

# Sprogtest, evalueringer og deres virkninger

Evalueringer og test har alle de forviklinger og kontroverser ved sig, som man kan forvente, når man tilstræber at bruge videnskabelig metode direkte i den politiske, pædagogiske og praktiske virkelighed.<sup>1</sup>

Hvad er konsekvenserne af sprogtest? Hvori består deres betydning, praktisk, organisatorisk, samfundsmæssigt og politisk? Det er enkle spørgsmål at stille. Men de er vanskelige at besvare. Konsekvenserne er formodentlig mangfoldige, og de spreder sig på mange niveauer fra det store samfundsmæssige plan og helt ned til det enkelte individ. Hvilke konsekvenser, man kan få øje på, afhænger som i enhver kompliceret samfundsvidenskabelig problemstilling af de(t) teoretiske perspektiv(er), man lægger ned over sagen.

Den teoretiske tænkning om sprogtest og den teoretiske tænkning om evaluering har i de senere år foretaget en parallel bevægelse (hvad angår test, se Shohamy 2007). På begge områder har den forskningsmæssige interesse udviklet sig fra primært at fokusere på måletekniske forhold til nu at handle om de samfundsmæssige, organisatoriske, sociale, politiske og praktiske omgivelser, som test/evaluering foregår i. Både test og evaluering er produkter af tidsånden og en række samfundsmæssige betingelser, men de virker også tilbage på mennesker og deres vilkår i konkrete situationer. Snarere end blot at spørge, om test og evaluering korrekt måler det, der bør måles (den intrinsiske side), er man i dag således optaget af den ekstrinsiske side, altså de mange måder, som evaluering og test virker på midt i den sociale virkelighed.



**PETER DAHLER-LARSEN**

Ph.d., professor, Institut for Statskundskab,  
Syddansk Universitet  
pdl@sam.sdu.dk

Der er selvfølgelig også forskelle på evaluering og sprogtest. Evaluering plejer at handle om vurderinger af offentlige indsatser og deres virkninger. Sprogtest har oftere fokus på den enkeltes sproglige udvikling, men bruges jo også som virkemidler og indikatorer i sprogpolitikken. Om en sprogtest er en evaluering, er derfor ikke et intrinsisk metodespørgsmål, men et ekstrinsisk praktisk, organisatorisk og politisk spørgsmål.

Artiklen består af to dele med hver sin ansats. I første del spørger jeg: Har der været nogle væsentlige udviklingstræk de senere år i organiseringen af evaluering, og herunder de værdier og tankegange, som er dominerende i tidsånden? I anden del drager jeg nogle konsekvenser for vore begreber om virkningerne af evalueringer og sprogtest. Jeg bruger i det følgende grundskolen og indskolingens som eksempel, men overførbareheden til unge- og voksenområdet ligger lige for.

## Tidstypiske stilarter i test og evaluering

Betragt to idealtyper, dvs. rendyrkede begrebslige måder at se virkeligheden på, som ikke skal forveksles med virkelighedens forvirrende kaos. Disse to idealtyper vil jeg kalder »den reflektive modernitet« og »revisionsfundet«.

I den reflektive modernitet har man nået den erkendelse, at samfundets fremskridt ikke længere har automatisk opbakning. Som Ulrich Beck (1992) er inde på, er der måske endda for meget fremskridt snarere end for lidt; for fremskridtet skaber nemlig en række bivirkninger. Heldigvis har man i reflektiv modernitet en række refleksionsværktøjer. Løbende evaluering er et af de vigtigste. På det pædagogiske område kender vi løbende evaluering som led i »at møde eleven, der hvor eleven er« såvel som i elevens »ansvar for egen læring«.

Refleksiv modernitet er også karakteriseret ved, at ingen længe tror på, at fremskridtet fører til samme verdensopfattelse hos alle. Det gælder på det store globale plan, hvor kultur- og troforskelle ikke forsvinder, blot fordi der kommer mere og mere modernitet til alle på jordkloden. Og det gælder på det lille lokale plan, hvor det bliver åbenlyst, at der er fundamentalt forskellige perspektiver på en offentlig indsats, alt efter om man er mand eller kvinde, læge eller patient, dansker eller nydansker, lærer eller elev osv. Ingen tror på, at et fremskridt vil få alle opfattelser til at konvergere i en stor fælles og fornuftspræget opfattelse. I stedet lyder dagens bud, at man må skabe en udvikling ved at bringe alle de forskellige opfattelser i spil med hinanden på en forhåbentlig konstruktiv måde.

Igen må man lytte og reflektere. Evaluering med inddragelse af brugere og andre interessenter har derfor været tidens løsen under refleksiv modernitet. Men den slags evaluering har ikke været uden problemer. Evalueringerne har været subjektive og derfor uforudsigelige og tillige usammenlignelige og vanskelige at sammenfatte. Ikke mindst fra et ledelses- og styringsperspektiv har man derfor haft et behov for at rydde op og skabe overblik.

Det overblik er kommet med det såkaldte »revisionssamfund« (eller The Audit Society (Power 1997)). Her er evaluering og test løbende, standardiserede, obligatoriske og sammenlignelige. Nu skal alt evalueres, og det systematisk. Alle skoler skal indsende kvalitetsrapporter. Alle elever skal have handleplaner. Og alle skal sprogtestes. Jo tidligere, jo bedre. I revisionssamfundet vil man foregribende håndtere de samfundsmæssige problemer ved hjælp af indikatorer og indrapporteringer, inden de viser sig som en risiko for beslutningstagerne.

I mange tilfælde betyder det nok, at lidet ophidsende og ikke-handlingskrævende forhold også dokumenteres grundigt. En hel del evalueringer og test bliver rutineprægede. Man kan kalde dem evalueringsmaskiner. De er baserede på indikatorer, manualer, håndbøger, skemaer, test og bureaukratisk afrapportering, således at ethvert tilsyneladende subjektivt element er elimineret. Derfor kører evalueringsmaskinerne relativt uafhængige af, hvem der er operatører i det daglige.

Mange af evalueringsmaskinerne fører ikke til kvalitetsforbedringer. Men de fastholder et afrapporteringsansvar. Det er klart og indlysende, hvem der har et ansvar for hvilke scorer på hvilke test, og hvem der har et ansvar for indrapporteringen. Den samfundsmæssige nøgle, der bedst forklarer et sådant evalueringsregime, har ikke at gøre med refleksivitet, udvikling og læring. Nøglen ligger i stedet i risikominimering og fastholdelse af ansvar gennem afrapportering. Allerede inden en ubehagelig afvigelse fra en opstillet testnorm sker, er den foregrebet gennem obligatoriske mekanismer, der mere ligner forsikring end kvalitetsudvikling. Ønsket om at eliminere vilkårlig og uforudsigelig evaluering spiller sammen med et ledelsesmæssigt ønske om forudsigelighed og overblik. Mediernes fokus på enkeltstager styrker kun politikernes behov for at sige, at der er installeret et overvågningssystem, så der er styr på sagerne. Over det hele svæver nogle aktuelle kulturelle værdier, der siden 2001 har betonet elimineringen af sjældne, men forfærdelige begivenheder. Overvågningssystemer både i direkte og i overført betydning skal derfor sikre samfundet mod ubehagelige overraskelser.

Sprogtest kan selvfølgelig være relevante i både refleksiv modernitet og i revisionsfundet. I den refleksive modernitet kan de bruges til at justere pædagogisk praksis ind på situationelle forhold og det enkelte barns kognitive og sproglige udvikling. Hovedvægten lægges dermed på læring og udvikling. Men jo mindre systematik der er i anvendelsen af test, jo mere sporadisk vil anvendelsen være.

I revisionsfundet lægges hovedvægten derimod på den systematiske kontrolfunktion. Opsynet er obligatorisk og vedvarende. Efterhånden som børn, pårørende, pædagoger og ledere indretter sig på testens sociale virkelighed, og brugen af test institutionaliseres, kan man forestille sig en række videre samfundsmæssige konsekvenser, som dårligt lader sig beskrive ved de kendte begreber »kontrol« versus »læring«.

## Anvendelser af sprogtest

Test gør andet og mere, end de siger, de gør. Det ser ud som om, der er styr på ting, når de måles. I en kultur, der lægger vægt på rationalitet, objektivitet og målbarhed, ser beslutningsprocesser fornuftige ud, hvis der lægges data til grund for beslutningen (Feldman og March 1981). Moderne organisationer bruger derfor styringskoncepter, der baserer sig på indgående afrapportering af store mængder data, også selv om de mange data ikke altid relaterer sig til de beslutninger, der tages. Mere end én politiker giver udtryk for, at hvis der er obligatoriske test, så er der styr på tingene. Ideen om, at en dokumenteret praksis også er en god praksis, er nærmest selvindlysende, i hvert fald i et afrapporteringsøjemed. Måske er formålet med evalueringer og test egentlig ikke at skabe nogen udvikling, men at udbrede en følelse af tillid til, at nogen har det store overblik, og at der gøres noget (Power 2004).

Imidlertid vil en rent symbolsk anvendelse af test og evaluering i længden ikke blive ved med at være rent symbolsk, lyder en indvending. Gradvist vil evaluering og test, specielt når de gentages, også begynde at præge virkeligheden på mere substantielle måder.

Sprogtest kan være instrumenter til betragtelig samfundsmæssig magtudøvelse. Sprogtest er i hænderne på samfundsmæssige autoriteter. Sprogtest virker med deres tekniske udtryk og kvantitative resultater som objektive størrelser, der er lige så sande som en temperaturmåling. Sprogtest knytter sig til det enkelte barn og kan dermed antages at sige noget væsentligt om ikke blot barnets sproglige udvikling, men i det hele taget om »hvor barnet står«.

En række professionelle og andre omkring barnet er parate til at handle og dermed øge de virkelighedsskabende effekter af sprogtest. Vor tids behandlingsideologi tilsiger, at jo tidligere barnet undersøges, jo bedre chancer har de professionelle for at gribe ind i barnets udvikling. På det store samfundsmæssige plan har sprogtest gode chancer for at gøre sig gældende som politiske instrumenter, fordi ideer om sprog og etnicitet har politisk vind i sejlene for tiden. Vi lever i et politisk og kulturelt regime, hvor vi har svært ved at håndtere balancen mellem sprog som et praktisk redskab og sprog som en kulturel markør, og det påvirker selvfølgelig hele det rum, hvor konsekvenser af evalueringer og sprogtest drages.

## Konstitutive virkninger

Til at betegne de mange uformelle måder, hvorpå evaluering og test griber ind i virkeligheden, har jeg introduceret betegnelsen »konstitutive virkninger« (Dahler-Larsen 2008). I begrebet ligger, at mennesker reagerer på at blive målt, og når reaktioner på reaktionerne tælles med, kan evaluering og test være med til at forme en ny virkelighed, der rækker videre, end hvad enkelte aktører har tænkt eller planlagt. De konstitutive virkninger folder sig ud på mange måder. I grundlæggende forstand er sprogtest med til at definere, hvad der er værd at tælle med som sprog. Et eller flere sprog fremhæves som værd at kunne, mens andre sprog ignoreres. En sprogtest kan desuden hive de sproglige kompetencer ud af den kontekst, hvor de bruges dagligt. Det kan medføre, at en sprogbrug, som er pragmatisk effektiv i en dagligdags situation, men som måske formelt og grammatisk er utilstrækkelig, vurderes negativt. Nogle gruppers sprogbrug anses for mere korrekt end andres. Dertil kommer, at sprogtesten fastholder og cementerer en bestemt forståelse af sproglig udvikling. I mange tilfælde udtrykkes resultatet af en sprogtest endimensionelt på en skala, mens nogle sprogforskere faktisk hævder, at sproglig udvikling rummer flere dimensioner og kommer til udtryk i flere forskellige domæner. En grundlæggende funktion af sprogtest er således, at de objektiverer og gyldiggør bestemte forståelser af, hvilke sprog der tæller, og hvordan sproglig udvikling skal måles og derfor også, hvordan den skal forstås.

Blandt de konstitutive virkninger af sprogtest hører også reaktionerne på testresultater. Det er en kendt sag fra eksamen og test i skolen generelt, at der kan være en række konstitutive virkninger i form af »teaching to the test«, altså at undervisningen gradvist indretter sig på, at der helst skal scores højt på testen. Nogle af

de pædagogiske adfærdsformer, lærere anvender i den forbindelse, kan have ganske negative virkninger på den pædagogiske praksis. Der kommer fokus på den type af viden, som kan gengives i test, og en del af undervisningen kommer til at handle om, hvordan man består test. Eksempelvis kan test, der alene handler om ordforrådets størrelse, have som effekt, at pædagoger entydigt retter opmærksomheden på at øge børnenes kendskab til nogle af de ord, der kan forekomme i testen. En pædagogisk teori om, at selve ordforrådets størrelse udtrykker sproglig udvikling som sådan, kan blive mindre troværdig som følge af en sådan adfærd.

Man taler inden for den pædagogiske forskning om 'washback-effects', dvs., at et bestemt koncept for evaluering eller test indvirker på den pædagogiske praksis, på undervisningens organisering og på læreplaner. Testbarhed bliver et kriterium ikke bare for stofets udvælgelse og organisering, men for hele rettetheden i den pædagogiske institution. På den måde bliver test ikke en tilbageskuende måling af udført arbejde, men et fremadrettet organisationsprincip, der præger det pædagogiske felt som sådan.

Blandt konstitutive virkninger tæller også, hvordan testresultater er med til at definere børnene. Ligesom diagnoser hænger fast på patienter, kan testresultater hænge fast på børn. Man glemmer måske tvivlen om testens pålidelighed og læser i stedet resultaterne som udtryk for iboende egenskaber ved det enkelte barn selv. Når man efterfølgende handler pædagogisk på sådanne resultater, kan der opstå en selvopfyldende praksis. Der findes i hvert fald pædagogisk forskning, der viser, at lærere lader sig påvirke af de forhåndsforventninger, de har til det enkelte barn (Rosenthal-effekten). Det kan være svært at slippe af med en etiket, når den først er klistret på. Den mekanisme er vi vant til, når det gælder større børn og patienter. Men vi har måske ikke tænkt igennem på samfundsplan, hvorvidt vi også ønsker, at sådanne mekanismer skal gælde meget små børn, der sprogtestes tidligt.

Normalitet bestemmes af socialt konstruerede forventninger til resultaterne. Et barn, som i en given aldersgruppe scorer lavere end de andre i gruppen, vil rent statistisk kunne rubriceres som ikke-normalt. Statistikken vil med andre ord definere nogen som »ikke-alderssvarende«. I vor tid synes der at være en række argumenter for, at de ikke-alderssvarende børn skal identificeres tidligt. Argumenterne for at vente med at identificere disse børn synes at stå svagere i billedet. Hvorom alting er, vil testresultater være med til at definere det pædagogisk gældende billede af det enkelte barn. Også barnet selv og forældrene vil med tiden på godt og ondt kunne tilegne sig dette billede af barnet og handle på det.

Blandt konstitutive virkninger tæller også, hvordan relationen mellem barn og voksen ændres, når testen kaster sit lys på barnet. En relation, der måske ellers fremtræder som mere umiddelbar, og hvor barnets styrker og svagheder måske fornemmes intuitivt eller slet ikke, vil nu blive omdefinert i lyset af den tilsyneladende skarpe og objektive viden, som testen har givet, og som ikke kan ruller tilbage. Det gælder ikke mindst, hvis de voksne (pædagoger eller forældre) tilordnes eller tiltager sig et ansvar for at følge op på resultaterne i forhold til næste gang, der skal måles.

I et større perspektiv kan sprogtest også kaste skygger på en klasse, en etnisk gruppe, et lokalområde eller et segment af befolkningen. Selvfølgelig kan sprogtesten også kaste lys på problemer, som ellers ville være skjulte, men som man nu er nødt til at gøre noget ved.

Muligheden for konstitutive virkninger af sprogtest leder frem til et begreb om det såkaldte »performance paradox«, nemlig en situation, hvor mere måling ikke fører til mere udvikling og kvalitet. Vi kan sprogteste mere uden at opnå en ønsket sprogdudvikling. Tolkningen af et performansparadoks kræver en del omtanke. Mange taler om negative og utilsigtede virkninger af test og evaluering. Tager vi begreberne seriøst, er der imidlertid ikke enighed om, hvad der er »negativt« og »utilsigtet«. Hvis f.eks. test og evalueringer fører til øget social segregering (f.eks. fordi velstillede flygter fra områder med dårlige test- og evalueringresultater), vil det i nogens øjne være en indlysende negativ og måske utilsigtet effekt. Hvis jeg er velstillet i et velstillet område, er øget social segregering imidlertid hverken nødvendigvis negativ eller utilsigtet. For mig personligt kan det være en god ting, hvis mine velstillede børn må tilbringe mere tid sammen med andre mere velstillede børn.

Dette argument afprøves her naturligvis alene for argumentets skyld. Det afgørende er, at hvis der findes forskellige synspunkter i samfundet, så vil test og evalueringer have en tilbøjelighed til at understøtte nogle af disse frem for andre. Og de konstitutive virkninger af evalueringer og test vil være mere velkomne hos nogle end hos andre.

## Afrunding

Det er analytisk utilstrækkeligt blot at hævde, at virkningerne af test er negative eller utilsigtede. Et mere analytisk holdbart syn er, at sprogtest blot er en avanceret måde at føre sprogpolitik på. Ny virkelighed defineres. Med begrebet konstitutiv evalueringssanven-

delse kan vi se, at evaluering og sprogttest er med til at skabe en ny virkelighed, uden at vi kræver, det foregår hverken på en tilsigtet eller en utilsigtet måde. Men det er ikke en virkelighed, som er neutral over for forskellige interesser og problemforståelser i samfundet. Sprogttest er sprogpolitik klædt som undersøgelse.

## Note

<sup>1</sup> Denne artikel er skrevet i december 2010.

## Litteratur

- Beck, U. (1992): *Risk society – Towards a New Modernity*. London: Sage.
- Dahler-Larsen, P. (2008): *Kvalitetens Beskaffenhed*. Odense: Syddansk Universitetsforlag.
- Feldman, M. and March, J. (1981): »Information in Organizations as Signal and Symbol«. I *Administrative Science Quarterly*, 26: 171-184.
- Power, M. (1997): *The Audit Society*. Oxford: Oxford University Press.
- Power, M. (2004): *The Risk Management of Everything*. London: Demos.
- Shohamy, E. (2007): »Language Tests as Language Policy Tools«. I *Assessment in Education*, 14(1): 117-130.