



Er nationale test et pædagogisk redskab?

JEPPE BUNDSGAARD, PROFESSOR MSO, DPU,
AARHUS UNIVERSITET

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 undersøge, om nationale test faktisk er blevet et sådant redskab. Fokus vil være på nationale test i dansk, læsning.

Artiklen genudgives i forkortet udgave efter tilladelse fra forfatteren og det norske tidsskrift Sakprosa. Den kan læses i sin fulde længde i Sakprosa, nov. 2018: <https://www.journals.uio.no/index.php/sakprosa/article/view/6007/5545>

Artiklen er forkortet af Viden om Literacys redaktører Henriette Romme Lund og Helle Bundgaard Svendsen.

De nationale test

Nationale test blev indført som obligatorisk test i 2010. Indtil da var der i Danmark ikke lovmæssigt krav om test af elevernes kompetencer, ud over afgangsprøverne i 9. og (den frivillige) i 10. klasse.

Nationale test er computerbaserede, adaptive og multiple choice eller udfyldning af tal. De indeholder 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. I nationale test og andre test, der anvender Rasch-modellen, måles testopgavernes (items) sværhedsgrader og elevers dygtighed begge på den såkaldte logitskala (for forklaring af Rasch-modellen og logitskala, se boks på s. 65). 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 0,55 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 sigens natur nogle grænser for, hvilke typer test-opgaver testen kan indeholde. Da testen er adaptiv, er der desuden behov for mange opgaver inden for alle sværhedsgradsniveauer, således at computeren kan vælge testopgaver ud med en passende sværhedsgrad. Konkret er der flere tusind opgaver i hver af de omkring 15 testbanker.

Rasch-modellen og logitskala

Den danske matematiker Georg Rasch udviklede i 1950'erne en ny måde at beregne elevers dygtighed på. Hvor den klassiske testteori talte rigtige svar op for en test og betragtede det samlede resultat som elevens dygtighed, var Raschs' pointe, at opgaver (items) har *forskellige* sværhedsgrader, og at den, der kan besvare 8 opgaver, ikke nødvendigvis er dobbelt så god som den, der kan svare på 4. Rasch gjorde opmærksom på, at man ikke enten kan eller ikke kan besvare en opgave, men at man har en vis sandsynlighed for at svare rigtigt på opgaven – en sandsynlighed, som bliver større, des bedre man er. Man siger, at Rasch-modellen er en *probabilistisk* model (altså at den er knyttet til sandsynlighed). På den baggrund skabte han en forbindelse mellem elevers dygtighed og opgavers sværhedsgrad ved at sige, at elever med en given dygtighed har 50 % sandsynlighed for at svare rigtigt på en opgave med samme sværhedsgrad som elevens dygtighed.

Det fortæller således ikke ret meget om elevens dygtighed, at hun kan svare på én opgave. Der skal flere opgaver til. Des flere opgaver eleven får, des mere sikker kan man være på sit estimat på, hvor dygtig hun er. Ingen test er helt præcis, men nogle test er mere præcise end andre.

Når man har fastsat opgavers sværhedsgrad, så kan man tage et udsnit af de opgaver, man har, og give dem til eleven og estimere dygtigheden ud fra dem. Næste gang man vil måle elevens dygtighed, kan man tage nogle andre opgaver fra samme opgavebank og derved undgå, at eleven kan huske opgaverne fra forrige gang, men alligevel få information om, hvor meget

bedre eleven er blevet i tiden mellem de to tests. Det er den egenskab ved Rasch-modellen, man benytter i nationale test og i internationale undersøgelser som PISA, ICILS, PIRLS og TIMSS.

Både elevers dygtighed og opgavers sværhedsgrad måles på den såkaldte logitskala, der er en logaritmisk skala, der går fra minus til plus uendelig, men for det meste har værdier mellem plus og minus seks eller mindre. Logitskalaen er kontinuert (i modsætning til når man bare tæller antal rigtige), og derfor kan man tage gennemsnit af elevers dygtighed, og man kan tale om afstande – så man fx kan sige, at dem, der har bevæget sig to logit, har bevæget sig dobbelt så meget som dem, der har bevæget sig én logit.

Hvilke muligheder er der for at bruge test som pædagogisk redskab?

For at kunne besvare spørgsmålet: *Er nationale test et pædagogisk redskab?* præsenterer jeg først mulighederne for at bruge standardiserede og normerede test som pædagogiske redskaber. Det sker i de nedenstående seks afsnit, hvor jeg, på baggrund af internationale review og ministeriets opstilling af krav til et pædagogisk værktøj, giver 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 test 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 test 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 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 zonen for nærmeste udvikling for eleverne inden for det målte område?
- ▶ Hvad kan jeg gøre som lærer i forhold til de forskellige grupper?

Jeg stiller [...] seks spørgsmål, som, hvis de kan besvares med testen, kan gøre den til et pædagogisk redskab.

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. For en test vil altid kun måle et uddrag af et fagligt område, og den vil altid gøre det fra et særligt perspektiv.

Når man skal udvikle test, så er det afgørende, at der udarbejdes en ramme for den. For det første skal der udarbejdes en præcis definition af det fænomen, der ønskes målt, og der skal foreligge en hypotese for progression inden for området, sådan at der skabes et kort over bevægelsen fra ikke at være så god til at have fuld kontrol over det målte fænomen (Wilson & Gochyyev, 2013). Dette arbejde er afgørende vigtigt for udviklingen af testen som helhed, for det skaber et grundlag for udviklingen af testopgaver, og det gør, at brugere af testen (i dette tilfælde læreren) efterfølgende kan forstå, hvad det er meningen, at testen måler – og ikke mindst, hvad den ikke måler.

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 test af andre af de faglige områder, man som lærer og skole lægger vægt på.

Hvordan klarer eleverne i klassen sig i forhold til kommunens eller landets elever?

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-test, 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 klasses 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 socioøkonomiske 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.

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 testopgavernes 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.

Hvilke elever har problemer i forhold til det faglige område, der måles?

En meget væsentlig viden for en lærer er, hvilke elever der har problemer i forhold til et fagligt område.

Ofte vil læreren have særlig opmærksomhed på elever, som har faglige udfordringer eller måske endda handicaps, fx ordblindhed. Mange test er udviklet til at identificere netop sådanne elever. Men også test, der ikke er udviklet med henblik på at identificere elever med specifikke vanskeligheder, kan bruges til at identificere elever, som skal undersøges nærmere gennem opfølgning, fx i form af mere specifikke test eller særlig observation af en fagperson.

Hvad er zonen for den nærmeste udvikling for eleverne inden for det målte område?

De hidtil nævnte spørgsmål 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 med test, der er udviklet efter Rasch-modellen og dermed på baggrund af et veldefineret fænomen for måling, og med testopgaver, der måler fænomenet 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 *empiriske kompetenceniveaubeskrivelser (described proficien-*

cy scales) (Bundsgaard, submitted; Fraillon et al., 2015; OECD, 2014).

En empirisk kompetenceniveaubeskrivelse udvikles ved at inddele skalaen i et antal intervaller og bede faglige eksperter undersøge de testopgaver, 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.

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 den nærmeste udvikling (Vygotskii, 1978, s. 86) er for elever med forskellige dygtigheder.

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.

Hvad kan jeg gøre som lærer i forhold til de forskellige grupper?

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 kompetenceniveaubeskrivelser, 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 et al., 2009a).

Pædagogisk brug af nationale test

I den følgende analyse af nationale test som pædagogisk redskab forsøger jeg at besvare hvert af de seks spørgsmål ovenfor i forhold til de nationale test i læsning.

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 der tre profilområder: 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 består af tre til fire kompetenceområder i hvert fag. I dansk er der fire kompetenceområder, og det er kun kompetenceområdet læsning, testen relaterer til.

Hvert kompetenceområde består af ni trin (et for hver årgang, men ikke knyttet direkte til et klasse-

trin). 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, og tre af dem er sammenfaldende med de tre profilområder for nationale test, Afkodning, Sprogforståelse og Tekstforståelse. Man kunne derfor forvente, at Fælles Mål for dansk, læsning udgjorde en definition af, hvad der måles i nationale test i dansk, 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, hvor eleven skal sætte to streger i en række af tre sammensatte ord, og derfor ikke kan teste, fx om eleven kan afpasse sin læsehastighed efter læseformål, eller om denne kan læse hurtigt og sikkert (flydende), og det gælder fx for Tekstforståelse, som ikke tester strategier ved manglende forståelse, brug af modeller, om eleven kan sammenholde forskellige tekster osv.

Det er ikke lykkedes mig at finde nogen yderligere angivelse af, hvad der måles, og hvordan der vælges konkret indhold i testen. 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 i forhold til kommunens eller landets elever?

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 klarer sig i forhold til populationen (i praksis er det en historisk population, typisk 2010 eller 2014). Den ene normbaserede score angives som procentilscore i form af en talværdi, der ligger mellem 1 og 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 gennemsnittet (kategorierne kaldes Klart over middel, Over middel, Middel osv.).

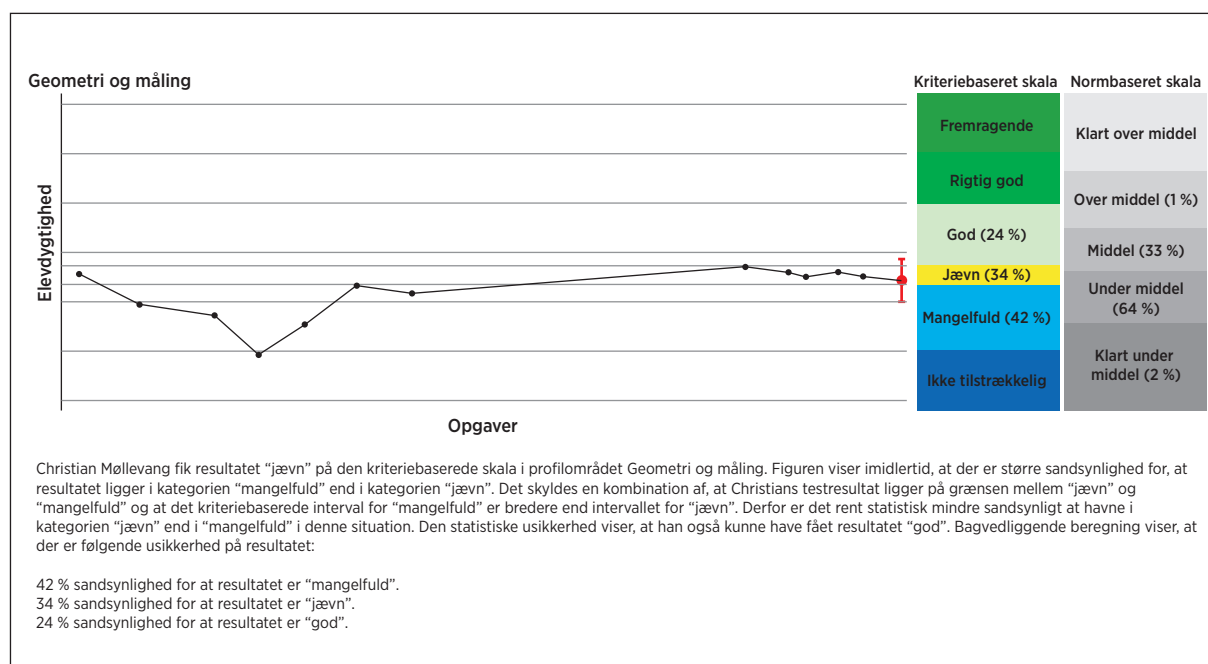
De normbaserede scores giver således læreren indsigt i, hvordan eleverne klarer sig i forhold til andre skolers elever. Men læreren får ikke information om, hvordan elever fra lignende skoler og socio-økonomiske baggrunde klarer sig, og derfor skal hun selv tage højde for, hvilken baggrund klassens elever 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 1 som er et uddrag fra *Vejledning til nye resultatvisninger i de nationale test*. Den røde horisontale streg angiver 68 procent konfidensintervallet, dvs. at elevens resultat ligger inden for det interval med 68 procent 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 procent 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.

Dette eksempel er 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?



Figur 1. Illustration af usikkerhed i estimat af en elevs dygtighed.

Er eleverne blevet bedre i løbet af den forgangne periode?

I dansk, læsning skal eleverne tage nationale test flere gange over deres skolegang, men hver test inden for det pågældende fag har sin egen bank af testopgaver, 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 test 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” test). 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 med det nye testresultat på x-aksen og det tidligere resultat på y-aksen. Dog fremgår 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). Derfor må læreren 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).

Hver test inden for det pågældende fag har sin egen bank af testopgaver, og der er ikke gennemført en statistisk analyse af, om de måler det samme.

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.

Hvilke elever har problemer i forhold til det faglige område, der måles?

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).

Hvad er zonen for den nærmeste udvikling for eleverne inden for det målte område?

De kriteriebaserede test er 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 zonen for deres nærmeste udvikling er.

Hvad kan jeg gøre som lærer i forhold til de forskellige grupper?

Da der som netop omtalt ikke gives information om, hvilke typer af faglige udfordringer eleverne på forskellige kriterieniveauer 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 af de færdigheder, der testes i nationale test, og for nogles vedkommende imiterer de sågar testopgaveformatet (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 (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 testenes 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 usikkerheder på resultaterne (se figur 1).

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; Under-

visningsministeriet, 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 skrev i indledningen, at det politiske argument for at indføre nationale test var, at lærerne 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.

I dag fungerer den samlede population af elevers testresultater som resultatmål for den skolereform, der blev gennemført i Danmark i 2014. Det bevirker, at testene får en anden rolle, end den oprindeligt var tildelt. 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 [...], for de fører til *teaching to the test* og til strategiske handlinger fra elever, lærere og skolelederes side.

Test, der er *high stakes*, er der stærkt belæg for at advare imod (Nordenbo et al., 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.

Referencer

- Bundsgaard, J., & Puck, M. R. (2016). *Nationale test – danske lærere og skolelederes brug, holdninger og viden*. DPU, Aarhus Universitet. Lokaliseret den 30. januar 2019 på: http://pure.au.dk/portal/files/102516329/Ebog_Nationale_test_FINAL_september_2016.pdf
- Bundsgaard, J. (2017). *Så upræcise er nationale test*. Folkeskolen.dk. Lokaliseret d. 12. februar 2019 på: <https://www.folkeskolen.dk/583103/saa-upraecise-er-nationale-test>
- Bundsgaard, J. (submitted). 21st Century Skills in Project-Based Learning. *Journal of Educational Assessment*.
- Clio Online (2018). *Clio Træningsøvelser* (kræver login). Lokaliseret d. 12. februar 2019 på: <https://mit.cloonline.dk/cloonline.myclio/archive/training/>
- 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). dansk0-2.gyldendal.dk Afkodning (kræver login). Lokaliseret d. 12. februar 2019 på: http://dansk0-2.gyldendal.dk/en/Indgange/Laerer/Om_traeningsdelen/Fagomraader/afkodning.aspx
- Nordenbo, S. E., Allerup, P., Andersen, H. L., Dolin, J., Korp, H., Larsen, M. S., ... Østergaard, S. (2009). *Pædagogisk brug af test. Et systematisk review*. København: Danmarks Pædagogiske Universitetsforlag.
- Norling, M. (2016). *De nationale test 2016 – hvor galt står det til?* (1). Folkeskolen.dk. Lokaliseret d. 12. februar 2019 på: <https://www.folkeskolen.dk/586828/de-nationale-test-2016---hvor-galt-staar-det-til-1>
- OECD (2014). *PISA 2012 – Technical Report*. Paris: OECD. Lokaliseret d. 6. marts 2019 på: <http://www.oecd.org/pisa/pisaproducts/PISA-2012-technical-report-final.pdf>
- 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). *Ekspertes dumper de nationale test*. Folkeskolen.dk. Lokaliseret d. 12. februar 2019 på: <https://www.folkeskolen.dk/572813/eksperter-dumper-de-nationale-test>
- Ravn, K. (2015b). *Skoleleder: Testresultater svinger, som vinden blæser*. Folkeskolen.dk. Lokaliseret d. 12. februar 2019 på: <https://www.folkeskolen.dk/572808/skoleleder-testresultater-svinger-som-vinden-blaeser>
- Styrelsen for It og Læring (2015). *Nationale test og folkeskolens digitale prøver*. Lokaliseret d. 11. februar 2019 på: <https://www.xn--testogprøver-ngb.dk/>
- Styrelsen for It og Læring (2017a). *Om nationale test*. Lokaliseret d. 11. februar 2019 på: <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. Lokaliseret d. 6. marts 2019 på: <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*. København: Undervisningsministeriet. Lokaliseret d. 6. marts 2019 på: <https://uvm.dk/folkeskolen/elevplaner-nationale-test-og-trivselsmaaling/nationale-test/vejledninger>

Styrelsen for It og Læring (2017d). *Vejledning om de nationale test – til skoleledere*. København: Undervisningsministeriet. Lokaliseret d. 6. marts 2019 på: <https://www.uvm.dk/-/media/filer/uvm/udd/folke/pdf17/mar/170315-vejledning-til-skoleledere.pdf>

Undervisningsministeriet (2004). *OECD-rapport om grundskolen i Danmark – 2004* (Uddannelsesstyrelsens temahæfteserie No. 5). København. Lokaliseret d. 6. marts 2019 på: <http://static.uvm.dk/publikationer/2004/oecd/index.html>

Undervisningsministeriet (2018). *Vejledninger – om baggrund, afvikling og brug af testresultater*. Lokaliseret d. 11. februar 2019 på: <https://www.uvm.dk/folkeskolen/elevplaner-nationale-test-og-trivselsmaaling/nationale-test/vejledninger>

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

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



Tegning: Anna Victoria Bingen-Svensson