



# TEST SKAL STØTTE PÆDAGOGIKKEN – IKKE SABOTERE DEN

De nationale test har mødt bred kritik, og politikerne har besluttet at skabe et nyt testsystem. Men har vi overhovedet brug for at teste eleverne i skolen? Ja, lyder det samstemmende fra tre forskere, som Asterisk har stillet spørgsmålet. Hvis systemet vel at mærke måler korrekt og tester meningsfuldt, og hvis formålet er klart og pædagogikken i højsædet.

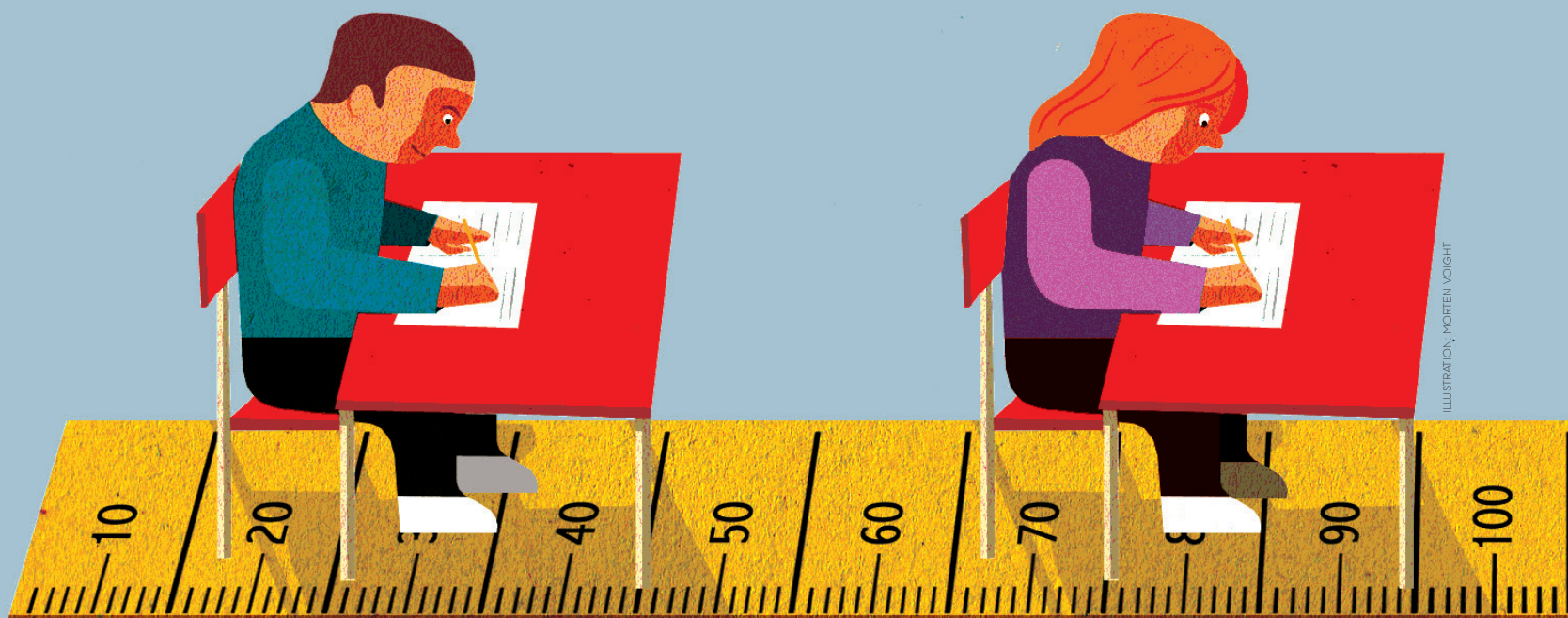


ILLUSTRATION: MORTEN VOIGHT

**I** ti år har de nationale test været en fast del af programmet for alle danske skoleelever. Og lige så længe har testene delt de skolepolitiske vande. På den ene side argumenterne for at styrke evalueringkulturen i folkeskolen med standardiserede test, på den anden side kritikken af testen for at måle de forkerte ting og måle dem forkert eller upræcist og for at være ubrugelig pædagogisk, for at være et politisk styringsredskab, der gør mere skade end gavn ude i landets folkeskoler og ikke mindst for at mangle transparens. Men selve idéen om en national test ser ud til at være kommet for at blive. I februar 2020 vedtog Folketinget at udvikle og indføre et nyt og forbedret nationalt evaluering- og bedømmelsessystem. Asterisk har spurgt tre testforskere, hvad det er vigtigt at tage højde for i det nye system.

Vi har brug for test i skolen af flere grunde, forklarer Christian Christrup Kjeldsen, viceinstituttleder for forskning ved DPU, Aarhus Universitet, og leder af Nationalt Center for Skoleforskning.

”Med test ønsker vi at give lærerne et didaktisk værktøj til at stikke en finger i jorden og se, hvordan eleverne responderer på undervisningen. Vi ønsker dels at kunne identificere elever, der ’hænger’ og har behov for særlig hjælp, dels at kunne få øje på under-

»En moderne test skal i langt højere grad, end vi ser i dag, afspejle det skolecurriculum, vi nu engang har vedtaget.«

Christian Christrup Kjeldsen

visningsområder, der skal have særligt fokus eller en ekstra indsats. Vi ønsker at kunne give forældre en tilbagemelding på, hvordan deres børn klarer sig i skolen. Og vi ønsker at give politikere, myndigheder og forvaltninger mulighed for at følge med i, hvordan det går med grundskolen samlet set. Det har været diskuteret, om én og samme test kan tjene flere herrer. Men der er ikke nogen ’one size fits all’-løsning,” siger han.

I stedet bør der udvikles et samlet test- og evalueringssystem med værktøjer, der hver især er anvendelige på forskellige områder i skolens hverdag, og som tilsammen kan give den indsigt, vi har brug for, mener han. Et sådant program får man ikke fra den ene dag til den anden.

”I mit perspektiv er det derfor okay, hvis vi begynder ét sted og så lægger til hen ad vejen,

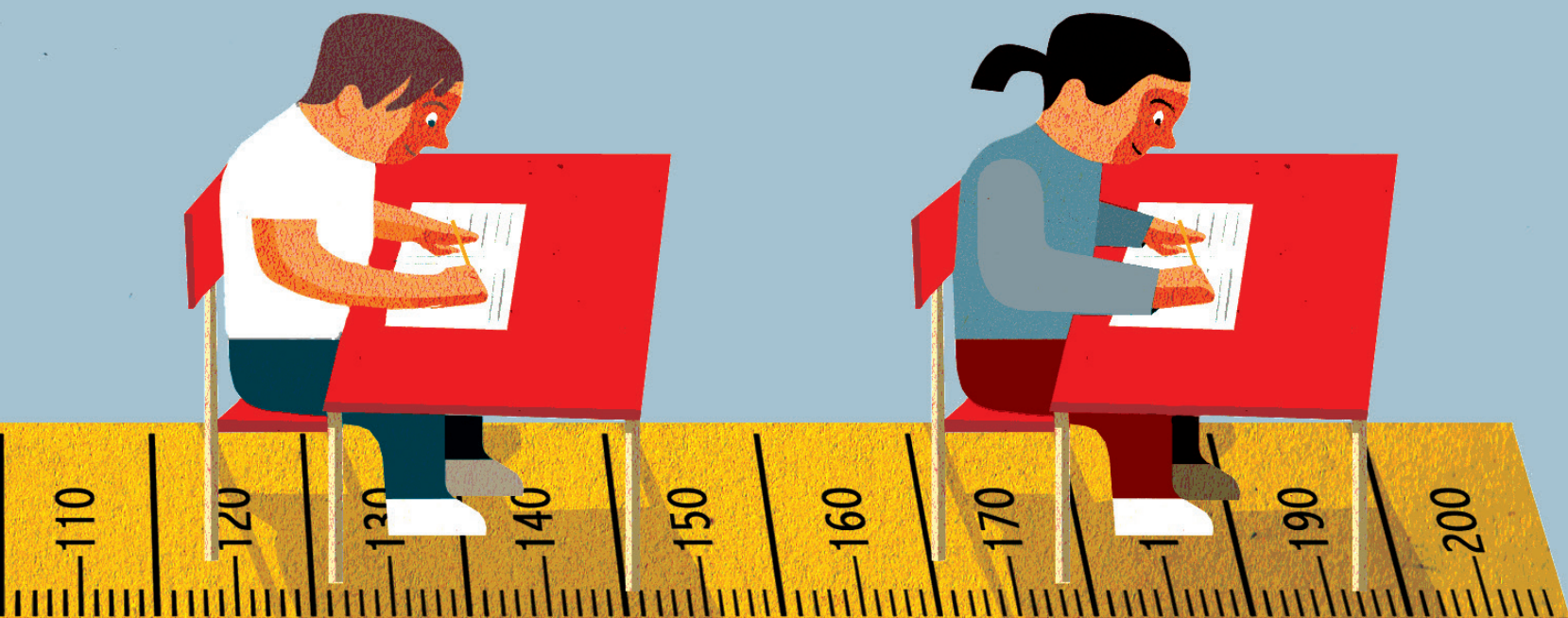
hvis vi vel at mærke er indstillede på at revidere udgangspunktet, når vi bliver klogere undervejs. Min anbefaling er, at vi beslutter, på hvilke klassetrin og i hvilke fag vi som minimum vil teste – og så gør det ordentligt dér. Ambitionen må være større end bare at smøre smørret tyndt ud over hele skolen. Så hellere gøre det lidt færre gange og i færre fag i starten, men til gengæld gøre det godt,” siger han.

#### Tættere på undervisningen

Indholdet i fremtidens test bør gentænkes, og det kræver enighed om, hvad det er for færdigheder og kompetencer, vi egentlig ønsker at teste, mener Christian Christrup Kjeldsen.

”En moderne test skal i langt højere grad, end vi ser i dag, afspejle det skolecurriculum, vi nu engang har vedtaget. Testopgaverne må være i tråd med den faglige virkelighed, eleverne møder i undervisningen,” siger han og peger samtidig på, at hvis fremtidens test skal tjene didaktikken, kræver det også større transparens:

”Det ville være ønskværdigt, hvis man efter hver testrunde kunne lægge opgaverne frem og få en faglig, didaktisk diskussion om, hvorvidt det er lykkedes at få opgaverne til at afspejle den skolevirkelighed, vi gerne vil teste. Når nu testen indgår som en central del af skolens praksis, og når vi nu har tradition for brede politiske forlig om folkeskolen, og når dens opgave i øvrigt er at danne eleverne



til demokratiske medborgere, så er det et problem, at opgaverne og de områder, man tester i, ikke kan diskuteres demokratisk.”

En model kunne ifølge Christian Chrstrup Kjeldsen være at teste dels i et kerne-curriculum, som alle elever skulle gennemføre, og dels inden for mere specifikke områder af curriculum, som udvalgte grupper af elever testes i.

”På den måde kan man få et gennemsnitligt billede af, hvordan eleverne samlet set eller grupper af elever med bestemte baggrunde klarer sig i forhold til fagenes samlede curriculum. Altså en mellemting mellem, at alle testes i det samme, og at hver elev har sin egen unikke test, som vi ser i de nationale test i dag. Det giver mulighed for at blive klogere på, om der fx er områder, der ikke er fokus nok på i undervisningen, eller hvor eleverne har særlige vanskeligheder. Det kan man så arbejde mere målrettet med, både på de enkelte skoler og i de enkelte klasser, men også nationalt på læreruddannelsen, i læreplanerne og i udviklingen af læremidler. Det er denne vekselvirkning mellem de forskellige niveauer, en moderne test bør bidrage til,” siger han.

### Testen er videnskabelig

Ifølge Christian Ydesen, professor MSO ved Aalborg Universitet, bliver skole- og uddannelsespolitiske diskussioner i dag alt for ofte til et spørgsmål om for eller imod nationale test.

”Den diskussion er gold. Vigtigere er grundlagsdiskussionen om, hvad det er, skolen egentlig skal kunne. Hvad vi vil med uddannelse i vores samfund. Så kan man herudfra diskutere og beslutte, hvad det er vigtigt at få dækket ind i et bedømmelsessystem,” siger han.

Nationale test bør nemlig ses i et større evalueringsspektiv.

”Test er jo en specifik måde at foretage en evaluering, dvs. en vurdering eller en bedømmelse, på inden for uddannelse. Så langt vi kan se tilbage, har der været interesse for at undersøge, om elever har lært det, de skulle. Vi ser eksempler på eksamen tilbage i den romerske republik i det første århundrede før vores tidsregning, og vi kender det fra katekismus-overhøringerne efter Reformationen i 1500-tallet. At bedømme præstationer og faglige fremskridt er en integreret del af uddannelse,” siger Christian Ydesen.

På dansk skelner vi mellem test, prøve og eksamen. Eksamen ligger i slutningen af et uddannelsesforløb, mens prøverne typisk ligger undervejs. Begge er faglige evalueringer

»At testen gør krav på at være videnskabelig og objektiv, giver den i nogles øjne en særlig autoritet, mens det i andres øjne netop er det, der gør den problematisk.«

Christian Ydesen



foretaget af lærere og censorer og adskiller sig derfor fra testen, der gør krav på at være videnskabelig.

”At testen er videnskabelig vil blandt andet sige, at den genererer standardiserede data, der gør det muligt at sammenligne resultater og afgøre, hvor en elev, en klasse, en skole eller et land ligger i forhold til andre. Testen forsøger at neutralisere betydningen af andre faktorer end det, den tester for. At testen gør krav på at være videnskabelig og objektiv, giver den i nogles øjne en særlig autoritet, mens det i andres øjne netop er det, der gør den problematisk,” siger Christian Ydesen.

### Intelligenstestens indtog

Videnskabelig i sin ambition var også intelligenstesten, der oprindeligt blev udviklet af den franske psykolog Alfred Binet i begyndelsen af det 20. århundrede, og som førte til, at den tyske psykolog William Stern i 1912 formulerede den første definition af intelligenskvotienten IQ.

”Intelligenstesten var båret af en forestilling om, at man objektivt kunne måle et menneskes intelligens. Hvor mange megahertz kører hjernen med derinde? På samme måde har man med faglige test en forestilling om, at der findes vidensområder eller færdigheder, der kan defineres sådan, at det er muligt at foretage nogle prøveboringer og ud fra resultaterne rekonstruere, hvordan de pågældende vidensområder eller færdigheder faktisk ser ud hos den pågældende person,” siger han.

I tiden mellem 1. og 2. verdenskrig havde danske reformpædagoger en idé om, at intelligenstesten skulle erstatte eksamen.

”De anså eksamen for at være subjektiv, fordi lærere og censorers forskelsbehandling af eleverne og fordomme om dem kunne spille ind på vurderingen. Eksamen var noget, der reproducerede de sociale uligheder i fx landsbyskolen, hvor storbondens børn sad på første række og husmandens på bageste. Den videnskabelige intelligenstest derimod gjorde det sammen med standpunktstest muligt på objektiv vis at afdække børnenes reelle evner og potentialer og fx identificere det intelligente arbejderbarn, der kunne blive statsminister,” siger Christian Ydesen.

Nationale test er et eksempel på en videnskabelig testpraksis, der er udsprunget af et politisk ønske om en styrket evalueringsskulptur i skolen. Intelligenstesten er et eksempel på det samme, men til gengæld udsprunget af et psykologisk og pædagogisk fagmiljø.

”I mellemkrigstiden var der ikke politisk vilje til at udskifte eksamen med det, man mente var en objektiv test. I dag er fortegnet omvendt. På den ene side var der et politisk ønske om på standardiseret måde at kunne følge med i, hvordan det gik i skolen, på den anden side et pædagogisk princip om, at evaluering måtte være forbundet med lærerens professionelle dømmekraft,” siger han.

### Datakyndighed er påkrævet

Hvor intelligenstesten med sit udspring i de pædagogiske og psykologiske fagmiljøer groede frem nede fra, er de såkaldte ’ministerielle prøver’, der blev søsat i 1915 og afskaffet i 1954, et eksempel på et kontrolinstrument, der ligesom nationale test udsprang fra politisk hold. Prøverne blev indført for at få viden om, hvilket fagligt niveau, eleverne egentlig havde, når de gik ud af folkeskolen. Dengang som nu var prøverne drevet af bekymring for, at danske elever ikke var dygtige nok. De ministerielle prøver foregreb nogle af de erfaringer, vi nu også har gjort os med de nationale test, mener Christian Ydesen.

Nemlig, at når man først er begyndt at producere testdata, risikerer de at blive brugt til noget andet end det oprindelige formål.

”Vi har eksempler på, at fx skoleledere dengang begyndte at rekvirere resultaterne fra ministeriet og brugte dem til at slå lærerne oven i hovedet med, hvis deres elever havde klaret sig dårligt. Det var jo ikke meningen med det, ligesom det heller ikke oprindeligt var meningen at lave ranglister over danske skoler på baggrund af data fra de nationale test, sådan som vi har set tænketanke og kommunale kvalitetsrapporter gøre det,” siger han.

Selve det politiske ønske om at vide, hvordan det går ude i skolerne, er dog helt legitimt, mener Christian Ydesen.

”Her vil jeg være fortalende for et system af de ministerielle prøver, hvor man udtager en statistisk repræsentativ del af skolerne til testen og samtidig sikrer, at den enkelte skole fx kun testes hvert femte år. Det vil være nok til at give den nødvendige pejling til myndighederne, og vi undgår så, at alle skoler skal teste hvert år med alle de uheldige sideeffekter, det giver,” siger han.

Ifølge Christian Ydesen er historiens lære til os fra de ministerielle prøver over inteligenstesten til de nationale test, at test altid kræver datakyndighed på alle niveauer af uddannelsessystemet.

”Om det er ministerielle embedsmænd, Børn- og Unge-udvalgsmedlemmer i kommunerne, skoleledere, lærere eller forældre – ja elever for den sags skyld – så er datakyndighed en vigtig kompetence. Hvad kan give data, og hvad kan de ikke, hvordan er de blevet til, hvad er formålet med dem, og hvad er det ikke meningen, de skal bruges til? Folk kan være tilbøjelige til at tolke data ud fra deres egen agenda. At være datakyndig er fx at vide, at testdata er abstraktioner, der forsøger at neutralisere betydningen af den kontekst, hvori de bliver til, men at resultatet skal ses i forhold til de betingelser, testen er taget under. Og at de er øjeblikksbilleder, der fx ikke får øje på skoler i udvikling,” siger han.

### Aldrig kun én chance

En test bør hverken være eller kunne opleves som uigenkaldelig, mener Christian Ydesen.

”En test som den kinesiske ’gao kao’-test, der fx afgør, hvilket universitet de unge kan komme ind på, er en ’high stakes’-test, fordi én og kun én test afgør ens livsmuligheder. Men en test bør altid give mulighed for at kvalificere sig, hvis ikke på én så på en anden måde, eller for at man kan tage testen igen

## »Man er nødt til at erkende, at uddannelse også handler om økonomi og geopolitik.«

Christian Ydesen

### GODT AT VIDE OM NATIONALE TEST

#### Adaptiv

De nationale test er adaptive. Testen skyder sig gradvist ind på elevens faglige niveau: De opgaver, hun får, afhænger af, hvordan hun har besvaret de forudgående opgaver. Hver gang en opgave er besvaret, beregnes et midlertidigt niveau for eleven, og testsystemet vælger derefter en opgave med en sværhedsgrad, der ligger så tæt som muligt på elevens niveau: En lidt lettere opgave, hvis den forrige er besvaret forkert, og en lidt sværere opgave, hvis den er besvaret korrekt. Dermed giver hver opgave mest mulig information til den næste beregning af elevens niveau. Når målesikkerheden på elevens niveau er lav nok, stopper testen. Hver elev får derfor sit eget unikke testforløb i modsætning til en lineær test, hvor alle elever får de samme opgaver. En adaptiv test kræver, at testen er computerbaseret.

#### Rasch-model

Det statistiske grundlag for beregning af sammenhængen mellem elevens dygtighed og deres besvarelse af opgaver i de nationale test bygger på Rasch-modellen, udviklet af den danske matematiker og statistiker Georg Rasch (1901-1980). Modellen, der angiver sandsynligheden for, at en person med en bestemt dygtighed svarer korrekt på en given opgave alene ud fra personens dygtighed og opgavens sværhedsgrad, er anerkendt og anvendt internationalt.

Kilde: Tine Nielsen

efter et stykke tid. Ved at lade forskellige bedømmelsessystemer spille sammen, kan man tage luften ud af den enorme byrde, en ’high stakes’-test ellers kan være,” siger han.

De nationale test er officielt ’low stakes’-test. Elevernes livsbaner afgøres ikke af, hvordan de klarer sig i den nationale test i matematik i 3. klasse. Men ifølge Christian

Ydesen kan selve oplevelsen af, at der er meget på spil, gøre testen til en ’high stakes’-test.

”I forskningen defineres det typisk sådan, at hvis en test opleves som ’high stakes’, så er den det også. Og når nogle elever har en oplevelse af, at der er meget på spil i testen i forhold til kammeraterne, læreren eller forældrene og bliver nervøse, kede af det eller frygtsomme forud for eller under testen, ja så er der meget på spil for dem socialt og psykologisk – også selvom det ikke officielt har konsekvenser for dem. Og når vi har set eksempler på, at skoleledere over for deres lærere italesætter det som et problem, at deres skole klarer sig dårligere end andre, og lærerne derfor begynder at nære frygt for, at deres elever scorer dårligt flere år i træk, fordi det måske i værste fald kan få konsekvenser for deres ansættelse, så er de nationale test blevet en de facto ’high stakes’-test. Det bør vi undgå, når fremtidens test skal konstrueres, og en af måderne at gøre det på er ved at sikre, at data kun bruges til det, der er meningen med testen,” siger han.

Han tilføjer, at det må være et kendetegn ved enhver god test af færdigheder, at den tager afsæt i og afspejler, at mennesker er forskellige og kan noget forskelligt.

”En test kan ikke være en test uden en vis standardisering af den menneskelige mangfoldighed. Testen må dog ikke begrænse diversiteten med rigide standarder for menneskelig udvikling, men skal tværtimod skabe rum for den, hvad enten den er kulturel, social eller kognitiv,” siger han.

### God pædagogik eller geopolitik?

Ranglister over, hvordan landenes uddannelsessystemer klarer sig i sammenligning med hinanden, går tilbage til den kolde krig, fortæller Christian Ydesen. I 1983, hvor Ronald Reagan var præsident i USA, udarbejdede hans regeringsadministration rapporten *A Nation at Risk*. Den advarede om, at USA ville tabe Den kolde krig, hvis ikke uddannelsessystemet blev bedre.

”Reagan-administrationen bad OECD udvikle indikatorer, der ville gøre det muligt at sammenligne landenes uddannelsesniveauer. Tankegangen var, at vi er nødt til at vide, hvor vi står, så vi kan styrke uddannelsessystemet og blive bedre end Østblokken. Dengang var der mange i OECD, der mente, at det var useriøst at udvikle standardiserede indikatorer, men projektet blev sat i søen, og det første skridt mod det, der senere blev til PISA-testen, var dermed taget,” siger han.

Han mener, at denne sammenhæng mellem uddannelse og geopolitik nogle gange

bliver overset i diskussioner om pædagogiske teorier og metoder. På den første OECD-konference om uddannelse i 1961 blev det fremført, at uddannelse er for vigtigt til, at man kan overlade til det til lærere og pædagoger – dertil er der alt for store økonomiske og geopolitiske interesser på spil.

”Sådan vil jeg selvfølgelig ikke formulere det, men jeg vil heller ikke plædere for pædagogikkens renhed ved at vende det om og sige, at uddannelse er for vigtigt til, at vi kan overlade det til økonomer og politikere. Man er nødt til at erkende, at uddannelse også handler om økonomi og geopolitik. Spørgsmålet om, hvordan vi afvejer styring af uddannelse gennem standardiserede test mod pædagogiske og didaktiske mål og ideer, bliver derfor centralt i udviklingen af et nyt testsystem,” siger Christian Ydesen.

### Hvad kan måles?

Tine Nielsen er lektor i psykologi på Københavns Universitet og en af de eksperter i psykometri og testning, der kritisk vurderede Børne- og undervisningsministeriets psykometriske dokumentation i forbindelse med VIVE's evaluering af de nationale test i begyndelsen af 2020. Psykometri er den videnskabelige disciplin, der beskæftiger sig med at måle mentale egenskaber og tilstande, og består af psykologisk statistiske metoder til konstruktion og afprøvning af test. Spørger man hende, hvad man meningsfuldt kan måle i forhold til skole og uddannelse, lyder svaret: ”Rigtig meget!”

”Man kan måle evner og færdigheder, men også læringspræferencer, stress eller trivsel. Det er først og fremmest et spørgsmål om at gøre det ordentligt. Både psykometrisk ordentligt, men også ordentligt i pædagogisk og didaktisk forstand. Kvaliteten af målingen skal naturligvis være meget høj, ligesom det bør være fuldstændigt klart, hvad det er, man vil måle, hvorfor man vil måle det, og hvad man vil bruge det til,” siger hun.

Det skal fx være klart, om målet med en test er at skaffe viden til politikere, eller om det er et pædagogisk projekt, der skal hjælpe lærerne til at styrke elevernes læring. Det er også vigtigt at se på, hvad de, der deltager i testen, får ud af det.

”Hvis elever og lærere gennemfører en masse test, der ikke giver dem noget i forhold til, hvad der er formålet med at gå i skole og undervise, så synes jeg, at det er spild af tid. Og ikke blot spild af tid. Det giver god mening at færdighedsteste eleverne, hvis det har et fornuftigt pædagogisk og didaktisk formål. Men det at blive testet kan have nega-

## »Når lærerne er bevidste om, at testen også tjener som kontrol, kan undervisningen uforvarende komme til at minde om testforberedelse.«

Tine Nielsen



tive sideeffekter i form af pres, konkurrence, nervøsitet, tab af selvtillid og endda angst hos nogle af eleverne. Disse sideeffekter kan man reducere ved kun at bruge testen som et pædagogisk værktøj,” siger hun.

### Det pædagogiske formål saboteres

Formålet med at lade alle børn i skolen gennemføre en standardiseret færdighedstest bør ifølge Tine Nielsen altså være pædagogisk og didaktisk. Testen skal bidrage til at opfylde folkeskolens formål. Et sådant testværktøj kan man sagtens lave, men fordi man kan opsamle data og bruge dem til alt muligt andet, risikerer formålet at skride.

”Når man følger debatten om og ser på undersøgelser af, hvordan lærere og elever oplever de nationale test, tyder meget på, at det er et helt andet formål, der er kommet i højsædet, nemlig styringsformålet. Når lærerne er bevidste om, at testen også tjener som kontrol, kan undervisningen uforvarende komme til at minde om testforberedelse. Også af godhed over for eleverne, fordi lærerne gerne vil have, at eleverne får en god testoplevelse. På den måde saboterer testen så at sige det pædagogiske formål, der

kunne være med at få et billede af, på hvilke områder eleverne har det svært eller klarer sig godt, og bruge det som pejlemærke i undervisningen,” siger hun og er enig med Christian Ydesen i, at det må være et krav til en god test med et pædagogisk formål, at den opleves som en ’low-stakes’ test af både lærere og elever.

Når den enkelte elevs testresultater ikke blot er et pædagogisk redskab, lærere og elever kan bruge sammen, men indgår som brik i en større evaluering af eleven, lærerne og skolen og i sidste ende hele det danske uddannelsessystem, pålægger man eleverne et unødigt pres, mener Tine Nielsen.

”Vi har jo i forvejen eksamen, prøver og karakterer, som eleverne skal hjem og tale med forældrene om, og der er ingen grund til at lægge nationale test oveni. Hvis man forestillede sig, at eleven fx fik at vide, at den nationale test havde vist, at hun skulle arbejde med de og de områder for at blive bedre, så var det en helt anden historie, hun havde med hjem til forældrene, end en historie om, at deres datter havde scoret ’en del under gennemsnittet’ i læsning,” siger hun.

Et nyt testsystem bør ifølge Tine Nielsen undlade at placere samtlige landets skoleelever i et færdighedshierarki i forhold til landsgennemsnittet.

”Det er ikke væsentligt, om en elev scorer over eller under skolens, kommunens eller landets gennemsnit. Det væsentlige er, at man psykometrisk faktisk kan konstruere en test sådan, at læreren kan få information ud om elevernes styrker, svagheder og udfordringer inden for de forskellige færdighedsområder og opgavetyper, som hun kan bruge aktivt sammen med eleven. Hvis man gerne vil have et redskab til styring og kontrol, må man definere præcist, hvad man gerne vil vide, og så udvikle en metode, der ikke nødvendigvis kræver, at eleverne involveres. Men at kaste så mange kræfter efter at udvikle et så avanceret testsystem som de nationale test er ikke umagen værd, hvis det ikke kan tilføre skolen noget pædagogisk og didaktisk,” siger hun.

### Adaptiv eller ej?

De nationale test er adaptive. Hver opgave, eleven præsenteres for under testen, afhænger af, om hun har svaret rigtigt eller forkert på den foregående opgave. Forenklet kan man sige, at hvis hun har svaret rigtigt, stiger sværhedsgraden, og har hun svaret forkert, falder den.

”Adaptive test er udfordrende for alle – ikke kun for børn. Når jeg underviser psykologistuderende i psykologisk testning, prøver de

en adaptiv test. De er nogle af dem med det højeste karaktergennemsnit i landet, og aligevel trykker testen på selvtilliden hos nogle af dem. For selvom man er god, ja måske netop *fordi* man er god, kan det være udfordrende, at opgaverne er relativt sværere, fordi de er tilpasset niveauet for den enkelte. Dette er man nødt til at kunne tage hånd om både før og under testen, også når det handler om børn,” siger Tine Nielsen.

En af fordelene ved adaptive test er til gengæld, at man behøver færre opgaver end i en lineær test til at opnå den målesikkerhed, der er påkrævet for at kunne vurdere elever individuelt. I det adaptive testdesign belastes eleven derfor mindre og skal heller ikke bruge så meget tid på test som i en lineær test.

Både det adaptive og det lineære testdesign har sine fordele og ulemper, pointerer Tine Nielsen. Om man skal vælge det ene eller det andet eller en kombination i en ny test, må bero på en afvejning af dem alle i forhold til formålet med testen.

”Det vigtigste er i virkeligheden, at beslutningen bliver taget af folk, der har forstand på det, og som kan tænke testteori sammen med folkeskolepraksis. Det kan være en kombination af psykometriske eksperter og kyndige folk med viden om eller erfaring med folkeskolen, som forholder testen til den kontekst, den skal indgå i, og til den viden vi har om elevers og læreres oplevelser med test,” siger hun.

Det adaptive princip kræver rigeligt med opgaver i alle sværhedsgrader og på alle færdighedsniveauer. Men det betyder også, at en masse opgavetyper og -formater går igen i den såkaldte opgavebank. Ifølge Christian Christrup Kjeldsen bør en ny test gå væk fra det adaptive princip, i hvert fald i dets fulde udstrækning.

”Opgaverne er i høj grad variationer over de samme temaer med forskellige sværhedsgrader. Man kan ikke gøre opgaverne vedkommende eller opbygge en sammenhæng mellem dem, når man skal kunne hoppe mellem sværere og nemmere opgaver, indtil man har ramt elevens niveau,” siger han.

De internationale IEA-undersøgelser kan tjene som inspiration for fremtidens test herhjemme, mener Christian Christrup Kjeldsen. Fx den næste PIRLS-undersøgelse i 2021 af skoleelevers læsefærdigheder, der kombinerer det adaptive princip med en lineær test.

”Efter den enkelte elev har skudt sig ind på et bestemt niveau, fortsætter testen på klassisk lineær facon, hvor alle efterfølgende opgaver omtrent ligger på dette niveau,

---

## »Vi lærer i sammenhænge, og vi underviser i sammenhænge – derfor skal vi også teste i sammenhænge.«

Christian Christrup Kjeldsen



uanset hvad eleven svarer. Denne kombination kunne være en god måde at teste på, hvis man ønsker at vurdere den enkelte elevs faglighed bedre, end man kan, hvis alle opgaver er ens for alle. For det kan samtidig give læreren mulighed for efterfølgende at se og vurdere grupper af elevers besvarelser af de samme opgaver i forhold til, om der fx er ’huller’ i curriculum, som skal dækkes,” siger han.

### Test i sammenhæng

De nationale test måler elevernes faglighed løstrevet fra de sammenhænge, de indgår i – og det er en fejl, mener Christian Christrup Kjeldsen.

Ud over de mere klassiske opgaver præsenterede IEA-undersøgelsen TIMSS, der tester elevers færdigheder i matematik og natur/teknologi, i 2019-udgaven eleverne for problemløsningsopgaver udformet som små historier. Hver ny opgave relaterede sig til den sammenhæng, der var opbygget undervejs. Det er en fornuftig måde at bygge en test op på, mener Christian Christrup Kjeldsen.

”Fagligheden er knyttet til undervisningssituationer og andre faktorer. Læreren kender

disse didaktiske sammenhænge, og testen må derfor kunne fungere som tilbagemelding ind i disse. I IEA-undersøgelserne spørger vi, om eleverne oplever, at de er blevet undervist i det fagområde, de testes i, og hvordan. Det giver et yderligere indtryk af sammenhængen mellem faglighed og undervisning,” siger han og fortsætter:

”Vi lærer i sammenhænge, og vi underviser i sammenhænge – derfor skal vi også teste i sammenhænge. I natur/teknologidelen af TIMSS 2019 lavede eleverne små interaktive forsøg. Det er en anderledes og mere avanceret testform. Eleverne syntes, det var spændende, og det ligner i højere grad den undervisningssituation, som elevernes faglighed bør testes i sammenhæng med,” siger han.

Selvom de nationale test afvikles digitalt, er opgaverne udformet sådan, at de lige så godt kunne have været løst som en klassisk papir og blyant-test, mener Christian Christrup Kjeldsen.

”Det er et spørgsmål om at tage online-testen et skridt videre og gøre brug af de særlige muligheder, computeren giver. Når eksperimentet fx er så vigtig en del af fysikundervisningen, bør en test i naturfag så godt det lader sig gøre i et online-univers, inkorporere nogle af de digitale laboratoriefaciliteter, der allerede findes som undervisningsform. Test bør være et billede af den undervisning, der faktisk foregår.” ■



**CHRISTIAN CHRISTRUP KJELDEN** er lektor i pædagogisk sociologi og viceinstituttleder for forskning ved DPU, Aarhus Universitet. Han er leder af Nationalt Center for Skoleforskning og forsker bl.a. i sammenhængen mellem elevers præstationer i skolen og deres sociale baggrund. Han underviser på Kandidatuddannelsen i pædagogisk sociologi.



**CHRISTIAN YDESEN** er professor MSO ved Aalborg Universitet og forsker i bl.a. uddannelseshistorie og brug af test i uddannelsessystemet.



**TINE NIELSEN** er lektor i psykologi ved Københavns Universitet og forsker bl.a. i psykometri, især udvikling og brug af psykologiske og pædagogiske måleskalaer og test relateret til uddannelse.