Jeppe Bundsgaard

Pædagogisk brug af test

Tidsskriftet Sakprosa

Bind 10, Nummer 2

© 2018

10.5617/sakprosa.6007

Sammendrag

De danske nationale test er computerbaserede, selvrettende og adaptive og er obligatoriske for elever i en række fag på 2.-8. klassetrin. Ved beslutningen om at indføre dem i 2006 var det politikernes genkommende begrundelse at de skulle fungere som et pædagogisk redskab og bidrage til at skabe en stærk evalueringskultur. Denne artikel undersøger om det med nationale test er lykkedes at skabe et sådant pædagogisk redskab.

Undersøgelsen indledes med en gennemgang af måder en test som nationale test potentielt kan fungere som et pædagogisk redskab, og nationale test undersøges derefter for om disse måder er realiserede. Undersøgelsen foretages som et dokumentstudie af baggrundsrapporter, vejledninger, evalueringer, hjemmesider mv. fra Undervisningsministeriet. Det viser sig i analyserne at nationale test kun i meget begrænset omfang kan fungere som et pædagogisk redskab. Det overvejes dernæst hvad den rolle nationale test spiller, kan have af konsekvenser for skolen og undervisningen af eleverne.

Abstract

The Danish national tests are computer-based, self-correcting and adaptive and are mandatory for students in a number of subjects in the 2nd to 8th grade. By the decision to introduce them in 2006, the politicians' recurring reason was that they should act as a pedagogical tool and contribute to creating a strong evaluation culture. This article examines whether the national test has succeeded in creating such a pedagogical tool. The study begins with a review of ways a test like national tests may potentially act as a pedagogical tool, and national tests are then examined regarding the question of whether these ways are realized. The survey is conducted as a document study of background reports, guides, evaluations, websites, etc. from the Ministry of Education. The analyzes show that national tests only

function to a very limited extent as a pedagogical tool. It is then considered how national test can have consequences for the school and the teaching of the students.

Nøkkelord: Standardiserede tests, Rasch, information om test, pædagogisk redskab.

Om forfatteren:

Jeppe Bundsgaard er Professor MSO ved Danmarks institut for Pædagogik og Uddannelse, Aarhus Universitet.

Pædagogisk bruk af test

På baggrund af en OECD-rapport der kritiserede det danske skolesystem for manglende evalueringspraksis (Undervisningsministeriet, 2004), blev det i 2006 besluttet at der skulle udvikles nationale test til brug i Folkeskolen i Danmark. I politikernes argumenter for indførelse af en sådan national test var det centralt at den skulle være et pædagogisk redskab for lærerne. I denne artikel vil jeg give et bud på hvordan standardiserede og normerede tests kan fungere som pædagogiske redskaber, og jeg vil undersøge om nationale test faktisk er blevet et sådant redskab.

Intentionerne med de danske nationale test

Indtil 2010 var der i Danmark ikke lovmæssigt krav om test af elevernes kompetencer ud over afgangsprøverne som finder sted i 9. og (den frivillige) 10. klasse. Med nationale test blev det besluttet at indføre jævnlige test inden for en række fag. Afgangsprøverne består både i mundtlige og skriftlige prøver; de mundtlige kan være individuelle og gruppebaserede, og de skriftlige består både i skrivning af længere sammenhængende tekst, løsning af problembaserede matematikopgaver, samt udfyldning og multiple choice (Undervisningsministeriet, 2018b). I modsætning hertil er nationale test alle computerbaserede, adaptive og multiple choice eller udfyldning af tal.

Loven om nationale test blev fremsat i 2005 af Undervisningsminister Bertel Haarder og diskuteret i Folketingets Uddannelsesudvalg som “Forslag til lov om ændring af lov om folkeskolen (Styrket evaluering og anvendelse af nationale test som pædagogisk redskab samt obligatoriske prøver)” (Uddannelsesudvalget, 2005). Som det fremgår af parenteser efter lovforslaget, så var målet at nationale test skulle blive et pædagogisk redskab. I sin fremsættelsestale præsenterede Bertel Haarder testene således:

Det foreslås endvidere, at der som led i evalueringen af elevens udbytte af undervisningen skal anvendes centralt udarbejdede test i udvalgte fag på bestemte klassetrin. Testene er et pædagogisk værktøj (Haarder, 2005).

Fire partier var fortalere for loven, og ordførerne for disse partier havde samme genkommende begrundelse som ministeren for deres støtte. Ordføreren fra det Konservative Folkeparti, Carina Christensen, beskrev lidt mere præcist hvad et pædagogisk redskab vil sige:

[...] med testene følger en række konkrete forslag og værktøjer, der vil kunne hjælpe lærerne med det allervigtigste, nemlig den efterfølgende opfølgning og planlægning af undervisningen, vejledningen af den enkelte elev og underretningen af forældrene (Uddannelsesudvalget, 2005).

Det Radikale Venstres ordfører, Marianne Jelved, var ligesom ordførerne fra Socialistisk Folkeparti og Enhedslisten imod indførelsen af tests, og mente at “[...] så kan den faktisk ikke bruges på den måde [som pædagogisk redskab]” (Uddannelsesudvalget, 2005). Men samlet set var det altså politikernes forventning at nationale test skulle fungere som et bidrag til en øget evalueringskultur og som pædagogisk redskab for lærerne, og at der som led i dette fra Undervisningsministeriets side ville følge en række konkrete forslag og værktøjer som lærerne kunne bruge i deres efterfølgende opfølgning og planlægning af undervisningen med henblik på at vejlede den enkelte elev.

Forskningsspørgsmålet i denne artikel er derfor: *Er nationale test kommet til at fungere som et pædagogisk redskab for lærerne?*

Dette spørgsmål besvares gennem en præsentation af mulighederne for at bruge standardiserede og normerede tests som pædagogiske redskaber, og en analyse af et udvalg af de forslag og værktøjer som lærerne faktisk har til

rådighed. Bemærk at jeg derved ikke forholder mig til politikernes ønske om at skabe en bedre evalueringskultur. Et sådant spørgsmål kræver en undersøgelse af lærernes og skolernes praksis. I en repræsentativ spørgeskemaundersøgelse om nationale test fra 2016 blev lærerne spurgt om nationale test havde bidraget til en bedre evalueringskultur på skolen. Kun 17 procent var lidt eller ret enige i at dette var tilfældet, 0 procent var helt enige (Bundsgaard & Puck, 2016, s. 19).

Forskning i pædagogisk brug af test i Norden

Der har i de seneste årtier fundet sted en del forskning om test og prøver i Norden. Således er nationale læsetest i Sverige, Norge og Danmark blevet sammenlignet (Tengberg, 2017), reliabiliteten af test er undersøgt (Tengberg & Skar, 2017; Vagle, Roe, & Narvhus, 2009), og reliabiliteten ved menneskelig vurdering af test og prøver er undersøgt og problematiseret (DAMVAD Analytics & Kreiner, 2016; Dolin, Nielsen, & Rangvid, 2018; Tengberg, Roe, & Skar, 2018). Denne artikel tager et andet perspektiv og spørger ikke om testen er reliabel, men om den er brugbar.

Man kan overordnet skelne mellem to typer af test: Test som gives ved afslutningen af et skoleforløb og indgår i systemets samlede karakteristik af eleven, ofte kaldet eksamen eller afgangsprøve, og test som gives undervejs i skoleforløbet og som indgår i monitorering af elevernes, lærernes og skolernes status og udvikling eller i lærernes arbejde med at identificere elevernes kompetencer og udviklingspotentialer, altså som et pædagogisk redskab. Det er denne sidste funktion som politikerne i Danmark brugte som begrundelse for indførelsen af nationale test.

I Norden er der foretaget en del forskning som relaterer sig til den første type af test og til den anden types første funktion, men der findes ikke megen nordisk forskning som undersøger potentialer for eller praksis med

pædagogisk brug af test. Tekster om pædagogisk brug af test er primært skrevet af ministerier (Styrelsen for It og Læring, 2017c), konsulenter eller lærere (Hansen, 2016; Weyergang & Ryen, 2014) og testudviklere (Pøhler, 2016; Roe, 2010) som vejledning til hvordan testene kan bruges.

I 2009 gennemførte Dansk Clearinghouse for Uddannelsesforskning et review af international forskning om pædagogisk brug af test. Her erfarer forfatterne blandt andet at “[...] ingen af de undersøgelser, der indgår i dette review, beskæftiger sig med lærerens tolkning af de informationer, som testdata yder” (Nordenbo m.fl., 2009b, s. 61), og senere (s. 66) konkluderes det at “[...] “pædagogisk brug af testdata” er svagt belyst forskningsmæssigt”.

I 2003 vedtog man i Norge at oprette et “Nasjonalt kvalitetsvurderingssystem for grunnsopplæringen” (NKVS) som blandt andet bestod i indførslen af nationale prøver som minder om de danske nationale test (om end de hverken er digitale eller adaptive). Formålet med prøverne kan sammenfattes således:

- Kortlægge i hvilken grad eleverne har nået læreplanens mål
- Give information som grundlag for forbedrings- og udviklingsarbejde
- Give information, som kan bruges i arbejdet med at forbedre kvaliteten i oplæringen

(Skov, 2009, s. 194).

Det er altså tanken at prøverne skal kunne bruges både som et styringsredskab og som et pædagogisk-diagnostisk redskab, og der er enighed blandt kommuner, rektorer og i nogen grad lærere om at disse mål er gode (Skov & Brøyn, 2010, s. 86). Men i en undersøgelse af de nationale prøver, konkluderer Poul Skov at “[d]en pædagogisk diagnostiske brug over for de enkelte elever er lille set i forhold til den styringsmæssige brug” (Skov, 2009,

s. 199). Der er ikke i forbindelse med undersøgelsen en yderligere udfoldelse af hvad det vil sige at test bruges pædagogisk.

Den statistiske teori bag nationale test

Test har været brugt igennem hundreder af år, og en række statistiske metoder er blevet udviklet til at sikre kvaliteten af disse tests. Ofte skelner man mellem en klassisk (Novick, 1966) og en moderne (Rasch, 1960) testteori. Mange test der bruges i skolen i Danmark, udvikles stadig på baggrund af den klassiske testteori. Det gælder fx læsetest (K. L. Jensen, Gellert, & Elbro, 2015; Møller m.fl., 2014) og test til matematik (P. E. Jensen & Jørgensen, 2007). Nationale test er udviklet med moderne testteori i form af Rasch-modellen (Rasch, 1960), også kaldet Item Response Theory (IRT), som både gør evalueringen af testen som helhed og de enkelte opgaver (items) væsentligt mere kvalificeret, og som giver en række yderligere muligheder for at gøre pædagogisk brug af resultaterne.

Rasch-modellen gør det muligt at opnå et mål for det enkelte items sværhedsgrad og for elevens dygtighed på samme skala (Rasch 1960; Bundsgaard, 2018). Modellen er probabilistisk (dvs. baseret på sandsynlighed), hvilket betyder at den hviler på en hypotese om at en elev med en given dygtighed vil have en vis sandsynlighed for at svare rigtigt på en given opgave. Hvis eleven har samme dygtighed som opgavens sværhedsgrad, er sandsynligheden for at eleven svarer rigtigt på opgaven ifølge Rasch-modellen 50 procent. Hvis eleven er dygtigere end opgavens sværhedsgrad, er sandsynligheden for rigtigt svar større, og er eleven mindre dygtig end opgavens sværhedsgrad, er sandsynligheden for rigtigt svar mindre.

Rasch-modellen gør det desuden muligt at bestemme sværhedsgraden af opgaver ved at lade en mindre gruppe (nogle få hundrede) elever besvare opgaverne. Når sværhedsgraderne er fastsat, og kvaliteten af opgaverne er

godkendt statistisk, kan man gå ud fra at andre elever vil opleve at opgaverne har samme sværhedsgrad. Dette betyder at man kan udvikle et stort antal opgaver inden for et givent område, men nøjes med at give en elev et mindre udvalg af disse opgaver, og alligevel efterfølgende kunne udtale sig om hvad elevens dygtighed er i forhold til det disse opgaver måler, og sammenlignet med andre elever.

Disse karakteristika betyder at man kan udvikle såkaldte adaptive test, dvs. test hvor man forsøger at stille opgaver af en sværhedsgrad der passer til elevens dygtighed. Derved skal man samlet set stille færre opgaver for at opnå en større sikkerhed i resultatet, og derved kan eleven opleve udfordringer på sit eget niveau (ikke for let, ikke for svært). Nationale test er adaptive.

Når man måler en elevs dygtighed med en test, er der som sagt forskellige sandsynligheder for at eleven svarer korrekt på de enkelte opgaver. Og det gælder samtidig at eleven får et endeligt antal opgaver, fx 20. Derfor er testens resultat ikke 100 procent præcist, men der er en måleusikkerhed som man kan udregne størrelsen af med Rasch-teorien. Man kan bruge denne måleusikkerhed til at beregne et såkaldt konfidensinterval inden for hvilket elevens sande dygtighed befinder sig fx med 95 procents sandsynlighed. Det er naturligvis ret vigtigt at dette konfidensinterval er så småt som muligt, men det er ikke let at opnå, for der skal ret mange opgaver til at opnå stor sikkerhed, og de skal ligge tæt på elevens dygtighed, for at give så meget information om elevens dygtighed som muligt. Det er det der gør adaptive test smarte, men denne fordel har en pris; det er nemlig nødvendigt at udvikle mange flere opgaver for at der er tilstrækkeligt med opgaver på forskellige sværhedsgradsniveauer.

Pædagogisk brug af test

Målet med nationale test var som sagt at de skulle kunne fungere som et pædagogisk redskab. Men hvad vil det sige? I det tidligere omtalte review af forskning om pædagogisk brug af test, påpegede forfatterne at

[...] man kan forestille sig, at test og testdata kan “bruges” pædagogisk på to måder. I det første tilfælde sker brugen, efter at testen er gennemført og testdata foreligger. I det andet tilfælde sker brugen, før testen er gennemført og testdata endnu ikke foreligger. I sidste tilfælde knytter brugen sig [til] annonceringen af hændelsen “der skal gennemføres test” (Nordenbo m.fl., 2009b, s. 31).

Denne pointe uddybede forfatterne i en figur (se figur 1) hvor mulige relationer mellem testdata og brugen af disse fremgår, hvor relation 4 er brug af viden om at testen skal foregå til at planlægge undervisningen med – det der også kaldes *teach to the test*, og relation 5 er elevernes reaktion på at de ved at de skal til test – som bl.a. kan føre til at de sætter sig grundigere ind i stoffet eller bliver angste for at klare sig dårligt. Politikerne der vedtog at indføre nationale test, gav udtryk for at de skulle “styrke den løbende evaluering”, “hjælpe lærerne med [...] vejledning af den enkelte elev”, samt give et “bedre grundlag for den videre planlægning af undervisningen”. Disse udsagn relaterer til relationerne 1 og 2 som handler om hvilke informationer testen stiller til rådighed for læreren, og hvordan læreren kan bruge disse informationer i sit arbejde med at undervise eleverne (relation 3). Det er derfor disse relationer jeg vil undersøge i nærværende artikel.

Figur 1. Model af sammenhænge mellem testning/testdata og elevlæring (Nordenbo m.fl., 2009b, s. 32)

Det danske undervisningsministerium udarbejdede i 2016 et notat “De nationale test som pædagogisk værktøj” (Styrelsen for Undervisning og Kvalitet, 2016) hvori de opstiller 6 punkter der gør nationale test til et pædagogisk værktøj. Tre af de seks punkter handler om anvendelse af testen som pædagogisk redskab, nemlig at testen skal kunne bruges i planlægningen af undervisningen (formativt), at læreren skal kunne følge elevernes progression og at testen skal relatere sig til undervisningens indhold.

I de følgende seks afsnit vil jeg på baggrund af det internationale review og ministeriets opstilling af krav til et pædagogisk værktøj give et bud på hvad en test der fungerer som pædagogisk værktøj, skal kunne. I det første afsnit argumenterer jeg for at for overhovedet at kunne tale om at en test kan bidrage til løbende evaluering, vejledning og planlægning, skal man vide hvad testen måler (bl.a. så man kan vide hvordan den relaterer sig til undervisningens indhold). I de følgende to afsnit undersøger jeg hvordan tests kan bruges til at få et sammenligningsgrundlag og følge elevernes progression. Derefter fokuserer jeg i to afsnit på hvordan læreren kan få grundlag for at vejlede den enkelte elev, først i forhold til hvor elever har udfordringer, og dernæst i forhold til hvad de kan forventes at skulle til at i gang med at lære.

Det sidste afsnit handler om hvordan læreren kan bruge tests til at planlægge sin undervisning.

Jeg stiller således følgende seks spørgsmål som, hvis de kan besvares med testen, kan gøre den til et pædagogisk redskab:

- Hvad måler testen?
- Hvordan klarer eleverne sig i klassen sig i forhold til kommunens eller landets elever?
- Er eleverne blevet bedre i løbet af den forgangne periode?
- Hvilke elever har problemer i forhold til det faglige område der måles?
- Hvad er den nærmeste zone for udvikling for eleverne inden for det målte område?
- Hvad kan jeg gøre som lærer i forhold til de forskellige grupper?

Efter at have beskrevet disse spørgsmål undersøger jeg i hvilket omfang de danske nationale test er et pædagogisk redskab set i forhold til disse kriterier.

Hvad måler testen?

En forståelse af hvad testen måler, er en forudsætning for at bruge den som redskab. Spørgsmålet kan forekomme unødvendigt, for en læsetest måler jo læsning, og en matematiktest måler matematik. Men en test vil altid måle et uddrag af et fagligt område, og den vil altid gøre det fra et særligt perspektiv.

I det førømtalte review af pædagogisk brug af nationale test henviser forfatterne da også til Tresch (2007) og konkluderer at: “[...] effektiv feedback på test forudsætter en præcis beskrivelse af testen, så læreren ved, hvad der skal måles [...]” (Nordenbo m.fl., 2009b, s. 62).

Mark Wilson har udviklet et rammeværk for udvikling af test, der består af fire faser: Udvikling af et construct map, udvikling af items, bestemmelse af outcome space, dvs. hvad der tæller som rigtige (samst evt. delvist rigtige) og forkerte svar, og endelig fastlæggelse af statistisk målemodel (Wilson & Gochyyev, 2013). Den første fase består i at udarbejde en præcis definition af det fænomen der ønskes målt (the construct), og at beskrive en hypotese for progression inden for området, altså at skabe et kort (et construct map) over bevægelsen fra ikke at være så god til at have fuld kontrol over det målte fænomen. Dette arbejde er afgørende vigtigt for udviklingen af testen som helhed, for det skaber et grundlag for udviklingen af items, og det gør at brugere af testen efterfølgende kan forstå hvad det er meningen at testen måler – og ikke mindst hvad den ikke måler. Store internationale tests som OECD's PISA, IEA's TIMSS, PIRLS og ICILS har alle af denne grund udgivelser der beskriver det såkaldte framework (fx Fraillon, Schulz, & Ainley, 2013; OECD, 2017). I sådanne frameworks indgår en præcis definition af fænomenet, afgrænsninger af hvad der måles, beskrivelser af de typer af items der anvendes osv. I Norden ses også enkelte eksempler på udvikling af construct maps i forbindelse med tests, fx er der i forbindelse med den norske skriveprøve udviklet en beskrivelse af såkaldte "mestringsnivåer på baggrund af eksperterers analyser, rangering og beskrivelse af elevtekster" (Skar, 2017, s. 9f.) og sådanne mestringsniveauer er også beskrevet for de norske nationale prøver i læsning, regning og læsning på engelsk (disse prøver finder ligesom de danske nationale test sted undervejs i skoleforløbet) (Skov, 2009, s. 112ff.).

Viden om hvad der testes, og hvad der tæller som rigtige svar, er naturligvis afgørende vigtigt for den lærer der ønsker at anvende en test til at få viden om sine elevers kompetencer inden for et givent område, og det er afgørende vigtigt for at læreren kan afgøre hvad der skal til for at forbedre sine elevers kompetencer inden for testens område.

Viden om hvad der testes, er også afgørende for at beslutte om det overhovedet giver mening at teste eleverne med denne test, og for at afgøre om der er brug for at supplere med andre tests af andre af de faglige områder man som lærer og skole lægger vægt på.

Hvordan klarer eleverne sig?

Det mest umiddelbare spørgsmål en test kan svare på, er hvordan eleverne klarer sig individuelt og som klasse i forhold til fx kommunens eller landets elever, og hvordan klassens elever fordeler sig i forhold til testens område. Når der er tale om Rasch-tests, kan elevernes resultat formuleres på en skala der minder om den metriske skala, dvs. at en elev der fx får 1 på skalaen, er lige så meget bedre end en der får 0, som den der får 2, er bedre end den der får 1. Hvis testen på et tidspunkt er givet til et repræsentativt udvalg af elever på et eller flere klassetrin, siger man at den er normeret. Det betyder at læreren kan forholde eleverne i sin classes resultater til det "normale", og altså se hvor klassens elever ligger i forhold til hvad man kan forvente. Denne fordeling vil afhænge af hvad elevernes socio-økonomiske forudsætninger er (Skolestyrelsen, 2011), hvordan de er blevet undervist, om de er motiverede for at svare godt på testen og meget andet, og det vil naturligvis være væsentligt for læreren at vide om elever fra lignende skoler klarer sig dårligere eller bedre end hendes egne elever.

Dette er måske den mest indlysende brug af test, men når læreren har fundet ud af hvordan hendes elever ligger, hvad så?

Er eleverne blevet bedre i løbet af den forgangne periode?

En anden indlysende brug af test er til undersøgelse af elevernes faglige progression. For at kunne det, er det nødvendigt at eleverne enten tager nøjagtig den samme test igen, eller at testen er udviklet efter Rasch-

principperne. Ulempen ved at tage den samme test to gange er at eleverne måske kan huske opgaverne, og at de kan have talt om dem med hinanden eller andre efter testen, og at de derfor ikke er bedre ved den anden test, fordi de har lært mere, men fordi de kan huske svarene. Fordelen ved Rasch-modellen er at man kan bestemme items' sværhedsgrad på forhånd og så give to forskellige uddrag af testen til eleverne ved den første og den anden test. På den måde kan man se om eleverne har udviklet sig på det faglige område. Hvis man vil vide om eleverne har udviklet sig mere end forventet, kan man holde klassens udvikling op imod den normale udvikling sådan som den kommer til udtryk hos et repræsentativt udvalg af elever der også har taget testen to gange med samme tidsinterval.

Også her kan man spørge: Og hvad så? Når læreren ved at hendes elever har udviklet sig som forventet (eller bedre eller værre), hvad kan hun så bruge den oplysning til?

Elever med faglige vanskeligheder

En meget væsentlig viden for en lærer er, hvilke elever der har problemer i forhold til et fagligt område. Men det kan være meget svært at overskue 24 elever på samme tid, og det kan være særlig svært at vide hvad en elev der fx ikke siger så meget, kan (Virinkoski, Lerkkanen, Holopainen, Eklund & Aro, 2017).

Ofte vil læreren have særlig opmærksomhed på elever som har faglige udfordringer eller måske endda handicaps, fx ord- eller talblindhed. Mange tests er udviklet til at identificere netop sådanne elever, og de vil ofte have mange opgaver med en sværhedsgrad lige omkring hvad testudviklerne har defineret som grænseværdien for fx ordblindhed. Derved bliver testen særligt præcis i forhold til at evaluere om eleven er over eller under denne grænse-

værdi. Det er særligt nødvendigt hvis der fx knytter sig særlige indsatser til handicappet (fx særlig støtte til ordblinde elever).

Men også tests der ikke er udviklet med henblik på at identificere elever med specifikke vanskeligheder, kan bruges til at identificere elever som måske skal undersøges nærmere gennem opfølgning fx i form af mere specifikke tests eller særlig observation af en fagperson.

Tests kan naturligvis også bruges til elever der ligger uden for et problemområde, fx til at vurdere om eleven egner sig til optagelse på en ungdomsuddannelse eller en videregående uddannelse. Men her er der ikke længere tale om egentlig pædagogisk brug.

Hvad er elevernes nærmeste zone for udvikling?

De hidtil nævnte spørgsmål der kan besvares, har set fra et fagdidaktisk perspektiv kun begrænsede muligheder for pædagogisk brug. De handler om sammenligning og identifikation af elever med særlige behov (eller dygtighed). Men test udviklet efter Rasch-modellen har også en helt særlig mulighed som hidtil ikke er udnyttet i særlig høj grad uden for de internationale såkaldte storskalatests som PISA og IEA-testene (de norske skriveprøver har dog en tilgang der kommer tæt på (Skar, 2018)).

Med tests der er udviklet på baggrund af et veldefineret construct og med items der måler constructet på nuanceret vis, er det muligt at udarbejde en empirisk begrundet beskrivelse af hvad elever med forskellige dygtigheder er i stand til og ikke er i stand til. Det kaldes described proficiency scales (Bundsgaard, Submitted; Fraillon m.fl., 2015; OECD, 2014), hvilket kan oversættes til empiriske kompetenceniveaubeskrivelser. En empirisk kompetenceniveaubeskrivelse udvikles ved at inddele skalaen i et antal intervaller, og bede faglige eksperter undersøge de items som ligger inden for hvert interval, og beskrive hvad der kendetegner dem. På den baggrund kan

man beskrive hvad elever der ligger inden for et givet interval, ved, og hvad de er i stand til. I PISA's læsetest har man fx valgt at beskrive seks niveauer, hvor eleverne fx på niveau 4 og over (dem var der godt 28 procent af globalt i 2009) bl.a. kunne løse opgaver der

[...] require the reader to locate and organise several pieces of embedded information [and] require understanding and applying categories in an unfamiliar context [and] require readers to use formal or public knowledge to hypothesise about or critically evaluate a text (OECD, 2010, s. 47).

Hvis læreren sammen med testresultatet får en beskrivelse af hvad elever der ligger på de forskellige niveauer, er i stand til, og hvad de stadig ikke er i stand til, så får hun en fagligt relevant viden, om hvad eleverne kan.

Denne viden kan hun bruge til at tilpasse sin undervisning til eleverne på de forskellige niveauer, så dem der fx ligger over niveau 3, ikke behøver bruge alt for meget tid på at forstå og træne de kompetencer der fx ligger på niveau 2, og så de heller ikke i alt for stor grad udsættes for udfordringer der kræver kompetencer fx på niveau 5. Der er således tale om en empirisk begrundet beskrivelse af hvad zonen for nærmeste udvikling (Vygotskii, 1978, s. 86) er for elever med forskellige dygtigheder.

Vejledninger og læremidler i tilknytning til test

Den sjette og sidste måde læreren kan bruge en test som pædagogisk redskab på, er ved at tage udgangspunkt i vejledninger eller læremidler som relaterer til den test der er taget. Hvis der er udviklet empiriske kompetenceniveau-beskrivelser, så er det også muligt at udvikle læremidler som relaterer til disse beskrivelser, og de kan give læreren forslag til hvordan hun kan tilrettelægge undervisningen så der er fagligt relevante og overkommelige udfordringer for elever på de niveauer der er repræsenteret i klassen.

Men desværre er mange læremidler der knytter sig til test, ikke udviklet i forhold til en kompetenceniveaubeskrivelse, så de kan bidrage til at kvalificere undervisningen efter testen. De er i stedet udviklet som materiale læreren kan give eleverne som optakt og forberedelse til testen. Det at score højt på testen bliver derfor et mål i sig selv. Sådanne materialer vil fx bestå af mange opgaver der minder om (eller er direkte taget fra) testen. Dette er et særligt stort problem hvis testen har karakter af high stakes, dvs. hvis eleverne eller læreren kan imødesee sanktioner (fx blive fyret, skulle gå en klasse om) eller opnå privilegier (fx få lønforhøjelse, komme ind på en særlig fin uddannelsesretning) på baggrund af testresultatet (Nordenbo m.fl., 2009a).

Pædagogisk brug af nationale test

Metode

Den følgende analyse af nationale test som pædagogisk redskab forsøger at besvare hvert af de seks spørgsmål jeg har stillet ovenfor. Besvarelsene bygger på studier af Undervisningsministeriets dokumenter og resurser om nationale test, samt på analyse af de testitems der er tilgængelige offentligt. Dokumenterne er fundet ved besøg på Undervisningsministeriets og Styrelsen for It og Lærings hjemmesider om nationale test (Styrelsen for It og Læring, 2017a, 2018, u.å.-b; Undervisningsministeriet, 2018). Dokumenterne består i vejledninger til lærere, skoleledere og kommuner, rapporter om evalueringer af og metoder bag nationale test, samt webinarer hvor ministeriets embedsmænd forklarer hvordan testen kan bruges, og svarer på spørgsmål. Da testitems er hemmelige, er den eneste mulighed for at undersøge de konkrete items i nationale test, de opgaver der ligger til afprøvning på hjemmesiden demo.testogprøver.dk. Undervisningsministeriet har i en mail til mig angivet at "Demotest er opdateret og omfatter alle opgavetyper, der aktuelt anvendes i nationale test" (personlig email-korrespondance). Lærere har adgang til en backend hvor de kan se deres elevers resultater og besvarelser. Denne

backend har jeg ikke adgang til, men funktionaliteterne er forklaret med skærmbilleder i de vejledninger der stilles til rådighed for lærerne. Endelig har jeg bedt Undervisningsministeriet om adgang til de materialer der lå til grund for udviklingen af nationale test, de evalueringer der er foretaget, og de materialer der stilles til rådighed for opgaveudviklere. Det udleverede materiale indgår også som grundlag for det følgende.

De danske nationale Test

Efter en række tekniske og faglige problemer med den første version af testen (Devoteam Consulting, 2007), blev nationale test indført som obligatorisk test i 2010. Nationale test indeholder i dag fire test til dansk, læsning (2., 4., 6. og 8. klasse), tre til matematik (3., 6. og 8. klasse), to til engelsk (4. og 7. klasse), samt én til fysik/kemi (8. klasse). Disse test er obligatoriske for alle elever, og de skal tages i foråret på de nævnte klassetrin. Testene kan desuden tages frivilligt en gang før den obligatoriske og en gang efter. Der er desuden frivillige test til biologi, geografi samt dansk som andetsprog.

Hver test forventes at tage 45 minutter at gennemføre. Læreren kan i sit kontrolpanel (kaldet "Afviklingssiden" (Styrelsen for It og Læring, 2017b, s. 41)) følge med i hvor langt den enkelte elev er i testen, og læreren kan se når eleven har nået et tilstrækkeligt sikkert resultat til at testen kan afsluttes (konkret ser læreren et grønt statusmærke ud for elevens navn). Hvis ikke resultatet er sikkert nok, skal (hvis status er rød) eller kan (hvis status er gul) læreren tildele eleven yderligere 15 minutter til testafviklingen.

Eleverne løser testen individuelt ved en computer eller tablet. Testen er selvrettende og adaptiv, hvilket vil sige at computeren udvælger opgaver der har en sværhedsgrad der svarer til elevens estimerede dygtighed. Konkret foregår det ved at eleven først stilles en opgave af cirka middel sværhedsgrad (Styrelsen for It og Læring, 2015). Hvis eleven svarer korrekt, stilles en opgave

der er mellem 0,5 og 1 logit sværere, og hvis hun svarer forkert, stilles hun en opgave der er tilsvarende lettere. Dette gentages for den tredje opgave. Efter de tre opgaver er besvaret (inden for hver af de tre profilområder der testes), estimeres elevens dygtighed, og hun stilles en opgave omkring dette niveau. Dette fortsætter indtil usikkerheden på estimatet af elevens dygtighed (Standard Error of Measurement) er under 0,5 logit.

Da testen er selvrettende, skal alle opgaver have et entydigt korrekt svar således at en computer kan afgøre om det er korrekt eller forkert. Det sætter i sagens natur nogle grænser for hvilke typer items testen kan indeholde, når der ikke kan indgå menneskelig vurdering af elevens svar i processen. Da testen er adaptiv, er der desuden behov for mange opgaver inden for alle sværhedsgradsniveauer, således at computeren kan vælge items ud med en passende sværhedsgrad. Konkret er der flere tusind opgaver i hver af de omkring 15 testbanker. I figur 2-5 ses fire eksempler på opgaver fra dansk, læsning og matematik. Opgaverne stammer fra de offentlige demo-tests. Opgaven i figur 2 tester afkodning gennem en såkaldt ordkædeopgaver hvor eleven skal sætte to streger i en række af tre sammensatte ord. Stregerne skal markere hvor et ord slutter og et nyt begynder. Opgaven i figur 3 tester tekstforståelse med en cloze-test, hvor eleven blandt fire forslag skal vælge det mest passende ord i sammenhængen i en tekst. Opgaven i figur 4 tester geometri ved at eleven skal angive hvad den ubekendte vinkel C er. Opgaven i figur 5 tester statistik og sandsynlighed ved at eleven skal udregne et antal ud fra en procent-andel.

Figur 2. Eksempelopgave fra dansk, læsning, tester afkodning

Figur 3. Eksempelopgave fra dansk, læsning, tester tekstforståelse

Figur 4. Eksempelopgave fra matematikk, tester geometri

Figur 5. Eksempelopgave matematik, tester sandsynlighed

Der har siden indførelsen af nationale test været en del diskussion i Danmark om nationale test og deres funktionsmåde, og blandt andet på baggrund af sådanne diskussioner har ministeriet udarbejdet eller fået udarbejdet en række rapporter om testenens reliabilitet og validitet (Styrelsen for It og Læring, 2016a, 2016b, u.å.-a). Blandt andet viste en undersøgelse af test-gentest-korrelationerne at ud af 30 forskellige dimensioner som nationale test

måler på, er der blot 12 som har en test-gentest-korrelation som er acceptabel (over 0,7) (Styrelsen for It og Læring, 2016b, s. 10), mens en anden rapport viste at der kan identificeres en sammenhæng mellem elevers score i nationale test i 6. og 8. klasse og deres score i PISA-undersøgelsens måling af elevernes kompetencer i matematik og læsning et til tre år senere (Damvad, 2014).

Hvad måler nationale test?

I hver test får eleverne opgaver inden for tre såkaldte profilområder. Testen måler således tre forskellige ting på 45 minutter. I dansk, læsning er de tre profilområder fx benævnt Afkodning, Sprogforståelse og Tekstforståelse. Der findes ikke definitioner af disse profilområder, men det angives at “Testopgaverne udarbejdes med udgangspunkt i færdigheds- og vidensmål fra Fælles Mål i faget. Hver testopgave er knyttet til målpar i det pågældende trinforløb” (Styrelsen for It og Læring, 2017c, s. 18). Fælles Mål er de officielle beskrivelser af hvad eleverne skal lære i fagene (curriculum). Fælles Mål består af tre til fire kompetenceområder i hvert fag. I dansk er der fire kompetenceområder, og det er kun kompetenceområdet læsning testen relaterer til.

Klassetrin	Kompetencemål	Faser	Færdigheds- og vidensmål							
			Finde tekst	Førberedelse	Afkodning	Sprogforståelse	Tekstforståelse	Sammenhæng		
Efter 2. klassetrin	Eleven kan læse enkle tekster sikkert og bruge den i hverdagsammenhænge	1.	Eleven kan vælge en tekst ud fra et mindre udvalg	Eleven har viden om teksters størrelse og formål	Eleven kan læse ord i tekster til klassetrinnet sikkert	Eleven har viden om bogstavernes kontekstbetydning	Eleven kan gen- give hovedind- holdet af tekster til klassetrinnet	Eleven har viden om fortællende og informerende teksters struktur	Eleven kan fortælle tekstens emne med egen viden, erfaring og ideer	Eleven har viden om samspil mellem tekst og læser
		2.	Eleven kan finde tekster ved at navigere på aldersstillede hjemmesider	Eleven kan anvende enkle fællesstrategier	Eleven har viden om stavemåde og betydning af ord i tekster til klassetrinnet	Eleven har forståelse af betydningen af ord i tekster til klassetrinnet	Eleven kan forstå betydningen af indholdet i ord i tekster til klassetrinnet	Eleven kan kom- binere tekst og baggrundviden til at skabe sam- menhængende forståelse	Eleven kan forholde sig til tekstens emne	Eleven har viden om valde reflek- sions-spørgsmål
Efter 4. klassetrin	Eleven kan læse mul- timediale tekster med beretning på tværs af faglig viden	1.	Eleven kan navigere ud fra søgespørgsmål på aldersstillede hjemmesider og på biblioteket	Eleven har viden om boggrundviden	Eleven kan læse ord i tekster til klassetrinnet hurtigt og sikkert	Eleven kan an- vende enkle og oplysninger af ord	Eleven har viden om funktion og oplysninger af oplysninger af ord	Eleven kan iden- tificere elemen- ter i teksten, som skaber sammen- hæng	Eleven har viden om tekstens påvirkende funk- tion	Eleven har viden om tekstens påvirkende funk- tion
		2.	Eleven kan vurdere teksten ud fra relevans i forhold til søge- spørgsmål	Eleven kan anvende enkle læsestrategier	Eleven har viden om ordklasser og regler for betydning af ord	Eleven kan anvende over- og underbegreber til at skabe sam- menhængende forståelse af teksten	Eleven har viden om over- og underbegreber	Eleven kan iden- tificere proble- mer med at forstå teksten	Eleven kan sætte teksten i sammen- hæng	Eleven har viden om tekstens påvirkende funk- tion
Efter 6. klassetrin	Eleven kan læse og forholde sig til tekster i faglige og offentlige sammen- hænge	1.	Eleven kan vur- dere relevans af søgespørgsmål på aldersstillede hjemmesider	Eleven har viden om søgestrategier	Eleven kan læse stikord og ord i tekster til klassetrinnet hurtigt og sikkert	Eleven har viden om ordklasser og regler for betydning af ord i tekster til klassetrinnet	Eleven kan anvende over- og underbegreber til at skabe sam- menhængende forståelse af teksten	Eleven kan iden- tificere elemen- ter i teksten, som skaber sammen- hæng	Eleven har viden om tekstens påvirkende funk- tion	Eleven har viden om tekstens påvirkende funk- tion
		2.	Eleven kan gen- nemføre billed- og faldteknisk opbygning	Eleven kan anvende enkle læsestrategier	Eleven har viden om ordklasser og regler for betydning af ord i tekster til klassetrinnet	Eleven kan anvende over- og underbegreber til at skabe sam- menhængende forståelse af teksten	Eleven har viden om over- og underbegreber	Eleven kan iden- tificere proble- mer med at forstå teksten	Eleven kan sætte teksten i sammen- hæng	Eleven har viden om tekstens påvirkende funk- tion
Efter 8. klassetrin	Eleven kan styre og regulere sin læse- proces og diskutere teksters betydning i deres kontekst	1.	Eleven kan vur- dere relevans af søgespørgsmål på aldersstillede hjemmesider	Eleven har viden om søgestrategier	Eleven kan læse stikord og ord i tekster til klassetrinnet hurtigt og sikkert	Eleven har viden om ordklasser og regler for betydning af ord i tekster til klassetrinnet	Eleven kan anvende over- og underbegreber til at skabe sam- menhængende forståelse af teksten	Eleven kan iden- tificere elemen- ter i teksten, som skaber sammen- hæng	Eleven har viden om tekstens påvirkende funk- tion	Eleven har viden om tekstens påvirkende funk- tion
		2.	Eleven kan gen- nemføre billed- og faldteknisk opbygning	Eleven kan anvende enkle læsestrategier	Eleven har viden om ordklasser og regler for betydning af ord i tekster til klassetrinnet	Eleven kan anvende over- og underbegreber til at skabe sam- menhængende forståelse af teksten	Eleven har viden om over- og underbegreber	Eleven kan iden- tificere proble- mer med at forstå teksten	Eleven kan sætte teksten i sammen- hæng	Eleven har viden om tekstens påvirkende funk- tion
Efter 9. klassetrin	Eleven kan styre og regulere sin læse- proces og diskutere teksters betydning i deres kontekst	1.	Eleven kan vur- dere relevans af søgespørgsmål på aldersstillede hjemmesider	Eleven har viden om søgestrategier	Eleven kan læse stikord og ord i tekster til klassetrinnet hurtigt og sikkert	Eleven har viden om ordklasser og regler for betydning af ord i tekster til klassetrinnet	Eleven kan anvende over- og underbegreber til at skabe sam- menhængende forståelse af teksten	Eleven kan iden- tificere elemen- ter i teksten, som skaber sammen- hæng	Eleven har viden om tekstens påvirkende funk- tion	Eleven har viden om tekstens påvirkende funk- tion
		2.	Eleven kan gen- nemføre billed- og faldteknisk opbygning	Eleven kan anvende enkle læsestrategier	Eleven har viden om ordklasser og regler for betydning af ord i tekster til klassetrinnet	Eleven kan anvende over- og underbegreber til at skabe sam- menhængende forståelse af teksten	Eleven har viden om over- og underbegreber	Eleven kan iden- tificere proble- mer med at forstå teksten	Eleven kan sætte teksten i sammen- hæng	Eleven har viden om tekstens påvirkende funk- tion
Efter 10. klassetrin	Eleven kan styre og regulere sin læse- proces og diskutere teksters betydning i deres kontekst	1.	Eleven kan vur- dere relevans af søgespørgsmål på aldersstillede hjemmesider	Eleven har viden om søgestrategier	Eleven kan læse stikord og ord i tekster til klassetrinnet hurtigt og sikkert	Eleven har viden om ordklasser og regler for betydning af ord i tekster til klassetrinnet	Eleven kan anvende over- og underbegreber til at skabe sam- menhængende forståelse af teksten	Eleven kan iden- tificere elemen- ter i teksten, som skaber sammen- hæng	Eleven har viden om tekstens påvirkende funk- tion	Eleven har viden om tekstens påvirkende funk- tion
		2.	Eleven kan gen- nemføre billed- og faldteknisk opbygning	Eleven kan anvende enkle læsestrategier	Eleven har viden om ordklasser og regler for betydning af ord i tekster til klassetrinnet	Eleven kan anvende over- og underbegreber til at skabe sam- menhængende forståelse af teksten	Eleven har viden om over- og underbegreber	Eleven kan iden- tificere proble- mer med at forstå teksten	Eleven kan sætte teksten i sammen- hæng	Eleven har viden om tekstens påvirkende funk- tion

Figur 6. Uddrag af Fælles Mål, dansk, kompetenceområdet læsning, færdigheds- og vidensmålpar for afkodning, sprogforståelse og tekstforståelse (horisontalt), for alle 9 klassetrin (vertikalt)

Hvert kompetenceområde består af 9 trin (et for hver årgang, men ikke knyttet direkte til et klassetrin). For hvert trin er der et antal såkaldte færdigheds- og vidensmålpar. Disse er benævnt med en fælles overskrift på tværs af de ni trin. I læsning er der seks sådanne overskrifter ("søjler"), og tre af dem er sammenfaldende med de tre profilområder for nationale test: Afkodning, Sprogforståelse og Tekstforståelse (se figur 6). Man kunne derfor forvente at disse Fælles Mål-søjler udgjorde en definition af hvad der måles i nationale test, men det er ikke tilfældet, for så vidt jeg kan vurdere, tester opgaverne kun ret begrænsede dele af disse i forvejen begrænsede dele af Fælles Måls læseområde (jf. Pøhler, 2016, s. 56). Det gælder fx afkodning som alene testes med ordkædeopgaver, og derfor ikke kan teste fx om eleven kan afpasse sin læsehastighed efter læseformål og ordkendskab, eller om hun kan læse hurtigt og sikkert, og det gælder fx for tekstforståelse som ikke tester strategier ved manglende forståelse, brug af modeller, om eleven kan sammenholde forskellige tekster osv. Det samme gør sig i større eller mindre grad gældende for de øvrige fag der testes i nationale test.

Det er ikke lykkedes mig at finde nogen yderligere angivelse af hvad der måles og hvordan der vælges konkret indhold i testen, heller ikke i det materiale jeg har fået fra det oprindelige testudviklingsarbejde. Det første kriterium for anvendelse af nationale test som pædagogisk redskab, er således kun opfyldt i meget begrænset omfang.

Hvordan klarer eleverne i klassen sig?

Dagen efter eleverne har gennemført testen, får læreren adgang til resultaterne. Der er overordnet set tre tilbagemeldinger. De to af disse er normbaserede, dvs. at de angiver hvordan eleven har klaret sig i forhold til

populationen (i praksis er det en historisk population, typisk 2010 eller 2014). Den ene normbaserede score angives som percentilscore i form af en talværdi der ligger mellem 1-100. Det vil sige at en elev der fx scorer 33 vil være på niveau med eller bedre end 33 procent af den historiske årgang og dårligere end 67 procent af denne årgang. Den normbaserede score er også delt op i seks kategorier i forhold til relationen til gennemsnittet (kategorierne kaldes Klart over middel, Over middel, Middel osv.).

De normbaserede scores giver således læreren indsigt i hvordan eleverne klarer sig i forhold til andre skolars elever. Men læreren får ikke information om hvordan elever fra lignende skoler og socioøkonomiske baggrunde klarer sig, og derfor skal hun selv tage højde for hvilken baggrund klassens elever har, når hun skal vurdere om fordelingen af scores er acceptabel.

Den tredje tilbagemelding kaldes kriteriebaseret og er tilgængelig i forbindelse med matematik- og læsetestene. Den minder om de empiriske kompetenceniveaubeskrivelser ovenfor, idet den hviler på vurdering af konkrete opgaver. En gruppe eksperter har udvalgt 30 opgaver som repræsentative for profilområdet og har vurderet hvor mange der skal svares på for at eleven er på forskellige faglige niveauer. Dette formuleres med ord som Fremragende, Jævn og Mangelfuld præstation. Der er således ikke nogen angivelse af hvad eleverne kan gøre på de forskellige faglige niveauer.

Når læreren vil vurdere den enkelte elevs resultat, skal hun forholde sig til usikkerheden på resultatet. Der er som forklaret ovenfor altid en usikkerhed på et testresultat, og dette mindskes jo flere opgaver eleven har besvaret, og i særlig grad hvis disse opgaver er tæt på elevens dygtighedsniveau.

Men i nationale test er antallet af opgaver ikke stort nok til at nå langt ned i usikkerhed. Dette illustreres i figur 7 som er et uddrag fra *Vejledning til nye resultatvisninger i de nationale test*. Den røde horisontale streg angiver 68

procents konfidensintervallet, dvs. at elevens resultat ligger inden for det interval med 68 procents sandsynlighed. Procentsatserne der er angivet i nogle af kategorierne i den kriteriebaserede og den normbaserede score, angiver hvor stor sandsynligheden (beregnet med udgangspunkt i 95 procents konfidensintervallet) er for at elevens dygtighed ligger i en af disse kategorier. Der er således ret store sandsynligheder for at den pågældende elevs dygtighed reelt er god, jævn eller mangelfuld.

Figur 7. Illustration af usikkerhed i estimat af en elevs dygtighed

Dette eksempel er som sagt valgt fra ministeriets egen vejledning, og det illustrerer hvor stor usikkerheden er på estimatet af den overvejende del af eleverne. Man kan med god ret spørge hvordan et sådant resultat kan bruges pædagogisk. Hvis resultatet ligger omkring hvad læreren ventede, så får læreren bekræftet sin antagelse. Men hvis resultatet ligger langt fra hvad læreren ventede, er det så blot fordi usikkerheden er så stor at elevens reelle dygtighed ligger i yderpunktet af konfidensintervallet? Eller er der faktisk behov for at læreren tager initiativer i forhold til denne elev?

Progression i nationale test

I dansk, læsning, samt matematik og engelsk skal eleverne tage nationale test flere gange over deres skolegang, men hver test inden for det pågældende fag har sin egen bank af items, og der er ikke gennemført en statistisk analyse af om de måler det samme. Så man kan ikke på statistisk validt grundlag sammenligne på tværs af disse tests på forskellige årgange. Men eleverne kan tage den samme test flere gange på frivillig basis (det er ikke nødvendigvis frivilligt for eleverne, og mange kommuner kræver af lærerne at deres elever tager de "frivillige" tests). Da testene er udviklet efter Rasch-modellen, behøver eleverne ikke at svare på de samme opgaver for at få et sammenligneligt resultat. Dette kendetegn anvendes til at give læreren information om elevernes progression. Progressionen vises i en graf (se figur 8) med det nye testresultat på x-aksen og det tidligere resultat på y-aksen. Som det fremgår gengives usikkerheden på resultaterne ikke. Der har i mange sammenhænge (Ravn, 2015a, 2015b) været peget på at et større antal elever oplever at få dårligere resultater til den obligatoriske test, end de fik uger eller måneder før til den frivillige test (hvilket er en følge af den statistiske usikkerhed på resultaterne). I eksempelvisningen er det ikke tilfældet for nogen af eleverne. Men læreren må under alle omstændigheder regne med at både det første og det andet resultat har så stor usikkerhed forbundet med sig, at progressionen ikke er statistisk signifikant (med mindre en elev har bevæget sig mere end to, måske endda tre kategorier).

Nationale test giver med andre ord mulighed for at læreren kan få indblik i elevernes progression, men set fra et statistisk synspunkt er disse oplysninger kun yderst sjældent statistisk signifikante.

Klassens progression fra tidligere test

Kun 19 ud af 20 elever har taget begge test

Resultatet fra følgende elev indgår ikke i graferne: Solvej Laursen

Eleverne i det blå område er gået tilbage siden tidligere test.

Eleverne i det grønne område er gået frem siden tidligere test.

Eleverne i det hvide område fik samme resultat i de to test.

Sammenlign med: Dansk/læsning 6. klasse - Frivillig test - 6. klasse ▾

Samlet resultat

Tidligere resultat Dansk/læsning 6. klasse - Frivillig test - 6. klasse	Fremragende						
	Rigtig god					Jonathan B.	
	God				Thure M. Sami R. Hassan D. Sofia K.	Rose M. Sasha B. Valentin H. Stefan S.	
	Jævn			Josephine T. Bjarke K. Louise K.	Muhammed S. Emil L.		
	Mangelfuld		Laura P. Cecilie M. Malte L.	Jesper F. Line H.			
	Ikke tilstrækkelig						
		Ikke tilstrækkelig	Mangelfuld	Jævn	God	Rigtig god	Fremragende
Nyeste resultat Dansk/læsning 6. klasse - Obligatorisk test - 2018/2017							

Figur 8. Præsentation af progression i nationale test.

Elever med faglige vanskeligheder

Usikkerheden på resultaterne har også betydning for muligheden for at identificere elever med vanskeligheder af faglig art. Men elever der er estimeret til at ligge i den nederste kategori vil efter al sandsynlighed have store vanskeligheder med at følge med i undervisningen. Det understreges dog at yderligere information skal indhentes fra egentlige diagnostiske tests (Styrelsen for It og Læring, 2017c, s. 16).

Den nærmeste zone for udvikling

De kriteriebaserede test er som sagt udviklet med samme grundlæggende ønske som de empiriske kompetenceniveaubeskrivelser i de internationale undersøgelser. Men i modsætning til disse er der ikke nogen faglig beskrivelse af hvad det vil sige at ligge på de forskellige niveauer på den kriteriebaserede skala. Af den grund er der ingen hjælp at hente i nationale test i forhold til at

forstå hvad eleverne kan, hvad de ikke kan, og hvad deres nærmeste zone for udvikling er.

Vejledninger og læremidler i tilknytning til nationale test

Da der som netop omtalt, ikke gives information om hvilke typer af faglige udfordringer eleverne på forskellige kritierieniveauer har, er der heller ikke udviklet egentlige vejledninger eller læremidler der kan hjælpe lærerne med at arbejde fagligt på baggrund af nationale test. Men der er udviklet mange læremidler som lægger op til at eleverne træner opgaverne i nationale test (Clio Online, 2018; Gyldendal, 2018; Pøhler, 2011). Disse læremidler lægger sig typisk tæt op ad de færdigheder der testes i nationale test, og for nogens vedkommende imiterer de sågar itemformatet (fx ordkædeopgaverne).

Nationale test er ikke velegnet som pædagogisk redskab

Som gennemgangen viser, så er det kun i et meget snævert perspektiv at nationale test kan betragtes som et pædagogisk redskab. Det er ikke klart hvad nationale test måler fagligt set, testen måler med så stor usikkerhed at elevens dygtighed kan befinde sig i tre eller flere af de seks niveauer, testen er af samme grund ikke i stand til at vise progression, læreren får ikke viden om hvad elever på de forskellige niveauer er i stand til og har af udfordringer, og læreren får ikke hjælp fra vejledninger eller læremidler til at planlægge undervisning der tager udgangspunkt i den viden der produceres i nationale test.

Man kunne nu spørge: Hvordan forholder Undervisningsministeriet sig til dette? Det gør de for det første ved at understrege at nationale test blot er ét instrument blandt flere i en "stærk evalueringskultur", således går dette citat igen i en række vejledninger og omtaler af nationale test:

En stærk evalueringskultur handler blandt andet om, at lærerne systematisk arbejder med at vurdere og reflektere over effekten af undervisningen med henblik på at øge kvaliteten. Eleverne bliver ikke fagligt dygtigere alene af at blive testet. Men de kan blive dygtigere, når der er en stærk evaluerings-kultur, og når eleverne bliver hjulpet med opfølgning, feedback og forældre-inddragelse. De nationale test er blot ét evalueringsværktøj, som kan indgå i dette arbejde (Styrelsen for It og Læring, 2017a, 2017d, s. 6).

For det andet negligeres problemerne, og lærerne får at vide at resultaterne kan bruges fx til at:

Få viden om klassens og elevernes faglige niveau inden for de områder af fagene, der testes i. Den viden kan du i samspil med dine øvrige evalueringer bruge til for eksempel at skabe overblik.

Få øje på elever, som klarer testen markant dårligere eller markant bedre end det, de øvrige evalueringer viser. Det kan give dig anledning til at overveje, om der er behov for en særlig opmærksomhed på de elever. [...]

“Få overblik over klassens og elevernes faglige progression inden for profilområderne fra den ene test til den næste. [...]” (Styrelsen for It og Læring, 2017c, s. 20).

Problemet med usikkerhed i resultaterne blev også negligeret i de første år af testens levetid, men måske som følge af at denne usikkerhed blev fremhævet som et problem af både lærere, forskere og skoleledere særligt i fagbladet Folkeskolen.dk (Bundsgaard, 2017; Norling, 2016; Ravn, 2015a, 2015b), valgte ministeriet i 2017 at tilføje en visning der angav af usikkerheder på resultaterne (se figur 7).

En anden måde at forholde sig til problemet på er ved i vejledningerne og i webinarer at lægge stor vægt på det der kaldes “elevernes testadfærd” (Styrelsen for It og Læring, 2017c, s. 27; Undervisningsministeriet, 2018). Det forklares således hvordan lærerne, når de oplever at en elev klarer sig anderledes end forventet, kan gå ind og undersøge elevens testforløb. Har eleven pludselig brugt væsentligt mindre tid på opgaverne, svarer hun pludselig forkert på lette opgaver osv. Set fra et statistisk perspektiv er elevernes testadfærd dog ikke en primær medvirkende årsag til testens usikkerhed, men til hvor godt Rasch-modellen beskriver elevens resultat.

Jeg citerede i indledningen ordførerne fra de partier der valgte at indføre nationale test i 2006. Deres argumenter var at lærerne med nationale test ville få et pædagogisk redskab der kunne bidrage til en bedre evalueringskultur, og at lærerne ville få konkrete forslag og vejledninger som de kunne bruge i deres efterfølgende opfølgning og planlægning af undervisningen. Nærværende undersøgelses konklusion er, at et sådant redskab har lærerne ikke fået med nationale test.

Ved 2. behandlingen af forslaget om at indføre nationale test (Uddannelsesudvalget, 2006) påpegede det Radikale Venstres Margrethe Vestager at nationale test i et andet lovforslag samtidig blev tiltænkt en central rolle i kommunernes kontrol af skolernes kvalitet, og at der var en uoverensstemmelse mellem denne brug og formålet om at det skulle være et pædagogisk redskab. I dag fungerer den samlede population af elevers resultater som resultatmål for den skolereform der blev gennemført i Danmark i 2014. Disse mål lyder blandt andet:

Mål 1: Andelen af de allerdygtigste elever i dansk og matematik skal stige år for år. Folkeskolen skal udfordre alle elever, så de bliver så dygtige, de kan. Mindst 80 procent af eleverne skal være

gode til at læse og regne i de nationale test.
(Undervisningsministeriet, 2017).

Disse nationale mål gengives i flere af vejledningerne der knytter sig til nationale test, og testene får derfor en anden rolle end den var tildelt i ordførernes begrundelser som de er gengivet i indledningen til denne artikel. For når staten holder kommunerne ansvarlige for om de bidrager til at eleverne bliver dygtigere målt gennem nationale test, og når kommunerne holder skolelederne ansvarlige, og når de dernæst holder lærerne ansvarlige, så er kimen lagt til at testene bliver *high stakes*, dvs. at de kommer til at spille en rolle for om skoleledere og lærere beholder deres jobs eller bliver forfremmet, og selv om de ikke er afgørende for om eleverne rykkes en klasse op eller kan komme ind på en ungdomsuddannelse, så er det lærernes opfattelse at mange elever oplever det at tage testene som en belastning (Bundsgaard & Puck, 2016). Test der er *high stakes*, er der stærkt belæg for at advare imod (Nordenbo m.fl., 2009a), for de fører til *teaching to the test* og til strategiske handlinger fra elever, lærere og skolelederes side. Og derved bliver nationale test ikke et pædagogisk redskab, men et styringsredskab der virker kontraproduktivt.

Litteratur

Bundsgaard, J. (Submitted). 21st Century Skills in Project-Based Learning. *Journal of Educational Assessment*.

Bundsgaard, J. (2017). Så upræcise er nationale test - Folkeskolen.dk. Hentet 12. februar 2018, fra <https://www.folkeskolen.dk/583103/saa-upraecise-er-nationale-test>

Bundsgaard, J. (2018). Test og måling i fagene. I T. S. Christensen, N. Elf, P. Hobel, & A. Qvortrup (Red.), *Udvikling i didaktik – didaktik i udvikling*. Aarhus: Klim.

Bundsgaard, J., & Puck, M. R. (2016). *Nationale test - danske lærere og skolelederes brug, holdninger og viden*. DPU, Aarhus Universitet. Hentet fra http://pure.au.dk/portal/files/102516329/Ebog_Nationale_test_FINAL_september_2016.pdf

Clio Online. (2018). Clio Træningsøvelser. Hentet 12. februar 2018, fra <https://mit.clioonline.dk/clioonline.myclio/archive/training/>

Damvad. (2014). *PISA-relateret af de kriterie-baserede nationale test. Delrapport 2 – teknisk rapport og dokumentation* (s. 89). København: Damvad.

DAMVAD Analytics, & Kreiner, S. (2016). *Forcensur ved folkeskolens 9. klasses afgangsprøver 2015* (s. 41). København: Styrelsen for Undervisning og Kvalitet, Ministeriet for Børn, Undervisning og Ligestilling. Hentet fra <https://www.damvad.com/uploads/Publications/Endelige-rapport-Forcensur-2015-08.04.2016.pdf>

Devoteam Consulting. (2007). *De nationale it-baserede test i folkeskolen. Rapport fra REVIEW-panelet* (s. 73). U.s.

Dolin, J., Nielsen, K., & Rangvid, B. S. (2018). *Følgegruppen for én bedømmer ved folkeskolens prøver* (s. 28). København: Undervisningsministeriet. Hentet fra

http://pure.au.dk/portal/files/127186747/180425_Rapport_fra_Foelgegruppen_for_een_bedoemmer_ved_FP.pdf

Fraillon, J., Schulz, W., & Ainley, J. (2013). *International Computer and Information Literacy Study: Assessment Framework*. Amsterdam, the Netherlands: International Association for the Evaluation of Educational Achievement (IEA). Hentet fra http://ifs-dortmund.de/assets/files/icils2013/ICILS_2013_Framework.pdf

Fraillon, J., Schulz, W., Friedman, T., Ainley, J., Gebhardt, E., Ainley, J., ... International Association for the Evaluation of Educational Achievement (IEA). (2015). *ICILS 2013: technical report*.

Gyldendal. (2018). dansko-2.gyldendal.dk | Afkodning. Hentet 12. februar 2018, fra http://dansko-2.gyldendal.dk/en/Indgange/Laerer/Om_traeningsdelen/Fagomraader/afkodning.aspx

Haarder, B. (2005). Fremsættelsestale. Hentet 8. februar 2018, fra http://www.ft.dk/samling/20051/lovforslag/1101/20051_1101_fremsaettelsestale.htm

Jensen, K. L., Gellert, A. S., & Elbro, C. (2015). *Rapport om udvikling og afprøvning af Selvtest af læsning – en selvtest af voksnes læsefærdigheder på nettet*. Center for Læseforskning, Københavns Universitet.

Jensen, P. E., & Jørgensen, I.-L. (2007). *Mat-prøver : pædagogisk analyse af matematik*. Virum: Dansk psykologisk Forlag.

Karner Hansen, K. (2016). Vejen fra test til undervisning. *Læsepædagogen*, 64(4), 4–9.

Lov om ændring af lov om folkeskolen, Pub. L. No. LOV nr 313 (2006). Hentet fra <https://www.retsinformation.dk/Forms/R0710.aspx?id=25068>

Møller, H. L., Arnbak, E., Petersen, Dorte Klint, Poulsen, M., Juul, H., & Elbro, C. (2014). *Teknisk Rapport om Ordblindetesten* (s. 45). København: Center for Læseforskning, Københavns Universitet, og Skoleforskningsprogrammet, IUP, Aarhus Universitet. Hentet fra <https://ordblindetest.nu/vejleder/pdf/TekniskRapportOrdblindetesten.pdf>

Nordenbo, S. E., Allerup, P., Andersen, H. L., Dolin, J., Korp, H., Larsen, M. S., ... Østergaard, S. (2009a). *Pædagogisk brug af test. Et systematisk review*. København: Danmarks Pædagogiske Universitetsforlag.

Nordenbo, S. E., Allerup, P., Andersen, H. L., Dolin, J., Korp, H., Larsen, M. S., ... Østergaard, S. (2009b). *Pædagogisk brug af test. Et systematisk review. Teknisk rapport*. København: Dansk Clearinghouse for Uddannelsesforskning.

Norling, M. (2016). De nationale test 2016 - hvor galt står det til? (1) - Folkeskolen.dk. Hentet 12. februar 2018, fra <https://www.folkeskolen.dk/586828/de-nationale-test-2016---hvor-galt-staar-det-til-1>

Novick, M. R. (1966). The axioms and principal results of classical test theory. *Journal of Mathematical Psychology*, 3(1), 1–18. [https://doi.org/10.1016/0022-2496\(66\)90002-2](https://doi.org/10.1016/0022-2496(66)90002-2)

OECD. (2010). *PISA 2009 Results: What Students Know and Can Do*. OECD Publishing. <https://doi.org/10.1787/9789264091450-en>

OECD. (2014). *PISA 2012 - Technical Report*. Paris: OECD. Hentet fra <http://www.oecd.org/pisa/pisaproducts/PISA-2012-technical-report-final.pdf>

OECD. (2017). *PISA 2015 Assessment and Analytical Framework*. OECD Publishing. <https://doi.org/10.1787/9789264281820-en>

Pøhler, L. (2011). Sådan klarer dine elever sig bedre i de nationale test. *Viden om Læsning*, 2011(9), 43–47.

Pøhler, L. (2016). Feedback fra Nationale test - er det en mulighed? *Kvan : et tidsskrift for læreruddannelsen og folkeskolen*, 36(105), 55–70.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danmarks pædagogiske Institut.

Ravn, K. (2015a). Eksperter dumper de nationale test - Folkeskolen.dk. Hentet 12. februar 2018, fra <https://www.folkeskolen.dk/572813/eksperter-dumper-de-nationale-test>

Ravn, K. (2015b). Skoleleder: Testresultater svinger, som vinden blæser - Folkeskolen.dk. Hentet 12. februar 2018, fra <https://www.folkeskolen.dk/572808/skoleleder-testresultater-svinger-som-vinden-blaeser>

Roe, A. (2010). Nasjonale leseprøver – hva de måler og hvordan resultatene kan brukes. *Bedre Skole*, 2010(1), 41–47.

Skar, G. B. (2017). *The Norwegian National Sample-Based Writing Test 2016: Technical Report* (s. 51). Trondheim: Skrivesenteret. Nationalt senter for skriveopplæring og skriveforskning. Hentet fra <http://hdl.handle.net/11250/2467346>

Skolestyrelsen. (2011). *Den socioøkonomiske reference for resultaterne af de nationale test – en vejledning til skoleledere og kommuner* (s. 8). København: Styrelsen for Evaluering og Kvalitetsudvikling af Grundskolen (Skolestyrelsen).

Skov, P. (2009). Evaluering af brugen af det nationale kvalitetsvurderingssystem (NKVS) i grundskolen. I A. P. Allerup, V. Kovac, G. Kvåle, G. Langfeldt, & P. Skov (Red.), *Evaluering av det Nasjonale kvalitetsvurderingssystemet for grunnsopplæringen* (s. 321). Kristiansand: Agderforskning. Hentet fra http://www.udir.no/globalassets/filer/tall-og-forskning/rapporter/2010/5/nkvs_evaluering.pdf

Skov, P., & Brøyn, T. (2010). Bruken av nasjonale prøver en evaluering. *Bedre Skole*, 2010(2), 84–89.

Styrelsen for It og Læring. (2015). *Den adaptive algoritme i De Nationale Test* (p. 8). København: Undervisningsministeriet, Styrelsen for It og Læring. Hentet fra <https://stil.dk/-/media/filer/uvm/udd/folke/pdf15/jan/150128-den-adaptive-algoritme-i-de-nationale-test.pdf?la=da>

Styrelsen for It og Læring. (2016a). *Notat om nationale tests måleegenskaber* (s. 8). København: Undervisningsministeriet.

Styrelsen for It og Læring. (2016b). *Notat om Undersøgelse af de nationale tests reliabilitet* (s. 24). København: Undervisningsministeriet.

Styrelsen for It og Læring. (2017a). Om nationale test. Hentet 11. februar 2018, fra <https://stil.dk/uvm-dk/folkeskolen/elevplaner-nationale-test-og-trivselsmaaling/nationale-test/om-de-nationale-test>

Styrelsen for It og Læring. (2017b). Test- og prøvesystemet – De nationale test. Brugervejledning for skoler. Hentet fra <https://stil.dk/-/media/filer/stil/pdf17/171020-dnt-vejledning-skoler-efteraar-2017.pdf>

Styrelsen for It og Læring. (2017c). *Vejledning om de nationale test - til lærere i alle fag* (s. 32). København: Undervisningsministeriet. Hentet fra <https://www.uvm.dk/-/media/filer/uvm/udd/folke/pdf17/mar/170315-vejledning-til-laerere-i-alle-fag.pdf>

Styrelsen for It og Læring. (2017d). *Vejledning om de nationale test - til skoleledere* (s. 20). København: Undervisningsministeriet. Hentet fra <https://www.uvm.dk/-/media/filer/uvm/udd/folke/pdf17/mar/170315-vejledning-til-skoleledere.pdf>

Styrelsen for It og Læring. (2018). Nationale test 2016-2017. Hentet 11. februar 2018, fra <https://stil.dk:443/support/test-og-proever/nationale-test>

Styrelsen for It og Læring. (2016a). *Bilagsnotat til: De nationale tests måleegenskaber* (s. 12). København: Undervisningsministeriet.

Styrelsen for It og Læring. (u.å.-b). Nationale test og folkeskolens digitale prøver. Hentet 11. februar 2018, fra <https://www.xn--testogprver-ngb.dk/>

Styrelsen for Undervisning og Kvalitet. (2016). *Notat om de nationale test som pædagogisk værktøj* (s. 4). København: Ministeriet for Børn, Undervisning og Ligestilling.

Tengberg, M. (2017). National reading tests in Denmark, Norway, and Sweden: A comparison of construct definitions, cognitive targets, and response formats. *Language Testing*, 34(1), 83–100.
<https://doi.org/10.1177/0265532215609392>

Tengberg, M., Roe, A., & Skar, G. B. (2018). Interrater reliability of constructed response items in standardized tests of reading. *Nordic Studies in Education*, 38(2), 20.

Tengberg, M., & Skar, G. B. (2017). Hur tillförlitligt är det nationella provet i läsning i åk 9? *Utbildning & demokrati*, 26(2), 113–137.

Tresch, S. (2007). *Potential Leistungstest. Wie Lehrerinnen und Lehrer Ergebnisrückmeldungen zur Sicherung und Steigerung ihrer Unterrichtsqualität nutzen*. Bern: H.E.P. Verlag.

Uddannelsesudvalget. (2005). 1. behandling af L 101 Forslag til lov om ændring af lov om folkeskolen. Hentet 8. februar 2018, fra <http://www.ft.dk/samling/20051/lovforslag/l101/beh1-38/forhandling.htm?startItem=-1>

Uddannelsesudvalget. (2006). 2. behandling af L 101 Forslag til lov om ændring af lov om folkeskolen. Hentet 8. februar 2018, fra <http://www.ft.dk/samling/20051/lovforslag/L101/BEH2-66/forhandling.htm>

Undervisningsministeriet. (2004). *OECD-rapport om grundskolen i Danmark – 2004* (Uddannelsesstyrelsens temahæfteserie No. 5). København. Hentet fra <http://pub.uvm.dk/2004/oecd/oecd.pdf>

Undervisningsministeriet. (2017). Om nationale mål. Hentet 12. februar 2018, fra <https://www.uvm.dk:443/folkeskolen/folkeskolens-maal-love-og-regler/nationale-maal/om-nationale-maal>

Undervisningsministeriet. (2018). Vejledninger – om baggrund, afvikling og brug af testresultater. Hentet 11. februar 2018, fra <https://www.uvm.dk/folkeskolen/elevplaner-nationale-test-og-trivselsmaaling/nationale-test/vejledninger>

Vagle, W., Roe, A., & Narvhus, E. (2009). *Den nasjonale prøven i lesing på 8. trinn, 2008. Rapport fra utvalgsundersøkelsen* (s. 24). Oslo: Institutt for lærerutdanning og skoleutvikling, Universitetet i Oslo. Hentet fra https://www.udir.no/globalassets/upload/nasjonale_prover/2008/analysen/5/np08_lesing8_analyse.pdf

Virinkoski, R, Lerkkanen, M.-K., Eklund, K., Holopainen, L., & Aro, M. (2017). Teachers' Ability to Identify Children at Early Risk for Reading Difficulties in Grade 1. *Early Childhood Education Journal*. DOI 10.1007/s10643-017-0883-5

Vygotskii, L. S. (1978). *Mind in society : the development of higher psychological processes*. Cambridge: Harvard University Press.

Weyergang, C., & Ryen, J. A. (2014). Nasjonale leseprøver – hvordan bruke resultatene. *Bedre Skole*, 2014(3), 39–44.

Wilson, M., & Gochyyev, P. (2013). Psychometrics. I T. Teo (Red.), *Handbook of quantitative methods for educational research*. Rotterdam: SensePublishers.