eleven behøver mere støtte.

[1] Den forskningsbaserede danske virksomhed NordicMetrics har for Københavns kommune gennemført en analyse af hvordan sammenhængen er mellem skalaerne i de forskellige klassetrins test i læsning (NordicMetrics, 2013a)

Er resultaterne retvisende?

Statistisk usikkerhed

I kommentarfeltet i slutningen af vores spørgeskema, har flere af lærerne peget på testens statistiske usikkerhed og stiller sig kritiske overfor, hvorvidt resultaterne er retvisende for elevens dygtighed. Her er det nærliggende at forveksle det abstrakte begreb *statistisk usikkerhed* (SEM=Standard Error of Measurment) med, hvorvidt testens score giver et retvisende billede af elevens dygtighed (Wandall, 2011). Den statistiske (u)sikkerhed er en teknisk beregning på grundlag af testens serie af besvarelser (dvs. serie af opgaver og elevbesvarelser jf. figur 1) – og på grundlag af denne serie af besvarelser beregnes resultatet som ét tal (en score).

Et grundlæggende princip er, at den statistiske sikkerhed altid øges med antallet af besvarede opgaver, uanset om eleven besvarer opgaverne svarende til sit niveau eller ej. Hvis man skal afgøre, om testens score er retvisende, er det altså ikke nok at se på den abstrakte statistiske sikkerhed; den fortæller alene noget om den statistiske risiko for at tage fejl. Hvis en elev konsekvent besvarer opgaverne i overensstemmelse med sin dygtighed, og opgaverne er designet og udvalgt optimalt, kan man bedømme en elev retvisende på grundlag af et meget lavt antal opgaver (fx under 10).

Forholdet mellem reliabilitet, antal besvarede opgaver, statistisk sikkerhed og hvorvidt testresultatet er retvisende for elevens dygtighed er en teknisk, statistisk diskussion, som ligger ud over, hvad vi beskæftiger os med i nærværende opgave. En populær beskrivelse af disse sammenhænge er gengivet i artiklen ”*Hvornår leverer de nationale test en reliabel måling? Og hvornår er den retvisende for elevens dygtighed”* på folkeskolen.dk (Wandall, 2014).

Ud fra dette forstår vi altså, at en af det helt store usikkerheder, når det kommer til DNT, er, hvorvidt eleven besvarer opgaverne svarende til hans eller hendes dygtighed, hvilket der heldigvis er mulighed for at undersøge nærmere, som forklaret i følgende afsnit.

Analyse af elevers testforløb

Foruden de forskellige scorer (norm-/percentil- og kriteriebaserede scorer samt progressionstabel), har læreren adgang til elevens testforløb, herunder til at se opgaven,

Lærerprofession.dk - et site for lærerpraksis og professionsudvikling

elevens besvarelse, hvad opgaven har drejet sig om, hvilket profilområde den hører til, hvor lang tid eleven har brugt, hvor lang tid opgaven er normeret til samt hvilken percentil opgaverne svarer til – opgavens sværhedsgrad. Ved hjælp af disse oplysninger, har læreren mulighed for at analysere elevens testforløb og få et klarere billede af, hvorvidt resultatet er retvisende. Vi har ved gennemgang af elevforløb oplevet, at det ikke er usædvanligt, at elever springer over (svarer blankt – ofte meget hurtigt) eller giver et ukvalificeret gæt fx på de tekstrige opgaver. Når eleven svarer blankt, tæller det i beregningen som et forkert svar, hvorfor man må være opmærksom på bl.a. den anvendte tid ift. den normerede. I tilfælde, hvor eleven bare vil videre til næste opgave, vil de måske have brugt ganske få sekunder på en opgave, der er normeret til betydeligt længere tid.

Formålet med DNT

Progression er et nøgleord, når vi taler om DNT. Et af de overordnede formål med testene er nemlig, at de skal fungere som et redskab, der hjælper læreren til at få indsigt i elevens faglige udvikling (UVM, 2017b). Derudover er de tiltænkt formativ anvendelse i folkeskolen; "*På baggrund af testresultaterne kan lærerne målrette undervisningen til den enkelte elev, og i højere grad arbejde med undervisningsdifferentiering*" (UVM, 2017a). At redskabet er tiltænkt læreren adskiller DNT fra PISA-undersøgelserne, som er en stikprøveundersøgelse af ca. 5000 danske skoleelever og skal fungere som et uddannelsespolitisk evalueringsredskab til UVM og andre fagfolk (Hummelgaard, 2015).

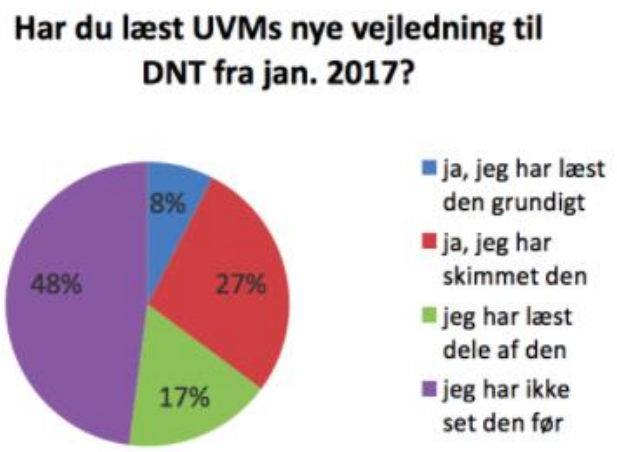
I det omfang, at det ikke er i modstrid til den formative anvendelse, kan DNT også bruges som grundlag for politisk og ledelsesmæssig monitorering af udviklingen i det faglige niveau (jf. fx den nationale præstationsprofil samt folkeskoleforligets målsætninger).

Sådan anvendes testene i praksis

Anvendelse af testresultater

To af vores hypoteser drejer sig om sammenhængen mellem testens formål, og måden hvorpå resultaterne anvendes i folkeskolen pt. (hypotese 1c[2] og 3a[3]). Med udgangspunkt i besvarelser af vores spørgeskemaundersøgelse vil vi nærme os en forståelse af, hvordan testresultaterne anvendes i praksis. Et af de første spørgsmål, vi stillede, gik på hvorvidt og hvor grundigt, de havde læst UVMs opdaterede vejledning for DNT. Der er flere årsager til, at vi kun spørger efter den nyeste vejledning: For det første er det den eneste, der er tilgængelig pt., og for det andet kan vi på denne måde få indblik i, hvorvidt lærerne holder sig ajour med UVMs nyeste informationer om brugen af DNT. Undersøgelsen viser, at knap halvdelen af respondenterne aldrig havde set vejledningen før og kun 8% havde læst den grundigt (se figur 4).

Selvom vejledningen er udførlig på mange punkter både ift. instruktion, afvikling og fortolkning, ser vi nogle afgørende mangler i informationer om de bagvedliggende mekanismer, som kan være afgørende for fortolkningen. Bl.a. hvordan de normbaserede resultater i virkeligheden er percentiler, der fungerer som en rangorden og derfor svarer til en ordinalskala, hvor hvert point ikke er lige meget værd, er ikke beskrevet i vejledningen.



Figur 4 - Diagram over besvarelser på spørgmål om hvorvidt lærerne har læst den nyeste vejledning fra UVM

[2] Hypotese 1c: Lærerne har manglende kendskab til, hvordan resultaterne hhv. fortolkes og anvendes hensigtsmæssigt
[3] Hypotese 3a: Testene bliver ikke anvendt formativt, som ministeriet beskriver er formålet

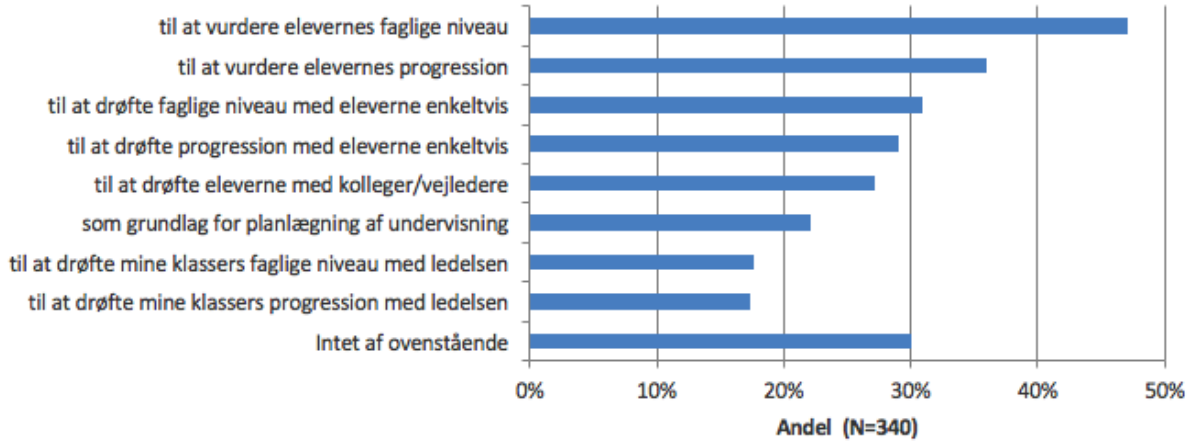
Formativ anvendelse af testresultaterne

Nu har vi sporet os nogenlunde ind på de responderende læreres forudsætninger for anvendelse af DNT, hvilket bringer os et skridt nærmere at afdække hypotese 1c[4]. Denne vil vi arbejde videre med løbende i opgaven, men først vil vi undersøge lærernes besvarelse af, hvordan de anvender testen i praksis og i hvilke sammenhænge.

Med henblik på at belyse, om testresultaterne anvendes formativt (hypotese 3a[5]), vil vi se på lærernes besvarelse af spørgsmålet om, hvorvidt de anvender testens resultater *som grundlag for planlægning af undervisning* på baggrund af vores førnævnte definition af formativ evaluering ("*den formative evaluering foretages med henblik på handlinger fremadrettet og skaber således grundlag for planlægning af undervisning i det, der evalueres"* (se afsnittet *Formativ/summativ* på s. 8)). Vi skal være opmærksomme på, at vores undersøgelse ikke nødvendigvis er repræsentativ for lærerstaben i den danske folkeskole, da den er baseret på selvudvalgte respondenter; det er plausibelt at antage, at (i hvert fald nogle af) lærernes incitament for at besvare spørgeskemaet har været at tilkendegive deres holdning til emnet med henblik på at gøre en forskel, hvilket især kommer til udtryk i en kommentar fra undersøgelsen; "*Rigtig godt initiativ! Jeg håber, det kan være med til at afskaffe DNT!"*. Vi må altså medregne sandsynligheden for, at en overvejende del af respondenterne har en kritisk indstilling til DNT. Når det er sagt, kan vi se (som det fremgår af figur 5), at der kun er lige over en femtedel af lærerne, der har svaret ja til spørgsmålet om, hvorvidt de anvender testresultaterne formativt.

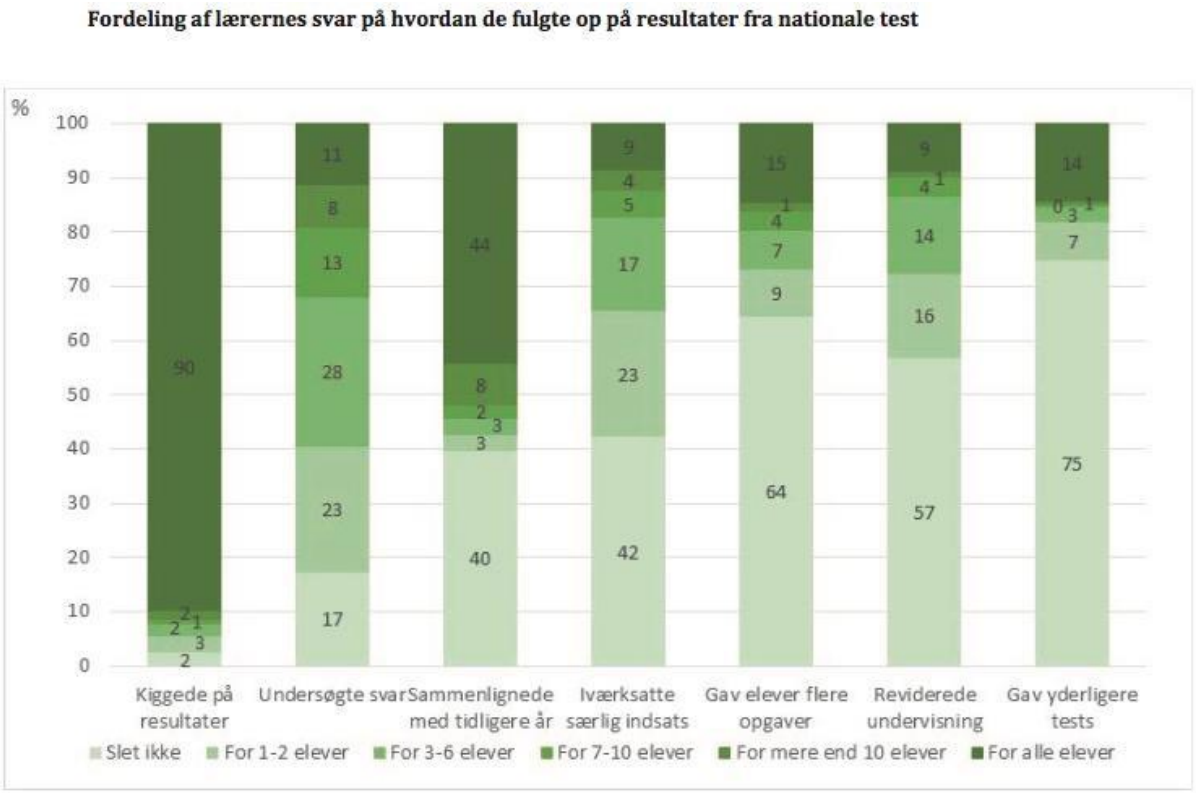
[4] Hypotese 1c: Lærerne har manglende kendskab til, hvordan resultaterne hhv. fortolkes og anvendes hensigtsmæssigt

[5] Hypotese 3a: Testene bliver ikke anvendt formativt, som ministeriet beskriver er formålet.



Figur 5 - Graf over besvarelser på spørgsmål om lærernes anvendelse fra vores undersøgelse

En lignende og repræsentativ undersøgelse af Jeppe Bundsgaard og Morten R. Puck giver et lidt mere nuanceret billede af lærernes anvendelse af testene (se figur 6). I rapporten spørger de om, hvorvidt lærerne følger op på resultaterne ved at revidere undervisningen (Bundsgaard & Puck:2016, s. 16). Dette mener vi kan sammenlignes med vores spørgsmål om, hvorvidt lærerne anvender testresultaterne som grundlag for planlægning af undervisning, da begge svarmuligheder indebærer en indtænkning af resultaterne i undervisningsplanlægningen og rummer mulighed for enten at ændre på indholdet af undervisningen eller fortsætte med noget, der allerede var tænkt/planlagt, hvis det stemmer overens med vurderingen af, hvad eleverne behøver ekstra støtte i.



Figur 6 - Læreres opfølgning på DNT (Kilde: Bundsgaard & Puck 2016, s. 16, figur 2)

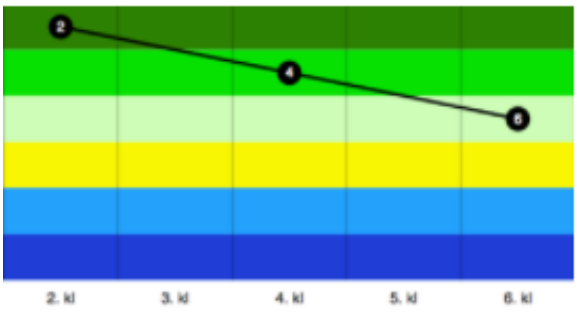
Af figur 6 fremgår det, at 57% af lærerne slet ikke reviderer deres undervisning på baggrund af resultaterne fra DNT, hvilket betyder, at 43% af lærerne gør. Dette er omtrent dobbelt så stor en andel, som i vores undersøgelse. Det kan evt. begrundes i, at vores spørgsmål kun har haft svarmulighederne ja/nej. Måske havde vi fået et mere nuanceret billede, hvis det havde været et vurderingsspørgsmål, hvor svarmulighederne eksempelvis hed *slet ikke, i lav grad, i nogen grad* og *i høj grad*. Der foreligger en risiko for, at dem, der i Bundsgaard og Pucks undersøgelse har svaret "*for 1-2 elever*" til spørgsmålet om revidering af undervisningen, i vores undersøgelse har svaret nej til det lignende spørgsmål, fordi de har vurderet, at det var mere passende end at svare ja. Hvis vi antog, at det var sådan, ville vores resultater ikke være så langt fra hinanden - 28% har hos Bundsgaard svaret, at de reviderer undervisningen for mindst 7 elever i deres klasser, hvor 22% har svaret ja til vores spørgsmål om at anvende resultaterne som grundlag for planlægning af undervisning. Uanset hvad, kan vi på baggrund af vores egen og Bundsgaard og Pucks undersøgelse afkræfte hypotese 1c[6], da de bliver anvendt formativt - bare ikke af *alle* lærere. Dette

[6] 1c: Lærerne har manglende kendskab til, hvordan resultaterne hhv. fortolkes og anvendes hensigtsmæssigt

bekræftes yderligere i figur 6, hvor 58% svarer, at de på baggrund af testresultaterne iværksætter særlige indsatser.

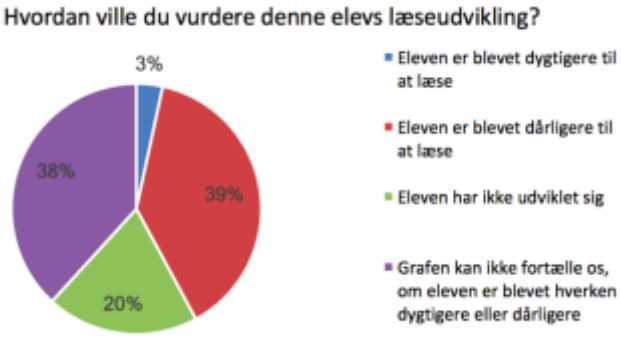
Fortolkning af resultater

I spørgeskemaet stillede vi lærerne nogle tjekspørgsmål, som tog udgangspunkt i forskellige elevers udviklinger for at få indblik i deres evner til at fortolke resultaterne. Resultaterne er taget fra læsetests, men ser ligesådan ud for matematik, hvorfor de besvarelser, vi har fået, fortæller om lærernes fortolkninger af begge tests. Det første, vi bad lærerne forholde sig til, var den udvikling, der er vist i figur 3.



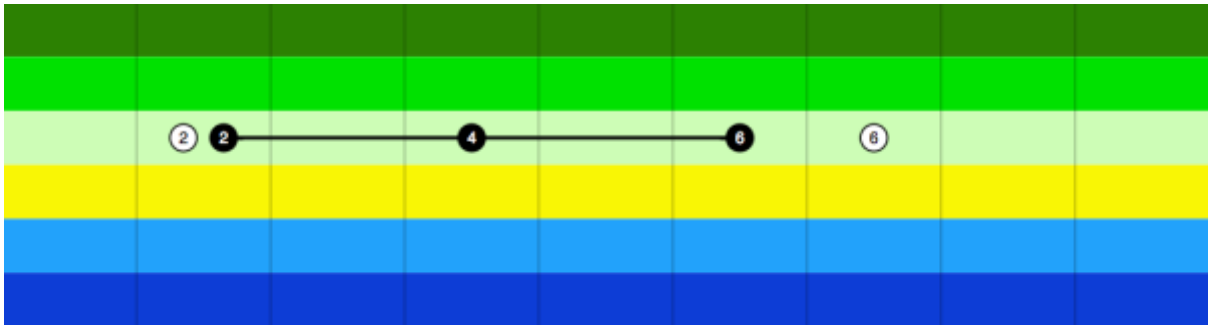
Figur 3

Det er ikke til at sige, om en *fremragende* læser i 2. klasse er bedre end en *jævn* læser i 6. klasse - grafen ikke kan fortælle os, om elevens udvikling. Alligevel svarede knap 40%, at eleven er blevet dårligere (se figur 7). Det er oplagt at antage, at en nedadgående kurve på en progressionsgraf svarer til en nedadgående progression i elevens udvikling. Desværre er det ikke så ligetil, hvilket kan have ærgerlige konsekvenser for den tilbagemelding, eleven får.



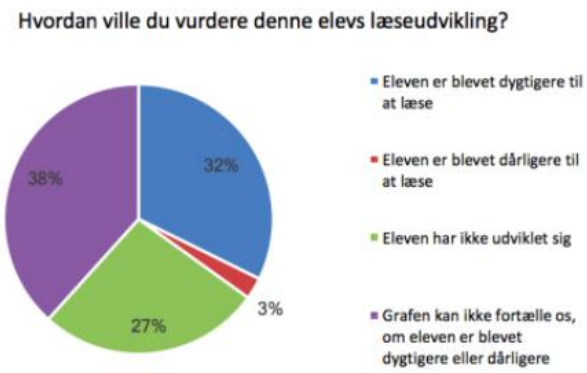
Figur 7 - Diagram over besvarelser på tjekspørgsmål om læseudvikling med nedadgående kurve

En femtedel af lærerne har svaret, at eleven ikke har udviklet sig og en lille andel af lærerne har svaret, at eleven er blevet dygtigere til at læse. De sidste 38% har svaret, at grafen ikke kan fortælle os om elevens læseudvikling. Vi vil i første omgang undlade at konkludere, om de har svaret sådan på baggrund af kendskab til testens teknik, da det senere vil vise sig, at samme procentdel har svaret ligesådan i næste tjekspørgsmål, hvor det korrekte svar er et andet.



Figur 8 - En 7. klasse elev fra Tove Ditlevsens skole – anonyme elevdata fra vores praktikforøb (Kilde: screenshot fra testogprøver.dk)

I figur 8 ses den næste graf, lærerne blev bedt om at vurdere. Her eksemplificeres en elevs progression, som passer til det, der forventes, man lærer i løbet af de fire år, der er testet over - eleven er altså blevet dygtigere år for år svarende til klassetrinnet. Dette var der også knap en tredjedel af lærerne der svarede, hvilket kun er lidt flere end de 27%, der svarede at eleven ikke havde udviklet sig. Der var kun en lille andel, der svarede, at eleven var blevet dårligere til at læse, mens 38% svarede, at grafen ikke fortæller os hverken det ene eller det andet. Dette tolker vi som udtryk for hhv. modstand til testen og mangel på valgmuligheden "*ved ikke*".



Figur 9 - Diagram over besvarelser på spørgsmål om læseudvikling med lige kurve

Hvilken tilbagemelding får eleven så? At vurdere fra vores resultater, har eleven mulighed for at få tre vidt forskellige tilbagemeldinger, alt efter hvilken lærer, der læser resultatet. Vi stillede også tjekspørgsmål, der drejede sig om percentiler - altså normbaserede scorer. Besvarelserne på disse illustrerer samme pointe; svarene stak i tre forskellige retninger. I Bundsgaards og Pucks undersøgelse stillede de lærere seks forskellige tjekspørgsmål, hvor de i gennemsnit svarede rigtigt på 2 - altså en tredjedel (Bundsgaard & Puck 2016, s. 31). e

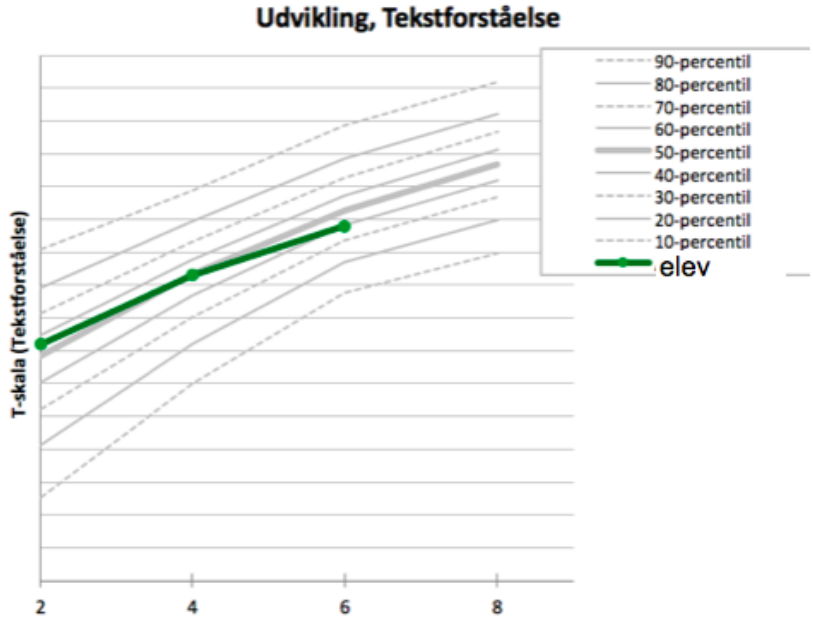
På baggrund af ovenstående kan vi bekræfte hypotese 1c[7], da størstedelen af lærerne enten ikke kan/vil sige noget om elevernes læseudvikling eller svarer noget, der er

[7] Hypotese 1c: Lærerne har manglende kendskab til, hvordan resultaterne hhv. fortolkes og anvendes hensigtsmæssigt

decideret forkert. Det understøttes i flere kommentarer fra vores undersøgelse; "*Da jeg er matematiklærer og ikke kender til DNT i læsning, har jeg svaret ens på alle spørgsmål herom. Savnede en 'Ved ikke'-mulighed"*. Forklaringen holder ikke helt vand, da graferne ser ens ud for matematik og læsning.

Selvom 43% af lærerne anvender resultaterne formativt i forbindelse med undervisningsplanlægningen (jf. figur 6), så er det ét fedt, hvis det sker på baggrund af fejlfortolkede resultater. Patienten bliver heller ikke rask, hvis behandlingen sker på baggrund af en fejldiagnosticering. Det samme gør sig gældende, når næsten 50% af lærerne fra vores undersøgelse svarer, at de på baggrund af testresultaterne vurderer elevernes faglige niveau og 35% vurderer progression.

I forbindelse med anvendelse af testresultaterne finder vi det relevant at nævne "*Beregneren"*, som godt halvdelen af landets kommuner bruger i større eller mindre omfang. Programmet er udviklet som en hjælp til læreren ift. at læse og fortolke resultaterne på mere overskuelig vis. Den omregner percentilerne til progressionscorer (Rasch-score), hvilket muliggør at beregne elevens progression på en egnet skala (intervalskala). Når den omregner til en intervalskala, er tallene betydeligt lettere at fortolke, da hvert point i sådan en *er* lige meget værd. Det er derfor muligt at regne sig frem til, hvor meget eleven har rykket sig fra én test til en anden som vist i figur 10. Her vises den samme elev som i figur 8.



Figur 10 - Screenshot fra Beregneren med anonyme elevdata fra vores praktikforløb

Introduktion til testen

Et af kritikernes hovedargumenter i diskussionen om brugen af DNT bygger på elevernes oplevelse af testen (Madsen, 2014). En afgørende faktor ift. dette er, hvad der sker *før testen* (hypotese 2c[8]). Vi vil undersøge elevernes oplevelser før og efter gennemførelse af DNT. Ved at se på sammenhængen mellem elevernes oplevelser af testen og lærerens introduktion, kan vi nærme os en handlingsorienteret løsning på kritikken (Bundsgaard & Puck, 2016, s. 7). Som belæg for at tale om denne sammenhæng bruger vi hhv. observationer fra Nærum Skole, hvor en lærer igangsætter en 4. klasse i DNT i læsning/dansk og sidestiller observationen med elevbesvarelser på spørgeskemaet, de fik både før og efter testens afvikling. Inden besøget på Nærum Skole havde vi sat os grundigt ind i UVMs og NordicMetrics' vejledninger fra 2017 og havde således en klar ide om, hvordan det anbefales, at testen instrueres og afvikles.

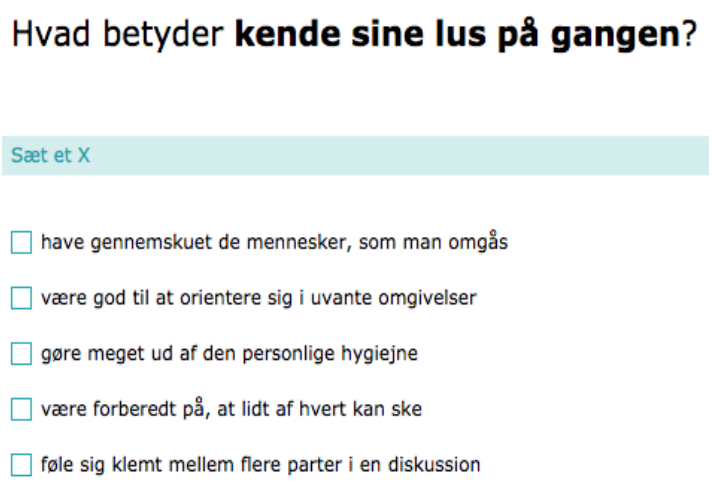
Kvalificerede gæt i DNT

Et punkt, hvorpå læreren fulgte vejledningernes anbefalinger, var i instruktionen; "*Det er meningen, at man skal svare forkert på nogle. Hvis man ikke kan finde ud af det, så gætter man bare! Skriv det I tror... Man skal prøve, hvis man kan!*" (Bilag 1). Her er der overensstemmelse med vejledningerne, der siger, at "*Når eleverne møder opgaver, som de ikke kan svare på, er det bedst, hvis de svarer med et kvalificeret bud."*, hvorefter der er tilføjet en kort begrundelse: *"Et kvalificeret bud indeholder mere information end et blankt svar. Et blankt svar tæller altid som et forkert svar."* (UVM, 2017b, s. 19). Årsagen hertil er, at når det kommer til sprog, har alle mennesker større receptive færdigheder (altså sprog, som man forstår), end produktive færdigheder (sprog, man kan producere). De receptive færdigheder er uobserverbare, da det er noget, der foregår inde i hovedet på eleven, og der, hvor der ikke er overlap med de produktive færdigheder, er det ikke muligt for eleven at udtrykke sig (Richards, 2008). Det kan illustreres med følgende eksempel: "*I can read great novels for example, but I could never write one"* (Richards, 2008). Richards påpeger en skelnen mellem passiv og aktiv sproglig viden, som især sættes i relation til ordforrådstilegnelse, hvor forventningen er, at eleven er i stand til at forstå langt flere ord,

[8] 2c: Elevernes oplevelse afhænger af lærerens introduktion til testen - den er også afgørende for deres endelige resultater.

end han eller hun kan anvende i praksis. Således kan en elev i 2. klasse godt have en idé om, at ordet *palindrom* knytter sig til et sprogligt fænomen, frem for fx en bestemt type bakterie eller en betegnelse for en figur, uden at kunne forklare præcist, hvad det er. DNT interesserer sig både for elevernes aktive og passive færdigheder, hvorfor et kvalificeret gæt indeholder relevante oplysninger om elevens dygtighed.

Det, at eleverne instrueres i at gætte, betyder også, at der er en sandsynlighed for, at de gætter rigtigt, uden at det nødvendigvis beror på hverken på aktiv eller passiv viden, men nærmere ren tilfældighed. Med fem svarmuligheder er der principielt 20% chance for at gætte rigtigt, og det er derfor vigtigt, at eleverne besvarer et højt antal opgaver som beskrevet i afsnittet om statistisk usikkerhed. Grunden til, at vi skriver, at der *principielt* er 20% chance for at gætte rigtigt er, at der i DNT er inkorporeret det, der hedder plausible distraktorer (Andersen m.fl. 2017:56-57). Det betyder, at der er lavet svarmuligheder, der nemt kunne forveksles med det rigtige, da det lyder sandsynligt. Et godt eksempel på en distraktor er sidste svarmulighed i spørgsmålet i figur 11, da udtrykket nemt kunne forveksles med det lignende udtryk *som en lus mellem to negle*, der betyder noget andet. Disse distraktorer gør sandsynligheden for at gætte rigtigt endnu mindre end 20%, hvis man ikke kender svaret i forvejen. En god distraktor er kendetegnet ved, at den "snyder" de mindre dygtige elever, men ikke de dygtigste.



Figur 11 - Eksempel på spørgsmål fra demoudgaven af DNT

Sabrina Ferguen - et site for lærerpraksis - er professionsvidtning

Forventningsafstemning i instruktioner

Et andet punkt, hvorpå læreren, vi observerede fulgte vejledningerne kom til udtryk i følgende: *"Det er ikke en prøve i at finde ud af, om I er gode til at danse, er søde, synge osv. - det er slet ikke det, som den måler*." (Bilag 1). Her understreger og tydeliggør læreren testens funktion og formål; den kan ikke sige noget om andet end lige præcis dét, der måles. Dette gør hun formodentlig med henblik på at afstemme elevernes forventninger til testens formål og hvad resultaterne kan vise.

At stille klare forventninger til eleverne, er vigtigt for deres faglige præstationer. Uddannelsesforskeren John Hattie mener, "[...]*at have høje, men passende forventninger og at give feedback på de rigtige niveauer er af afgørende betydning for opbygning af selvtillid til at gå i gang med udfordrende opgaver"* (Hattie, 2014, s. 98). Nøgleordet her er "*passende"*. Hvis det ikke pointeres, at eleverne med stor sandsynlighed vil møde opgaver, de ikke kan svare på, er der risiko for, at de har forventninger om, at denne test er ligesom dem, de plejer at tage, hvor de bliver testet i det, der forventes de kan - det er den ikke! Testens adaptive princip betyder, at alle vil blive udfordret. Hvis eleverne ikke er indstillet herpå, kan især de dygtige elever få en følelse af ikke at leve op til forventningerne og dermed føle sig utilstrækkelige og blive kede af det. Med andre ord; Det, der forventes af eleverne, skal fremgå tydeligt, da det har en afgørende betydning for deres præstationer og selvtillid. De fagligt svage elever vil med større sandsynlighed være vant til at blive mere udfordret end de dygtige elever. Når læreren fra fortæller sine elever, hvordan testen måler, får de bedre forudsætninger for at have en god oplevelse med testen.

Skolelederen Claus Drachmann fra Aalborg kommune mener ligeledes, at lærerens instruktion til testen spiller en afgørende rolle for elevernes oplevelse. Han siger bl.a., at "*nogle af de opgaver, man får i skolen, er svære at løse. Det skal vi voksne hjælpe med, så selvfølgelig skal vi ikke afvikle test uden, at der har været grundig instruktion inden, så vi sikrer os at eleverne ikke får dårlige oplevelser"* (citeret i Medom 2017).

Vi gav 10 elever fra 4. klasse spørgeskemaer, som de skulle besvare før og efter testen blev afviklet. Spørgsmålene fokuserede på elevernes motivation for at blive testet, de følelser, der opstod inden de skulle testes samt deres forventninger til, hvordan de klarede testen. For at undersøge elevernes motivation stillede vi spørgsmålene "*Hvordan har du det med at blive testet?"* og "*Hvordan har du det lige inden du skal tage en test?".* Herunder

skulle de tage stilling til forskellige udsagn og vurdere, om de var "*helt enig", "lidt enig", "lidt uenig"* eller *"helt uenig*". Eleverne skulle svare på hvilke følelser, de havde lige før testens afvikling. Her har 7 elever sat kryds ved "*helt enig*" og 3 ved "*lidt enig*" til udsagnet "*jeg bliver spændt"*. Det anser vi som en positiv tilgang, da det at være spændt kan forstås som værende forventningsfuld og nysgerrig (som beskrevet i Den Danske Ordbog). Den positive tilgang kan yderligere bekræftes i elevernes besvarelser til udsagnet *"jeg frygter det"*, hvortil alle svarer, at de er hhv. helt eller lidt uenige. Vores indtryk er, at eleverne er nysgerrige efter at vide, hvordan de klarer testen og ikke har noget imod at blive udfordret - 6 elever er helt enige i, at det er rart at finde ud af, hvor godt de klarer sig.

Efter testen fik de samme elever et nyt spørgeskema, denne gang med bagudskuende spørgsmål, som henvendte sig til deres oplevelse af at tage testen. De fleste elever følte sig lettede og glade og få var lidt nervøse efter de havde taget testen - formodentlig for at se resultaterne. Generelt havde eleverne stadig en positiv indstilling til testen, hvilket ikke stemmer i overens med hypotese 2a[9]. Vi kan ikke sige noget generelt om elevers oplevelser af tests på baggrund af denne undersøgelse, da den kun tager udgangspunkt i 10 elever. Det vi til gengæld kan konkludere er, at det ikke er alle elever, der ikke bryder sig om at blive testet.

Eleverne var enten helt eller lidt enige i, at testen var svær, men størstedelen mente, det var rart at blive udfordret. I instruktionen blev eleverne forberedte på at møde spørgsmål, som de ikke ville kunne svare på. Vi kan ikke konkludere, at det er på grund af lærerens instruktioner, at størstedelen af elevernes oplevelser med testen har været positive, men vi kan med sikkerhed sige, at lærerens instruktioner har været i overensstemmelse med vejledningerne, og eleverne har haft en positiv oplevelse med testen. Vi finder det nærliggende at tro, der er en sammenhæng derimellem, hvilket stemmer overens med hypotese 2c[10].

[9] Hypotese 2a: Eleverne bryder sig ikke om at blive testet.
[10] Hypotese 2c: Elevernes oplevelse afhænger af lærerens introduktion til testen - den er også afgørende for deres endelige resultater.

Kritik af DNT

Lærere, elevers og forskeres kritik

For at nærme os en besvarelse af problemformuleringen må vi undersøge de udfordringer, DNT bærer med sig. Her vil vi redegøre for, hvad kritikken af testene fra både lærere, elever og forskere går på, fortolke diverse udsagn og diskutere løbende. Vi har udvalgt fire artikler fra henholdsvis Politiken og Folkeskolen.dk: *"Den nationale test i dansk er den rene slagterbænk for børn"* fra Politikken, skrevet af dansklæreren Anne Madsen, "*Eksperter dumper nationale test"* skrevet af journalist Pernille Mainz, som bl.a. inddrager Peter Allerup og Tomas Højgaard, *"Professor: De nationale test er spild af tid og penge"* af John Villy Olsen (professor, Jeppe Bundsgaard) samt "*Skoleelev: Drop dog de nationale test. De spilder min skoletid og skaber unødig og uproduktiv stress"* af Nikoline Prehn fra 8. klasse, formand for Aalborg Ungebyråd. Artiklerne belyser DNT fra tre forskellige vinkler: Et forsknings-/ekspertperspektiv, et lærerperspektiv og et elevperspektiv. Sådan kan vi nærme os et nuanceret billede af, hvad kritikken går på. Vi vil sammenligne pointerne fra artiklerne med de holdninger, der tilkendegives i hhv. vores spørgeskemaundersøgelse til lærere og en undersøgelse af NordicMetrics. I vores undersøgelse er lærerne blevet bedt om at bedømme, hvor enige de var i forskellige udsagn (se figur 10). Endvidere inddrages udtalelser fra kommentarfeltet i vores undersøgelse. Spørgeskemaet fra NordicMetrics belyser DNT fra et elevperspektiv.

I artiklen "*Eksperter dumper nationale test"* citeres Tomas Højgaard, Ph.D. og lektor på DPU: "*Læreren kan ikke se ud af testen, hvad eleven samlet set trænger til at få gjort mere ud af. Den viden går i realiteten tabt, fordi testen er adaptiv, og det er den viden, der er værdifuld for læreren. I sin nuværende form er man bedre stillet ved at skrotte dem. De gør mere skade end gavn"*. Her stilles skarpt på testens adaptive princip, hvor han påpeger en sammenhæng mellem lærerens muligheder for at undersøge, hvilke områder eleverne skal styrkes på og testens adaptivitet. Det står os ikke helt klart, hvad det er, han mener, da testens adaptivitet ikke afholder læreren fra at undersøge generelle problemer på profilområde-niveau. Hvis han derimod mener, at det kan være vanskeligt at finde ud af, om klassen har et generelt problem med konkrete opgaver inden for et profilområde, siden alle

elever ikke har fået præcis de samme opgaver, så har han jo ret. Som lærer skal man kigge på alle opgaver, eleverne har svaret rigtigt eller forkert på fx under profilområdet *tal og algebra*, hvis man vil undersøge, om de har en udfordring relateret til fx division. Selv da, nytter det ikke ret meget, da læreren samtidig skal forholde sig til opgavernes sværhedsgrad.

Dansklærer, Anne Madsen, fortæller om sin oplevelse med at afvikle DNT med sine 2. klasse-elever. Hendes kritik retter sig mod sværhedsgraden i testen, som hun mener stiller for høje krav til elevernes faglighed og dermed udfordrer elevernes motivation. "*Selvom prøven er adaptiv (tror jeg nok, det hedder) og tilpasser sig elevens niveau, er det i mine øjne en pædagogisk og motivationsmæssig fejl, at testen ikke starter på et overkommeligt niveau og langsomt stiger i sværhedsgrad. Så eleverne får læst sig varme ligesom i en sportskamp"*. Ifølge Madsen bør testen stige i sværhedsgrad, hvilket udraderer det adaptive princip. Vi undersøgte testforløbene i den klasse, hvis resultater, vi har adgang til, hvor første opgave for de 22 elever i 2. klassetesten varierede fra at ligge på 10. percentil til 29. percentil i sprogforståelse. I afkodning varierede det fra 13 til 32 og i tekstforståelse varierede de fra 27 til 63. Percentilen er indikator på sværhedsgrad. Madsen giver et eksempel på en opgave, et testforløb begyndte med: "*Istanbul har en betragtelig befolkning af kurdere, der stammer fra det sydøstlige Tyrkiet. De anerkendes ikke officielt som minoritet, og man indsamler ikke data om, hvor mange de er. - Giv teksten en overskrift."* I et af de testforløb, vi undersøgte, stødte vi på samme spørgsmål, som Madsen nævner - det ligger på 63. percentil. Vi er enige i, at det ikke er hensigtsmæssigt at begynde med et spørgsmål, der ligger så højt i sværhedsgrad. Dog kan vi godt se pointen i, at testen er adaptiv, da man på denne måde nærmere kan danne sig et billede af, præcist hvor eleven ligger.

Eleven, Nikoline Prehn stiller spørgsmålstegn ved formålet med DNT og forstår ikke hvorfor hun og *"500.000 andre folkeskoleelever i Danmark* [...] *hvert eneste år skal spilde flere timer på at lære ingenting".* Hun er ikke kritisk over for prøver og test generelt, men mener ikke, at hun lærer noget af DNT: *"Da jeg i sidste uge gennemførte fysik/kemi-testen, lærte jeg ingenting"*. Det tyder på, at der er en uoverensstemmelse mellem elevens forventninger til testen og testens egentlige formål; Eleven anser testen som et undervisningsmateriale, hun skal lære noget af frem for at se på den som et formativt evalueringsværktøj til læreren. Dog forstår vi godt, at der er denne uoverensstemmelse, da

tests i folkeskolen oftest resulterer i, at eleven får testen tilbage med flueben ud for de rigtige svar. På den måde bliver eleven bevidst om hvad, hun har svaret hhv. rigtig og forkert på - og lærer måske noget deraf. I DNT er der som sagt fortrolighedspligt, hvorfor læreren pr. lov ikke må gennemgå opgaverne med eleven. Det kan virke utilfredsstillende for eleven - især, hvis det ikke er blevet tydeliggjort af læreren fra start.

Af artiklen fremgår det, at flere af elevens lærere omtaler testen som ubrugelig. Det kan have den effekt, at eleverne tager testen useriøst, hvilket kan være af stor betydning for, hvordan de sætter deres krydser, og i sidste ende hvilket resultat, der kommer ud. Det lader til at læreren bruger testen ud fra et *teaching to the test* princip, hvilket yderligere påvirker resultaternes retvisning: "*En uge inden testene kom lærerne farende med adskillige opgaver om svære ordsprog og ord, som for de færreste voksne giver mening.* [...] *På den måde blev vores undervisning de sidste uger inden testene fyldt med såkaldt teaching to the test*". Hvis resultaterne ikke er retvisende, er det vanskeligt at bruge dem i en formativ sammenhæng, hvilket Nikolines lærere tilsyneladende heller ikke gør; "*Heldigvis er jeg heldig. Jeg har gode lærere, som bakker os op. Gode lærere, som fortæller os, at vi ikke skal tage os af de tests, som politikerne ellers har bestemt, vi skal igennem.* [...] *For de gider næsten ikke kikke på resultaterne. De kan ikke bruge dem til noget. De ved godt, hvad vi kan og ikke kan".* Her bliver det tydeligt for os, at elevens lærere er kritisk indstillede til DNT, og ikke lægger skjul på det overfor deres elever.

I kronikken er én elevs perspektiv fremstillet, og som supplement hertil vil vi referere tilbage vores egen undersøgelse af eleverne fra Nærum Skoles holdninger til test, der er overvejende positivt indstillet til testen. Dernæst vil vi inddrage en undersøgelse, der blev lavet på 140 elever fra hhv. 6.,7. og 8. klasse på en skole i Nordjylland i forbindelse med DR-programmet *Folkeskolen Forfra*[11]. Af undersøgelsen fremgår det, at knap end en fjerdedel af eleverne ikke bryder sig om at blive testet, hvor de resterende 76% enten har svaret både/og eller helt eller delvist enig i, at de godt kan lide at blive testet, hvilket belyser vores hypotese 2a[12]. Det samme gør sig gældende ved spørgsmålet om, hvorvidt de er spændte på resultaterne, hvor mindre end 10% har erklæret sig helt eller delvist uenige. Hverken vores eller den omtalte undersøgelse er repræsentativ for alle elever i den danske folkeskole, men vi synes, der er noget interessant ved, at de to undersøgelser viser det samme; at eleverne

[11] Ikke tidligere offentliggjort data indsamlet i 2014 af NordicMetrics.
[12]Hypotese 2a. Eleverne bryder sig ikke om at blive testet.

ikke har så stor modstand til at blive testet, som kritikken ellers ofte går på, og som eleven i kronikken udtrykker.

I artiklen "*Professor: De nationale test er spild af tid og penge*" citeres Jeppe Bundsgaard således: "*Testene giver ikke andet end en meget, meget utydelig strømpil om den enkelte elev. Man kan faktisk ikke vide, om en elev, der ligger helt oppe i toppen, har klaret testen bedre end én, der ligger under middel. Altså en elev, der får 12 point kan have klaret testen lige så godt, som en elev, der får 32 point. Eller omvendt"* (Olsen, 2016). Der skabes her et unuanceret billede af et komplekst regnestykke:

Den adaptive algoritme har større udsving i starten af testens forløb, end i slutningen, da resultatet beregnes på baggrund af alle opgaver. Hvis du kun tager 5 opgaver inden for samme profilområde, betyder hver opgave 20% af den samlede score. Hvis man tager 20 opgaver inden for hvert profilområde, udgør den sidste opgave kun 5% af datagrundlaget. Det betyder at udsvingene bliver meget mindre, når der er større mængde data som følge af en enkelt opgave. Om der er svaret korrekt eller forkert på én opgave er altså af mindre betydning for resultatet, hvis man har besvaret mange opgaver, og elevens niveau kan beregnes mere nøjagtigt. Hvis testen eksempelvis stoppes efter seks opgaver inden for hvert profilområde, kan forskellen på, om man svarer rigtigt eller forkert i det sidste spørgsmål godt betyde en forskel, der svarer til omkring et års læring (et halvt år op hvis rigtigt og et halvt år ned, hvis forkert). Hvor meget, det er i percentiler, kommer an på, hvor på skalaen man er, men forskellen på 12 og 32 er svarende til ca. et års progression - i hvert fald på sprogforståelsesskalaen (NordicMetrics, 2013b). Derudover foreligger den usikkerhed, der gør ved alle evalueringer; har eleven sovet ordentligt natten forinden? Er der lige sket noget i frikvarteret, som distraherer? Elevernes forudsætninger og motivation varierer fra dag til dag.

Vi skal huske på, at på trods af, at DNT fylder meget i den offentlige debat, er det kun et begrænset udsnit af den danske lærerstand, der deltager i debatten. Vi kan ikke sige noget generelt om læreres holdninger til DNT alene på baggrund af kronikker og debatindlæg, hvorfor vi stillede holdningsprægede spørgsmål til lærere i vores undersøgelse. I undersøgelsen har omkring 50 af lærerne efterladt kommentarer, mange af disse udtrykker kritiske holdninger, men der er også få, som ser potentialer i at teste: *"Jeg går klart ind for at benytte test årligt. Vi gjorde det også før de Nationale test kom, da var det blot nogle andre. Test kan aldrig stå alene, og er altid et øjebliksbillede, men sammen*

Lærerprofession.dk - et site for lærerpraksis og professionsudvikling

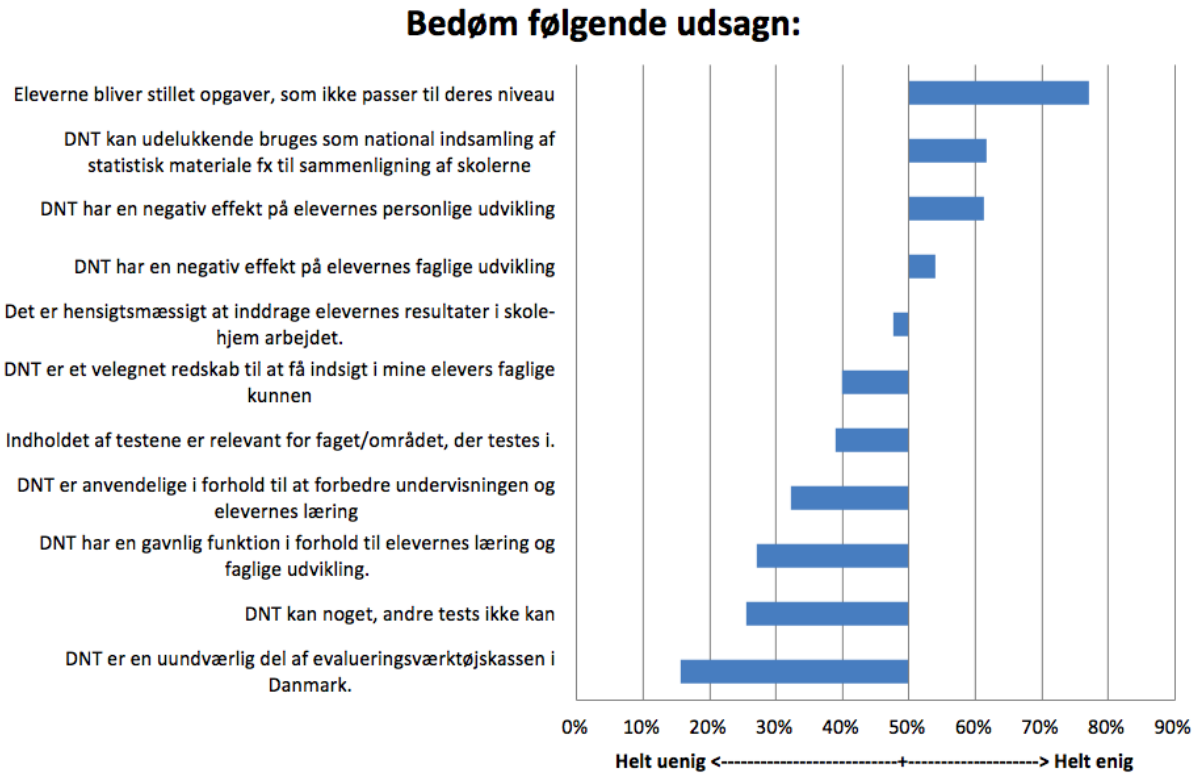
med andre observationer giver det et bidrag til lærerens indsigt i især de meget svage og meget dygtige elevers læseudvikling". Her ser vi en lærer, der er positivt indstillet til at teste og som anser det som en hjælp for læreren.

Nogle af lærerne påpeger også, at de ikke nødvendigvis har et problem med test i sig selv, men at der er nogle af aspekterne ved DNT, som er problematiske - især med fokus på elevernes oplevelser: *"Test er ok, hvis de tilpasses det niveau som eleven er på, uden at præsentere opgaver for meget højere niveau. Det er virkeligt et problem, at i normal undervisningssituation, der lærer vi eleverne at gøre sig umage og at forsøge at svare rigtigt på opgaverne i alle fag. Når de så skal testes i DNT, så skal de "bare" klikke videre og springe over noget de synes er for svært, eller de skal bare gætte på en løsning... Stik modsat det vi normalt forventer af dem*[...]". Dette forstår vi bl.a. som en henvisning til det adaptive aspekt; ideen er jo netop, at testen tilpasses elevens niveau, men for at den kan gøre det, må der også indimellem komme spørgsmål, eleven ikke kan svare på. Derudover påpeger læreren dét, vi allerede har været lidt inde på, nemlig det kulturfremmede i, at eleverne indimellem skal gætte. Dog opstilles der et modsætningsforhold, som vi ikke køber; læreren påpeger, at eleverne til dagligt får besked på at forsøge at svare rigtigt på de opgaver, de møder og stiller det, at de i testen skal gætte, som modsætning hertil. I testen skal eleverne jo stadig forsøge at svare rigtigt - indimellem ved at give et kvalificeret gæt, hvor deres receptive færdigheder kan komme i spil.

Et andet kritikpunkt, lærerne påpeger i kommentarerne, er det tidsmæssige aspekt ved DNT. De mener ikke, at det er muligt for eleverne at besvare så mange opgaver, som fx er anbefalet af beregneren (mindst 60 opgaver – gerne 120) på de 45 minutter, testen gerne skulle tage. En lærer skriver bl.a. "*Beregneren tager udgangspunkt i at eleverne besvarer 60 opgaver. På min skole når kun 38% af alle testede elever de 60 opgaver*" og en anden lærer *"Beregneren skriver i deres vejledning, at der helst skal tages 20 opgaver inden for hvert profilområde. Hvis det skal afholdes, kommer nogle elever til at sidde i over 2 timer. Sluttes testen efter de 40 minutter hvor eleverne kan gå i grønt, risikerer nogle at blive bedømt på omkring 5 opgaver i hvert profilområde, hvilket jeg mener er meget usikkert".* Der er ingen tvivl om, at det må prioriteres, at få et retvisende billede af elevens kunnen for at kunne give feedback og bruge testen formativt. Om man bruger beregneren eller ej, er testen bygget op således, at eleven er nødt til at svare på så mange spørgsmål som muligt for at resultatet bliver retvisende, hvorfor vi igen må understrege vigtigheden i at lærerens

introduktion til testen indbefatter dette. Vi kan ikke på baggrund af kommentarerne sige, hvordan disse lærere har instrueret sine elever, men som vi også pointerede i afsnittet *introduktion til testen* er det essentielt, at læreren ikke blot fortæller, at eleverne må yde deres bedste, men også, at de ikke bør dvæle ved samme spørgsmål i længere tid, men hellere komme et kvalificeret bud og komme videre til næste spørgsmål.

For at undersøge det tidsmæssige aspekt i selve testen nærmere, undersøgte vi et testforløb, hvor en elev havde besvaret 72 opgaver. Her sammenlagde vi den normerede tid for hver opgave og fandt frem til, at forløbet ville have taget knap en halv time, hvis den normerede tid var blevet holdt. Det vil sige, at hvis eleven kun havde brugt den normerede tid, var hun nået igennem 72 opgaver på en halv time. Den normerede tid er - ligesom sværhedsgraderne - regnet ud på baggrund af de 700 elever, som testen er afprøvet på inden implementering. Med den rette instruktion er det altså ikke urealistisk at nå op på mindst 60 opgaver på 45 minutter.



Figur 12 - Besvarelser fra spørgeskemaundersøgelse. Læreres bedømmelser af holdningsprægede udsagn

Ovenstående figur viser i hvor høj grad lærerne er enige i udsagnene, hvor 0% svarer til *helt uenig*, 50% svarer til *hverken eller* og 100% svarer til *helt enig*. Her fremgår det, at der er fire af de 11 udsagn, som lærerne er overvejende enige i; disse fire udsagn er de eneste, der er negativt ladet, hvor resten af udsagnene er positivt ladet. Især har lærerne erklæret sig enige i udsagnet om, at de opgaver, eleverne bliver stillet, ikke passer til deres niveau - her vil vi igen henvise til det adaptive princip, vi før har belyst.

Derudover er lærerne overvejende enige i, at DNT "*udelukkende* [kan] *bruges som national indsamling af statistisk materiale fx til sammenligning af skolerne"*. Her finder vi det relevant at referere til vores hypotese 1b[13]. Ordet *udelukkende* er i dette udsagn afgørende. Det er bl.a. med til at implicere en skjult dagsorden ovenfra, da regeringen har gjort DNT obligatoriske, hvor det formative formål anses som skalkeskjul fra politiske niveauer, der kun er interesserede i at indsamle statistisk materiale. Det, at der ikke er tillid til den politiske brug af testresultaterne, ser vi som et kæmpe problem. Det er i sig selv med til at gøre testene ubrugelige (i hvert fald ift. det formative formål), hvis dem, der skal afvikle testen og anvende resultaterne i praksis, føler de gør et stykke arbejde, de ikke kan stå inde for. Det et redskab, der som udgangspunkt er tiltænkt lærerne, og hvis de ikke føler, at DNT kan fungere som en hjælp i praksis, må noget ændres.

De sidste to udsagn, lærerne er overvejende enige i er, at DNT har en negativ effekt på hhv. elevernes personlige og faglige udvikling. Dog er der større enighed om, at de har en negativ effekt på elevens personlige udvikling frem for faglige udvikling. Hvorfor det er sådan, kan vi kun gisne om, men vi forestiller os, at der igen refereres til elevernes oplevelser af testen, da flere deltagere i den offentlige debat antyder, at eleverne får en følelse af at være utilstrækkelige og dermed et hak i selvtilliden og motivationen for faget, når de udfordres på den måde, som de gør i DNT. Ift. den faglige udvikling, kan besvarelserne måske knyttes an til den tid, der bruges på DNT, som kunne være brugt mere konstruktivt, hvis resultaterne alligevel ikke bliver brugt til noget. Summa summarum er, at en overvejende del af de responderende lærere mener, at DNT har en negativ effekt på elevernes faglige og personlige udvikling, hvilket stemmer overens med besvarelserne på resten af udsagnene, hvor lærerne erklærer sig mere eller mindre uenige i de udsagn, der er positivt ladet. Alt i alt er vores indtryk, at de overordnede kritikpunkter kan kategoriseres

[13] Hypotese 1b: Lærerne ser testene som et udtryk for mistillid fra det politisk niveau (Ministeriet og kommunerne)

ved følgende: 1) testens adaptivitet, 2) det tidsmæssige aspekt, 3) elevernes oplevelse af at blive stillet svære spørgsmål, 4) anvendelsesmuligheder for læreren og 5) regeringens anvendelse af resultaterne.

Vores kritik af DNT

Vi har i ovenstående forsøgt at imødekomme og diskutere de kritikker, der hidtil har været fra både forskere, lærere og elever og vil i følgende afsnit opstille egne kritikpunkter af testen.

- Resultaterne kommer ud til læreren i en ufortolkelig form. Tallene i testene er usammenlignelige, da hver test måler på forskellige ordinalskalaer.
- Der kan være elever som mister tålmodighed og koncentration under testen, eksempelvis børn med ADHD eller andre diagnoser med koncentrationsbesvær som følge, der besværliggør det at skulle sidde stille i minimum 40 min - hvor andre har langt bedre forudsætninger for at tage testen.
- Især de dygtigste og svageste elever vil ofte skulle bruge meget længere tid, før testen går i grøn, da testsystemet vil have svært ved at lokalisere elevens hhv. øvre og nedre faglige barre. Det vigtigste her er, at eleverne får besvaret et højt antal opgaver.
- DNT i de naturvidenskabelige fag er uhensigtsmæssige. Der er meget stor forskel på elevers viden indenfor det felt - enten kan man det, ellers kan man det ikke (Entwistle, 2005: James 2009, s. 45). Derfor er det svært at lave noget fornuftigt med resultaterne, da de er mere emne-bestemte (fx celledeling, evolution eller tyngdekraft) end fx læsning. I matematik og dansk er der grundlæggende regler, som gør det muligt at løse opgaverne.
- I og med, læreren har adgang til sine elevers testforløb, er det muligt for læreren at samle en stor opgavebank på baggrund af én national test. Det er som tidligere nævnt ulovligt at anvende testene som undervisningsmateriale, men det bliver fristende, når regeringen udlover millionbonusser til 100 dårligt præsterende skoler i Danmark, hvis de løfter deres elevers resultater i DNT tilstrækkeligt. At opstille en økonomisk gevinst ved gode testresultater kan give incitament til *teaching to the test* og manipulation af resultater, jf. Campbell's Lov (Nichols m. fl., 2007).

- I testsystemet får læreren fire forskellige resultater; et for hvert af de tre profilområder og en samlet vurdering, som er et gennemsnitstal af de tre andre. Det strider imod grundlæggende matematiske regler at regne gennemsnit på ordinalskalaer. De skal først omregnes til en intervalskala, hvor hvert point er lige meget værd.
- I matematiktesten er der mange tilfælde, hvor opgaven skal besvares med et konkret beregnet resultat, som eleven selv skal skrive ind i en tekstboks. Det giver ikke mulighed for at afprøve overslagsregning, hvilket besværliggør det tidsmæssige aspekt i, at eleverne skal besvare mange opgaver. I matematik giver det faktisk mening at kigge på opgaverne i længere tid, fordi man kan komme tættere på svaret ved at regne sig frem til det. Det kunne man undgå ved at lave multiple-choice.

Vi ser ovenstående punkter, som nogle områder, hvorpå DNT er vanskelige at arbejde med i praksis. Dog mener vi ikke, det gør testen ubrugelig, vi ser det nærmere som nogle punkter, man må være opmærksom på ift. brugen af testen. Vores oplevelse er, at der i de fleste evalueringsværktøjer er muligheder og begrænsninger og ved at være bevidst herom, kan man bruge værktøjet eller metoden til det, som det er beregnet til.

Ligesom DNT er det de færreste evalueringsredskaber, som kan stå alene ift. at vurdere den enkelte elev, men testen kan give et fingerpeg om elevens faglighed og bør være et supplement til læreres fagprofessionelle vurdering.

Diskussion

Diskussionen om DNT er kompleks med flere dimensioner, hvor vi har valgt at fokusere på to facetter: Dels et ideologisk aspekt, som består i, at test som policyinformation og styringsredskab ofte forbindes med et bestemt menneske- og læringssyn, og dels et praktisk aspekt, hvor testen anses som en række opgaver, eleven har svaret rigtigt eller forkert på, og som kan give information om, hvad eleven kan og har brug for undervisning i. De to aspekter har et dialektisk forhold, og vil blive diskuteret således det følgende afsnit.

Ser man på undervisning i et historisk perspektiv, er man i Danmark gået fra et behavioristisk læringssyn til et socialkonstruktivistisk læringssyn som præger diskussionen; Behaviorismen mener, at adfærd reguleres ved ydre påvirkninger og stimuli, og viden på

samme måde kan 'overføres' fra ét hoved til et andet bl.a. gennem repetition (Løw & Skibsted, 2012, s. 171). Det konstruktivistiske syn tager udgangspunkt i teorien om, at læring ikke alene består af tilvækst i færdigheder og/eller kundskaber, der kan overføres, men er en aktiv proces, hvor mennesket selv er med til at konstruere viden og erkendelse.

Det er nærliggende at forbinde test med behaviorisme, da videnskabsteorien bygger på en analytisk filosofisk tilgang med vægt på dataindsamling. At man kan bedømme dygtighed alene ud fra, hvor eleven sætter krydsene, antages sædvanligvis som udtryk for et behavioristisk læringssyn. Der kan ikke sættes lighedstegn mellem et redskab (test) og en filosofisk retning (behaviorisme), hvorfor testen i sig selv ikke kan kategoriseres som værende hverken det ene eller det andet. Test er et redskab, som først kan forbindes med en videnskabsteoretisk orientering, når vi ser på anvendelsen af resultaterne. Bruger man testen som grundlag for belønning (fx i form af ros), straf eller resultatløn, kan det siges at være præget af et behavioristisk læringssyn. Hvis de derimod bruges som kvalificering af lærerens pædagogisk didaktiske overvejelser, er udgangspunktet i højere grad i familie med konstruktivismen.

Begrebet *evaluering* blev for første gang introduceret i dansk folkeskolelov i 1993 (Folkeskoleloven §13), hvilket var udtryk for et gradvist paradigmeskift: Det var starten på en proces, hvor det politiske søgelys i stigende omfang blev flyttet fra klassen og input (læreren, undervisningen, antal lektioner) til at fokusere på individ og outcome (elevens udbytte, læring, kvalifikationer). Derved blev evaluering en nødvendig del af lærerens praksis, og med det individuelle fokus, blev test et mere udbredt evalueringsredskab. Det var ikke længere kun folkeskolens ansvar at undervise eleverne, det lå også inden for folkeskolens og lærerens ansvar, at eleverne faktisk lærte noget, og at det kunne måles. Dette gradvise paradigmeskift er stadig på vej, hvilket kan eksemplificeres i *fælles mål*, som er gået fra at være beskrevet som undervisningsmål (Fælles Mål 2003 og 2009) til at være formuleret som læringsmål (Forenklede Fælles Mål i 2014).

I forbindelse med den danske evalueringskulturs udvikling (kickstartet af OECD, 2004), begyndte regeringen at fokusere på test som evalueringsværktøj (Regeringsgrundlaget 2005, s. 17). I den forbindelse blev DNT etableret, med det formål at fungere som en hjælp til læreren (jf. folkeskoleloven bemærkninger i 2006). Implementeringen af DNT har dog affødt stor kritik fra mange læreres side, da der er uenighed om den pædagogiske værdi. Et klart flertal af de lærere, der deltog i vores

undersøgelse mener, at det er uhensigtsmæssigt at udsætte elever for en sådan test og oplever kun i lav grad testen som en hjælp til at kvalificere deres praksis (se figur 12 på s. 30). Alligevel er testene obligatoriske, og lærere er dermed tvunget til at bruge værktøjet, uanset om de finder det brugbart eller ej. Det leder til antagelser om skjulte dagsordner, for når lærere nu tilkendegiver, at testen ikke er en hjælp, hvem får så gavn af resultaterne? Vi må igen understrege, at vores undersøgelse ikke er repræsentativ og sandsynligvis med stærk bias. Andre undersøgelser påviser en langt mere positiv indstilling overfor DNT. Fx Rambølls evaluering af DNT, hvor flertallet af lærerne mener, at testen giver dem overblik over eleven og klassens faglige niveau, og meget få lærere siger at de ikke anvender DNT ved tilrettelæggelse af undervisningen.

"Tænk tilbage på, da du fik resultaterne af de nationale test i foråret 2012. Hvad har du brugt resultaterne af de nationale test til?"

Angivet i pct.	Slet ikke	I mindre grad	I nogen grad	I temmelig høj grad	I meget høj grad	I alt (N=2160)
Til at få et samlet overblik over **klassens** faglige niveau	2%	8%	32%	41%	17%	100%
Til at få et overblik over de **enkelte elevers** faglige niveau	2%	11%	34%	38%	16%	100%
Til at tilrettelægge en undervisning, der tager udgangspunkt i den **enkelte elevs** faglige niveau	9%	22%	42%	23%	5%	100%
Til at tilrettelægge en undervisning, der tager udgangspunkt i **klassens** faglige niveau	9%	21%	44%	22%	4%	100%

Figur 13 - Spørgeskemaundersøgelsen til lærere (Kilde: Rambøll, 2013, tabel 4.1 og 4.5)

Evaluering i Danmark er i høj grad betinget af læreren, der evaluerer. Læringsmålene i fælles mål kan gradbøjes og er ikke målbare (eks. "*Eleven kan vurdere teksters perspektiv på et emne"*), hvorfor læreren må nedbryde dem yderligere til konkrete mål med faglige niveauer, som man kan vurdere, om er nået. Ligeledes er 7-trinsskalaen til fri fortolkning, fx "*Karakteren 7 gives for den gode præstation, der demonstrerer opfyldelse af fagets mål,*

med en del mangler" (Holm, 2017a) - hvordan defineres hhv. "*en del mangler"* og *"fagets mål"*, hvis både fagformål og fælles mål kan gradbøjes?

To journalststuderende lavede et eksperiment i 2013, hvor de hver sendte en stil, de havde skrevet, ud til 20 dansklærere, som skulle give dem karakterer. Lærerne, der ikke på forhånd kendte forfatterne bag stilene, vurderede begge stile vidt forskelligt - den ene stil fik bedømmelser, der varierede fra 4-12 og den anden fik bedømmelser, der varierede fra 02-12 (Politiken, 2013).

Hertil kan DNT bidrage med en måling, hvor eleverne bedømmes, på grundlag af deres svar, af en matematisk model, og læreren får et indblik i, hvordan eleven ligger sammenlignet med resten af klassen og landsgennemsnittet. Har eleven levet op til sit bedste, får læreren et mere neutralt grundlag for bedømmelsen og for evt. at give eleven karakterer. Det er ikke nødvendigvis hensigtsmæssigt at sammenligne landets elever, men sådan er det gældende karaktersystem skruet sammen (Holm, 2017b). Et godt argument for, at testene er obligatoriske er, at alle elever efter vores vurdering har ret til en bedømmelse, der består af mere end lærerens subjektive vurdering. Dette skal ikke forstås som en underminering af lærerens fagprofessionelle, løbende vurdering af eleven, da denne stadig er mindst ligeså vigtig som testen, men måske udelukker det ene ikke det andet. Vores indtryk er, at debatten er præget af en forståelse om, at det enten hedder *kun* test eller *kun* lærerens subjektive evaluering. Som vi ser det, er de to evalueringer ikke uforenelige. Testen skal netop fungere som supplement til lærerens helhedsvurdering af eleven.

Gert Biesta taler om en stigning i interesse for måling inden for evalueringskulturen, som har haft indflydelse på uddannelsesområdet, hvor bl.a. test afspejles og anses som en måde at effektivisere samfundet, i denne forbindelse uddannelsessektoren, på (2009). I forlængelse af det, er det interessant at kigge på den danske regerings anvendelse af testen. Her kan regeringens nationale mål for folkeskolen om, at 80% af eleverne skal score *god* eller derover i DNT (UVM 2017b, s. 20) med fordel inddrages. Her sættes fokus på elevernes slutresultat frem for deres progression. Samtidig sættes samme mål for alle danske folkeskoleelever, uanset deres udgangspunkt. Dette kan sammenlignes med, at man vurderede elevers evner til at hoppe højdespring over samme overligger, selvom den ene elev er 1.50 høj og den anden 1.90. Eleverne har ikke lige forudsætninger for at nå samme slutresultat, da der er stor forskel i elevgrundlag, hvilket er mere afgørende for det gennemsnitlige niveau i DNT end undervisningen (Hattie, 2014)

Biesta fokuserer overordnet på den politiske anvendelse af (inter-)nationale testsystemer og hævder, at politiske instanser misbruger testresultater til sammenligning og er således konkurrenceorienterede, hvilket også afspejles i ovenstående eksempel om regeringens 80%-tiltag. Han argumenterer for, at testene hovedsageligt bruges til at identificere såkaldte "*failure schools"* og i nogle tilfælde også "*failure teachers"*. Dette sker ofte under banneret at 'hæve standarden' for uddannelsessektoren (2009, s. 1). Derudover diskuterer han, hvad der er ønskværdigt at uddanne eleverne til, hvor værdier som fx. demokratisk dannelse inddrages som en vigtig faktor. Med den testkultur, der er i fuldt flor og stadigt voksende, mener han, at fokus fra sådanne værdier negligeres. Han stiller i den forbindelse et vigtigt spørgsmål om, "[...]*whether we are indeed measuring what we value, or whether we are just measuring what we can easily measure and thus end up valuing what we (can) measure*" (2009, s. 2). Her indikeres, at det, der måles på, i sidste ende vil være det, der værdsættes mest i uddannelsessektoren, hvorfor testkulturen kan have en skadelig effekt på uddannelsessystemet og elevernes dannelse. Biestas pointe om, at værdier som fx demokratisk dannelse bør prioriteres i uddannelsen, understøttes i den danske folkeskoles formålsparagraf, som er en af de få i verden, der er to-delt. Her er både fokus på elevens kundskaber og faglige færdigheder (Folkeskolens formålsparagraf, første linje) og på den personlige, sociale og alsidige udvikling (Resten af folkeskolens formålsparagraf). Det tiltag, regeringen har lavet, som går ud på at belønne dårligt præsterende skoler økonomisk, hvis de løfter deres score i DNT tilstrækkeligt (Olsen, 2011), ser vi som en fin eksemplificering af Biestas pointe om risikoen for, at folkeskolens øvrige formål negligeres til fordel for det, der måles gennem test og prøver. Man kunne i den forbindelse undre sig over, hvorfor det kun er demonstration af faglig kunnen, der udløser karakterer i folkeskolen, når formålsparagraffen vægter det andet så meget tungere? Man kan også undres over, at der er så meget polemik om DNT (hvor formålet er formativ og low-stake[14] anvendelse – og de er fortrolige), når der tilsyneladende ingen modstand er mod folkeskolens afsluttende prøver (som er summative og high-stake[15] - og hvor resultaterne offentliggøres)?

[14] Tests kan kategoriseres som low-stakes, når testen er *"en del af den interne evaluering og uden konsekvenser for* [fx] *skoler og kommuner"* (Andersen m.fl. 2017, s. 48)

[15] Tests kan kategoriseres som high-stakes, "*når bedre testresultater bliver vigtige at opnå af andre end rent faglige hensyn, fx ved at skolerne, leder- ne, lærerne eller eleverne opnår fordele ved bedre testresultater*" (Andersen m.fl. 2017, s. 48)

I formålsparagraffen har man forsøgt at retningsangive, hvilke værdier vi bør bestræbe os på at undervise ud fra og bidrage til, hvor bl.a. elevens *lyst til at lære mere* nævnes (Folkeskolens formålsparagraf §1). Dette mener John Hattie er muligt at gøre ved at synliggøre elevernes progression for både lærere og elever - især gennem test. Han bekender sig til evidens (baseret på vurdering af elevers besvarelser af opgaver uden lærervurdering) og mener, at læring kan måles – han anser progression og læring som synonymer. Når progression måles med test, bliver noget, der ellers er usynligt (læring) observerbart for begge parter. Det giver læreren bedre mulighed for at anerkende og synliggøre elevers fremskridt - også hos dem, hvor det ikke er tydeligt til hverdag. Således kan eleverne følge med i deres egen udvikling, og det mener Hattie kan øge motivationen for at lære mere. Det stiller dog krav om, at de skalaer eleven får sine resultater på er egnet til formålet, og her er percentilskalaer (som anvendes i DNT) ikke altid egnede (Andersen m.fl., 2017). Eksempelvis vil en elev, der har scoret 1 i én national læsetest og 10 i den næste i begge omgange få resultatet *utilstrækkelig*, selvom progressionen her er, hvad der svarer til 3 læringsår. Det er ikke fjernt at forestille sig, at elevens skole/hjem-samtale på baggrund af kriteriescoren ville lyde således: "*Du scorede utilstrækkelig sidste år, og du scorede utilstrækkelig igen i år. Vi skal til at finde en måde, vi kan løfte dit faglige niveau på"*. Hvis man i stedet får mulighed for at anerkende eleven for udviklingen, skabes der grundlag for at arbejde med elevens (selv-)narrativer i en positiv retning, frem for at fastlåse sig på, at eleven ligger på niveauet *utilstrækkelig* (Løw, 2009)[16]. Derudover er der mulighed for at lokalisere elever, der "flatliner" (ikke har nogen udvikling) og således skabe grobund for at udfordre disse og bidrage til, at eleven udvikler/fastholder gode arbejdsvaner. Især hos de dygtigste elever kan det være svært at gennemskue udviklingen eller manglen herpå. Hvis vi antager, at en elev har ligget på 10. klasseniveau siden 6. klassetrin og således scorer *fremragende* i samtlige tests, vil dette højst sandsynligt udløse ros, som ikke nødvendigvis afspejler deres arbejdsindsats.

Når testen er afsluttet, kommer resultatet som nævnt i form af en percentil og en kriteriescore, som i sig selv er endimensionelle størrelser, hvilket kan være årsagen til en generel opfattelse af, at resultatet kun består heraf. Hvis man som lærer vil anvende testresultaterne til at få et mere detaljeret indblik i elevens faglighed, kan man imidlertid

[16] Hypotese 1d: Evalueringen kan fungere som en støtte for læreren i forbindelse med at løsrive sig fra sine narrativer om eleverne (Hattie 2014, Løw 2009).

analysere de besvarelser, der ligger til grund for resultatet - dvs. den serie af opgaver, som eleven har svaret enten rigtigt eller forkert på. Heri ligger der et fortolkningsarbejde, som kan give læreren større indsigt i elevens faglighed, hvis man altså ved, hvad man skal kigge efter. Hertil er især én kommentar fra vores undersøgelse relevant at inddrage: "*Det kunne være interessant at vide, hvor mange lærere, der er uddannet/har taget kurser i at læse DNT. Jeg har aldrig modtaget nogen form for uddannelse til at læse/forstå dem"*. Der er altså dårlige forudsætninger for de lærere, der, ligesom ovenstående, ikke har modtaget nogen form for uddannelse i fortolkning af testresultaterne, for at benytte testene til deres fulde potentiale.

Hattie er ofte blevet nævnt i forbindelse med den nye folkeskolereform, men det er ikke på alle punkter, der har været overensstemmelse med Hatties ideer og regeringens behandling og anvendelse af data, eksempelvis i kvalitetsmålingerne, hvor testresultater for fx en 4. klasse i 2015 sammenlignes med testresultaterne for 4. klasse 2016. Det er altså forskellige elever, der sandsynligvis har haft forskellige lærere, der sammenlignes. Det er misvisende at inddrage dette i en kvalitetsrapport, da den sammenligning hverken kan fortælle om kvaliteten af skolen eller af undervisningen. Det eneste vi får informationer om er, i bedste fald, årgangenes elevgrundlag.

Vi oplever indimellem, at der, i diskussionen omkring test, bliver opstillet et modsætningsforhold, der går ud på, hvorvidt man skal have fokus på det faglige niveau (kognitive færdigheder) eller børnenes trivsel (non-kognitive færdigheder). Vi ser dog ingen grund til, at de ikke kan gå hånd i hånd. International forskning viser, at det hænger nøje sammen – glade og trygge børn lærer bedst, motivation og selvtillid styrker læring, og elever, der lærer, bliver mere tilfredse med sig selv (Heckman 2008, Duckworth 2007). Også herhjemme er det er en udbredt teori, at en af forudsætningerne for elevernes faglige og personlige udvikling er, at de trives i skolen (Illeris, 2007). Modsætningsforholdet mellem det kognitive og non-kognitive skriver sig ind i en større debat om dannelse versus uddannelse, som Knud Romer udtalte sig om i debatten 2011: "*Man har fjernet dannelse fra uddannelse* og *hvad har man tilbage? Ud! Ud på arbejdsmarkedet. Ud og tjen penge. Ud, ud, ud"* (Citeret i Aastrøm, 2011). En holdning, der ofte repræsenteres i den offentlige debat, hvor det hævdes, at uddannelsespolitikernes formål med den danske folkeskole i højere grad er, at skabe arbejdsmus, der kan træde ind i landets økonomiske hamsterhjul, end det er at fremme nysgerrighed og kritisk stillingtagen hos eleverne (Aastrøm, 2011).

Lærerprofession.dk - et site for lærerpraksis og professionsudvikling

I formålsparagraffen er indikeret en årsag/virkning-effekt mellem faglige færdigheder og kundskaber (det kognitive) og hvad vi kategoriserer som dannelse eller non-kognitive færdigheder: "*Folkeskolen skal i samarbejde med forældrene give eleverne kundskaber og færdigheder, der: forbereder dem til videre uddannelse og giver dem lyst til at lære mere, gør dem fortrolige med dansk kultur og historie, giver dem forståelse for andre lande og kulturer, bidrager til deres forståelse for menneskets samspil med naturen og fremmer den enkelte elevs alsidige udvikling*". Det dannelsessyn, indbefatter altså at folkeskolen via faglige kundskaber og færdigheder skal bidrage til elevernes dannelse, livsfærdigheder og personlige færdigheder, der bl.a. gør dem i stand til at håndtere at blive udfordret til grænsen af deres formåen. International forskning (fx Heckmann 2008) peger på en omvendt årsag-virkning sammenhæng: Man skal starte med det non-kognitive (dannelse, livs- og personlige færdigheder fx selvtillid, resiliens, motivation og vedholdenhed) og derigennem blive god til at erhverve faglige kundskaber og færdigheder. Man skal lære at møde til tiden, holde sine aftaler, ikke at give op ved modgang, hjælpe andre, være kreativ, positiv og samarbejdsorienteret for at blive god til at tilegne sig ny viden – at lære.

I relation til DNT vil det konkret betyde, at hvis eleverne har veludviklede non-kognitive evner, uden fornemmelsen af nederlag kan besvare DNT, hvor opgaverne netop fører dem til grænsen af deres formåen og indimellem også udover det, hvilket går godt i spænd med flere anerkendte læringsteorier, bl.a. Vygotskys *zone for nærmeste udvikling*, hvor det bl.a. handler om, at udfordre eleverne der, hvor de er rent fagligt (Vygotsky, 1987).

Konklusion og handleperspektiv

I arbejdet med denne opgave har vi beskæftiget os med spørgsmålet om, hvilke pædagogiske udfordringer og potentialer, der ligger i arbejdet med DNT. I løbet af processen har vi belyst emnets kompleksitet både ift. afvikling, fortolkning og anvendelse af test og deres resultater. Når der implementeres et komplekst redskab, vil der normalt også blive afsat ressourcer til at uddanne dem, der skal bruge det på den ene eller anden måde. Man kan ikke forvente, at lærerne på forhånd ved, hvordan testen skal bruges, hvis de ikke har fået det at vide. Hvis redskabet ikke anvendes hensigtsmæssigt, er det tvivlsomt, om det kommer til gavn for læreren og/eller eleverne. Vores indtryk er, at der i den offentlige debat

er tendens til at fokusere på, den politiske kontekst, frem for at forholde sig anvendelsesorienteret og undersøge, om den egentlig har potentiale som et pædagogisk redskab for læreren. Den politisk orienterede diskussion kan føre til urimelig kritik, fx at læsetesten hverken kan vise elevernes analytiske evner eller kammeratskab (hvilket er sandt men irrelevant, da det aldrig har været intentionen).

Den kritik, vi har mødt i løbet af processen, må ikke underkendes – det er vigtigt, at man som lærer ytrer sine holdninger og ikke blot underlægger sig tiltag, man ikke finder brugbare. Ministeriet har stillet sig selv en opgave om at få implementeret DNT, som de efter vores vurdering ikke har løst ordentligt. De modarbejder fx testens formål ved at tilbyde økonomiske bonusser til dårligt præsterende skoler, ved at fastsætte fælles nationale mål for slutpunkt for samtlige danske folkeskoleelever og ved at tvinge et redskab over hovedet på lærerne, der som udgangspunkt er tænkt som en hjælp. Vi har dog en formodning om, at hvis lærerne får viden om, hvordan testene virker og kan bruges i praksis, samt hvordan resultaterne fortolkes og anvendes, vil modstanden mod testene som pædagogisk redskab kunne reduceres eller forsvinde helt, mens lærerprofessionens fagpolitiske modstand mod testens som kontrolredskab nok aldrig forsvinder.

Først og fremmest skal lærerne have tillid til, at testen måler noget relevant. Uanset hvad og hvordan der måles, er der fejlkilder, som kan føre til misvisende konklusioner. Disse fejlkilder kan principielt findes tre forskellige steder i processen, når man måler. Der kan enten være fejl i måleredskabet, i anvendelsen eller i fortolkningen. Test kan sammenlignes med en blodtryksmåler: hvis den ikke fungerer ordentligt fra start, er resultatet ikke retvisende. Ligeledes vil man få et misvisende resultat, hvis man monterer blodtryksmåleren forkert eller hvis man ikke er i stand til at fortolke tallet, der kommer ud i den anden ende. Selve testen fungerer fint som måleredskab, hvis den anvendes korrekt; eleverne får den rette instruktion og besvarer tilstrækkeligt antal af opgaver. Derfor må vi konkludere, at i de tilfælde, hvor resultaterne eller tilbagemeldingen til eleven ikke er retvisende eller afspejler deres dygtighed, må fejlkilden ligge i anvendelsen eller fortolkningen. I makroperspektiv ser vi disse muligheder for DNTs fremtid: Enten skal 1) kommuner og ministeriet afsætte endnu flere ressourcer (tid og penge) til at lærere og andre fagfolk, folkeskolens organisme består af, kan blive fortrolige med et komplekst redskab, 2) testen gøres frivillige, med risiko for, at elever får misvisende tilbagemeldinger på testresultatet, eller 3) testen afskaffes helt og vi

går således tilbage til, at evalueringerne udelukkende afhænger af hvilken lærer man nu engang ender med at få.

Vi er af den overbevisning, at UVM i bund og grund har samme interesse som de ildsjælde, der deltager i debatten om pædagogisk anvendelse af DNT - at fremme elevernes læring og personlige alsidige udvikling. Vi har ikke selv beskæftiget os med DNT i lærerpraksis, men det skal vi inden længe, hvorfor vi mener, det er vigtigt at finde frem til et handleperspektiv, der kan hjælpe os med at bruge testen mest hensigtsmæssigt. For os handler dette grundlæggende om 2 ting: At udnytte de potentialer, vi ser i DNT som evalueringsværktøj og at sikre det optimale grundlag for anvendelsen af testen.

At udnytte DNTs potentialer kræver mulighed for en mere detaljeret analyse af testresultaterne for den enkelte elev. Vores bedste bud herpå er at anvende et program som 'Beregneren', der imødekommer den største praktiske udfordring, vi hidtil er stødt på; at resultaterne i testsystemet er ufortolkelige. Beregneren omregner percentilerne til Raschscore - dvs. resultaterne går fra at være på en ordinalskala til en intervalskala. Her fokuseres der på at lave en ipsativ reference, hvor eleven sammenlignes med sig selv og testen kan således anvendes som et pædagogisk redskab i overensstemmelse med Hatties teori om synlig læring (2014).

For at sikre det optimale grundlag for anvendelse af testen, således at testresultaterne er retvisende og dermed kan anvendes i en formativ evaluering af den enkelte elev, er det afgørende at være omhyggelige med vores instruktioner til - og italesættelse af testen.

Litteraturliste

- **Aastrøm, M.I.** (2011) *Hvad er uddannelse uden dannelse?* Lokaliseret 21. maj 2017 på Politiken: http://politiken.dk/debat/art5478420/Hvad-er-uddannelse-uden-dannelse

- **Andersen M., Wahlgren B. & Wandall J.** (2017) *Evaluering af Læring, Undervisning og Uddannelse*, København: Hans Reitzel (in press)

- **Biesta, G.** (2009). Good education in an age of measurement: On the need to reconnect with the question of purpose in education. *Educational Assessment, Evaluation and Accountability (formerly: Journal of Personnel Evaluation in Education)*, *21*(1), 33-46.

- **Brandt, A.** (2009) *Det åbne samfunds ven* I: M. Hansen (red.) 50 Samfunds tænkere (s. 391-402), 1. udgave, 1. oplag. København: Gyldendal

- **Bundsgaard, J., & Puck, M. R.** (2016). *Nationale test: Danske lærere og skoleledere brug, holdninger og viden.*

- **Drachmann, C.** (2015) *Jakob Wandall på Tårnborgkonferencen 2015*. Lokaliseret 15. Maj 2017 på SlideShare: https://www.slideshare.net/drachmann1/wandall-taarnborg-off

- **Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R.** (2007). Grit: perseverance and passion for long-term goals. Journal of personality and social psychology, 92(6), 1087.

- **Dudley, P., & Swaffield, S.** (2008). *Understanding and using assessment data. Unlocking assessment: Understanding for reflection and application*. Sue Swaffield (Edit)

- **EMU**. *Nationale test*. Lokaliseret http://www.emu.dk/modul/nationale-test-s%C3%A5dan-l%C3%A6ser-du-resultaterne

- **Entwistle, N.** (2005), Learning outcomes and ways of thinking across contrasting disciplines and settings in higher education. *The Curriculum Journal, 16 (1):* 67-82.

- **Hattie, J.** (2014). *Synlig læring-for lærere*. 1. udgave, 6. oplag, Fredericia: Dafolo A/S

- **Heckman, J. J.** (2008). Schools, skills, and synapses. Economic inquiry, 46(3), 289-324.

- **Holm, V.** (2017a) *Karakterer på 7-trinsskalaen*. Lokaliseret 24. maj 2017 på UVM: http://www.uvm.dk/uddannelsessystemet/7-trins-skalaen/karakterer-paa-7-trins-skalaen

- **Holm, V.** (2017b) *Anvendelse af 7-trinsskalaen*. Lokaliseret 24. maj 2017 på UVM: http://www.uvm.dk/uddannelsessystemet/7-trins-skalaen/anvendelse-af-7-trins-skalaen

- **Hummelgaard, H.** (2015) *Formålet med PISA, og hvordan PISA 2015 foregår?* Lokaliseret 23. maj 2017 på Danmarks Statistik: http://www.dst.dk/extranet/pisa2015/html/ac5e94f5-ab5d-48d2-92c7-f66f49a7ca3d.htm

- **Illeris, K.** (2007) *Barriere mod læring - læringsteoriens elementer hvordan hænger det hele sammen*, Roskilde Universitetsforlag, s. 165-184

- **James, M.** (2016). Evaluering, undervisning og læringsteorier. *CEPRA-striben*, (04), 44-57.

- **Løw, O.** (2009). *Lærerens fortællinger om elever, pædagogisk analyse og ledelse af skoleklasser* I: Jensen, E. & Løw, O. (red.) Klasseledelse; nye forståelser og handlemuligheder (s. 40-53). København: Akademisk forlag

- **Løw, O. & Skibsted, E.** (2012) *Psykologi for lærerstuderende og lærere*, København.Akademisk forlag

- **Madsen, A.** (2014) *Den nationale test i dansk er den rene slagtebænk for børn*. Lokaliseret 23. maj 2017 på Politiken: http://politiken.dk/debat/art5621183/Den-nationale-test-i-dansk-er-den-rene-slagteb%C3%A6nk-for-b%C3%B8rn

- **Mainz, P.** (2016) *Eksperter dumper nationale test.* Lokaliseret 25. maj 2017 på Politiken: http://politiken.dk/indland/uddannelse/art5607621/Eksperter-dumper-nationale-tests

- **Medom, M.** (2017), *Trods Aalborgs kursændring om nationale test: Skoleleder vil stadig bruge 'Beregneren'*. Lokaliseret 17. maj 2017 på folkeskolen.dk: https://www.folkeskolen.dk/608147/trods-aalborgs-kursaendring-om-nationale-test-skoleleder-vil-stadig-bruge-beregneren

- **Mortimore, P., David-Evans, M., Laukkanen, R., & Valijarvi, J.** (2004), *OECD-rapport om grundskolen i Danmark.* Lokaliseret 25. maj 2017 på UVM: http://static.uvm.dk/Publikationer/2004/oecd/

- **Nichols, S. L., Berliner, D. C., & Noddings, N.** (2007). Collateral damage: How high-stakes testing corrupts America's schools.

- **NordicMetrics** (2013a), *Appendiks 1: Om baggrund og teori bag valg af skala*. Lokaliseret 19. maj 2017 på NordicMetrics: http://www.nordicmetrics.com/wp-content/uploads/2015/08/Appendiks_1_baggrund-og-skalaer.pdf

- **NordicMetrics** (2013b), *Vejledning til gennemførelse af test: Hvad enhver lærer, som gennemfører nationale læsetest bør vide og gøre!* Lokaliseret 25. maj på NordicMetrics: http://nordicmetrics.com/wp-content/uploads/2014/05/Hvad_enhver_laerer_boer_vide_om_at_gennemfoere_DNT.pdf

- **Olsen, J.V.** (2011). *Høj score i nationale test udløser penge.* Lokaliseret 25. maj 2017 på folkeskolen.dk: https://www.folkeskolen.dk/68530/hoej-score-i-nationale-test-udloeser-penge

Lærerprofession.dk - et site for lærerpraksis og professionsudvikling

- **Olsen, J.V.** (2016) *Professor: De nationale test er spild af tid og penge*. Lokaliseret 25. maj 2017 på Folkeskolen.dk: https://www.folkeskolen.dk/579362/professor-de-nationale-test-er-spild-af-tid-og-penge

- **Politiken** (2013) *Dansklærere gav samme danske stile mellem 02 og 12*. Lokaliseret 24. maj 2017 på Politiken: http://politiken.dk/indland/uddannelse/art5450152/Danskl%C3%A6rere-gav-samme-danske-stile-mellem-02-og-12

- **Prehn, N.L.** (2017), *Skoleelev: Drop dog de nationale test. De spilder min skoletid og skaber unødigt og uproduktivt stress*. Lokaliseret 18. maj 2017 på Politiken: http://politiken.dk/debat/kroniken/art5956771/Drop-dog-de-nationale-test.-De-spilder-min-skoletid-og-skaber-un%C3%B8digt-og-uproduktivt-stress

- **Rambøll** (2013), *Evaluering af de nationale test i folkeskolen*. Lokaliseret 28. maj 2017 på Rambøll: http://www.ramboll.dk/~/media/Files/RM/Rapporter/Nationale%20test%202013.pdf

- **Retsinformation** (2016), *Bekendtgørelse af lov om folkeskolen* af Ministeriet for Børn, Undervisning og Ligestilling. Lokaliseret 28. maj 2017 på Retsinformation: https://www.retsinformation.dk/forms/r0710.aspx?id=182008

- **Richards, J. C.** (2008). *Moving beyond the Plateau*. Cambridge University Press.

- **UVM** (2017a) *Om nationale test*. Lokaliseret 13. maj 2017 på UVM: http://www.uvm.dk/folkeskolen/elevplaner-nationale-test-og-trivselsmaaling/nationale-test/om-nationale-test

- **UVM** (2017b) *Vejledning om de nationale test – til lærere i alle fag*. Undervisningsministeriet

- **UVM.** *Fælles Mål*. Lokaliseret 27. maj 2017 på UVM: http://www.uvm.dk/folkeskolen/fag-timetal-og-overgange/faelles-maal

- **VK Regeringen II** (2005) *Nye mål - Regeringsgrundlag*. Lokaliseret 26. maj 2017 på Statsministeriet: http://www.stm.dk/multimedia/Nye_M_l__Regeringsgrundlag.pdf

- **Vygotsky, L.** (1987). Zone of proximal development. *Mind in society: The development of higher psychological processes*, *5291*, 157**.**

- **Wandall, J.** (2011). National tests in Denmark–CAT as a pedagogic tool. *Journal of Applied Testing Technology*, *12*(1).

- **Wandall, J.** (2013). Education, testing, and validity: a Nordic comparative perspective in Chatterji, M. (2013). *Validity and Test Use: an international dialogue on educational assessment, accountability and equity*. Bingley: Emerald.

- **Wandall, J.** (2014) *Hvornår leverer de nationale test en reliabel måling? Og hvornår er den retvisende for elevens dygtighed*. Lokaliseret 14. maj 2017 på folkeskolen.dk: https://www.folkeskolen.dk/540420/hvornaar-leverer-de-nationale-test-en-reliabel-maaling-og-hvornaar-er-den-retvisende-for-elevens-dygtighed