

Jesper Wittrup,

Programleder, KREVI (Det Kommunale og Regionale Evalueringsinstitut)

KREVI's analyser af folkeskolernes faglige kvalitet og effektivitet:

# En talknuser møder folkeskolen



**KREVI har i foråret/sommeren 2011 publiceret to analyser, der måler folkeskolernes faglige kvalitet samt foretager vurderinger af effektiviseringspotentialet på området. Artiklen fokuserer på de bagvedliggende overvejelser bag undersøgelserne, og deres forhold til den generelt øgede brug af kvantitative metoder til evaluering i den offentlige sektor. Artiklen argumenterer for, at sådanne analyser vil vinde mere indpas, men pointerer også, at de, der udfører analyserne, har et betydeligt ansvar i forhold til den måde, hvorpå tallene bliver præsenteret og brugt. I forhold til de konkrete analyser af folkeskolen er der særligt lagt vægt på at understøtte skolernes læring og erfaringsudveksling.**

### **KREVI-analyser af folkeskolen**

Ifølge det "rejsehold", der i 2010 foretog en analyse af den danske folkeskoles tilstand og udfordringer (Skolens rejsehold, 2010), lider skolerne generelt af et manglende fokus på resultater, en svag evalueringkultur og en dårlig evne til at lære af hinandens gode erfaringer. Bl.a. med udgangspunkt heri har Det Kommunale og Regionale Evalueringsinstitut (KREVI) besluttet at understøtte skolernes muligheder for at sammenligne og lære af hinandens resultater ved hjælp af tre nye tiltag.

For det første har KREVI offentliggjort en ny og bedre måling af folkeskolernes faglige resultater (KREVI, 2011a). For det andet er disse resultater blevet koblet med oplysninger om skolernes ressourceforbrug med henblik på at afdække mulighederne for gennem udveksling af erfaringerne at forbedre skolernes effektivitet (KREVI, 2011b). For det tredje lanceres sidst på efteråret 2011 en software-løsning, der vil kunne hjælpe skolerne med Interaktiv Benchmarking.

KREVI afdækker i sin nye rapport om skolernes faglige kvalitet de enkelte danske folkeskolelæreres selvstændige indflydelse på, hvor gode eksamens karakterer deres afgangselever får, hvilket kan tages som et udtryk for skolernes faglige kvalitet. Det er set før, men den metode, som er anvendt her, er mere avanceret end dem, der tidligere er anvendt i Danmark. Desuden giver målingen et indblik i skolernes evne til også at samle de socialt svageste elever op.

Undersøgelsen fokuserer på eksamensresultater, men den målte faglige kvalitet er sammenholdt med målinger af elevernes trivsel. Undersøgelsen viser, at der er en positiv sammenhæng mellem skolernes faglige kvalitet og elevernes trivsel.

Undersøgelsen kan også bruges til at advare mod forsimplede ranglister over skolernes kvalitet. Med multilevel-modellering får vi et mere realistisk billede (sammenlignet med tidligere gennemførte analyser) af den usikkerhed, der naturligt knytter sig til måling af skolernes faglige kvalitet. Usikkerheden er så betydelig, at det ikke giver megen mening at operere med f.eks. en top ti-liste over de "bedste" folkeskoler i landet. Vi kan nemlig ikke med stor sikkerhed afgøre, om den faglige kvalitet på top ti-skolerne er bedre end på skolerne på pladserne 11 til 40. På den anden side viser analysen, at nogle grupper af skoler med stor sikkerhed gør det væsentlig bedre end andre – også inden for samme kommune.

I KREVI's rapport om effektivisering er skolernes faglige kvalitet koblet med oplysninger om skolernes ressourceforbrug. Ved hjælp af "Data Envelopment Analysis" (DEA) identificeres "forbilleder" for den enkelte skole, det vil sige andre skoler, der er karakteriseret ved enten at opnå væsentligt bedre faglige resultater uden et højere ressourceforbrug, eller ved at have et væsentligt lavere ressourceforbrug uden at opnå dårligere faglige resultater. En mere detaljeret teknisk gennemgang af metoden findes hos Bogetoft og Wittrup (2011).

Analysen viser, at folkeskolerne kan spare 13 % i snit uden at ødelægge fagligheden. Der er store lokale forskelle bag gennemsnittet: Fra under 2 % i Ikast-Brande til over 30 % i Høje-Tåstrup. Hvis man vil effektivisere ved at øge fagligheden frem for at svinge sparekniven, er det muligt på de fleste skoler: Gennemsnitligt kan karaktererne for afgangseleverne i 9. klasse øges med mellem 5 % og 20 %. Potentialet er størst i Faxe og Gribskov; over 18 % i gennemsnit. Det er mindst i Hørsholm og Holstebro; under 3,5 %.

KREVI's analyse kaster også lys over, hvilke effektiviseringsstrategier der vil være mest oplagte i de forskellige kommuner. Nogle kommuner kan med fordel overveje at øge skolestørrelsen, mens andre i højere grad bør overveje at øge lærernes undervisningsandel. Endelig er der kommuner, som skal skruer på helt andre håndtag end skolestørrelse og undervisningsandel.

### Talknuserne i evaluering

Mange i "evalueringsbranchen" har en vis skepsis over for rent talbaserede analyser. Også på CEPRA-stribens netside kan man læse, at redaktørerne ser tidsskriftet som "en platform for seriøse debatter om evaluering, politik, forskning og erfaringer, hvor evaluering og dokumentation er mere end det, der kan måles og vejes".

Mens en sådan skepsis ofte kan være velbegrundet, så er der næppe tvivl om, at tendensen er, at vi netop får flere og flere helt overvejende kvantitative analyser. Ian Ayres, professor i økonomi og jura på Yale, har i sin bog "Super Crunchers: Why Thinking-by-Numbers Is the New Way to Be Smart" (Ayres, 2007) leveret et dybfølt forsvar for Excel-regnearkets indtog på snart sagt alle områder. Ifølge Ayres befinder vi os i et hi-

storisk "hest vs. lokomotiv-kapløb", hvor intuition og traditionel professionel ekspertise uværgeligt taber til talknuserne.

Ayres præsenterer en perlerække af eksempler, hvor statistiske modeller har vist sig overlegne i forhold til professionelle vurderinger. En hovedrolle tildeles økonomen Orley Ashenfelter, ophavsmand til vintidsskriftet "Liquid Assets", der ved brug af data om vejrforhold m.v. har vist sig langt bedre end alskens traditionelle vineksperter til at forudsige den endelige pris (og dermed underforstået kvalitet) på vine. Næsten helt parallelle historier fortælles om andre moderne "helte", herunder idrætstrænere, advokater, læger og minsandten også skoleledere, der har frembragt imponerende præstationer og overgået eller udmanøvreret mere traditionelt tænkende kolleger, konkurrenter og eksperter ved at udnytte statistisk information.

Denne indtrængning af talknuser har typisk mødt heftig modstand fra etablerede eksperter og professioner. Vinskribenten Robert Parker har på et tidspunkt beskrevet Ashenfelter som "a total sham". Ayres' pointe er imidlertid, at talknuserne vinder igen og igen. Traditionelle eksperter har en overdreven tillid til deres egen intuition, og derfor vil de blive trængt



af den talbaserede tilgang. Ashenfelter forudså med sine data, at en årgang 1986 Bordeaux ville blive ordinær, mens Parker spåede den til at blive exceptionelt fremragende. Ashenfelter fik ret.

Ayres' optimistiske fremstilling vil glæde enhver talknuser, men der er nok grund til at spise brød til. Det er også relativt nemt at finde eksempler på brug af dårlige modeller, baseret på utilstrækkelige data. Dertil kommer den omfattende forskning, der har afdækket utilsigtede (mest negative) konsekvenser af talbaseret styring i den offentlige sektor – se f.eks. Hood m.fl. (2006). Her er problemet oftest, at ufuldstændige modeller bruges uden hensyntagen til de incitamenter, de afføder, som når målingen fører til negligering af aspekter, der er vigtige, men svære at måle.

Som følge af eksempler på dårlige erfaringer samt de potentielle negative virkninger af brug af resultatmåling kan man lege med tanken om at rulle tiden tilbage og fjerne alle former for målinger og resultatkrav og på den måde "frigøre" offentligt ansatte fra den slags bindinger. Selv om denne tanke af og til luftes af nogle af de fagforeninger, der organiserer dem, der måles, er det en utopi. Som en OECD-rapport om emnet har formuleret det:

„ This is a superficially attractive prospect, but an unrealistic and undesirable one. The increases in accountability and transparency brought about by the last twenty years of performance measurement have been valuable. Information is now available that cannot and must not be suppressed. Open government demands that people have the right to know how well their services are being delivered, and professionals and managers need to be held to account. The aim must be to build on these developments, while reducing any negative effects... ” (OECD, 2009: 43)

Vi kan ikke rulle tiden tilbage. I et moderne samfund skjuler man ikke information om offentlig service for folk, men man kan gøre en indsats for at ruste dem bedre til at forholde sig til og bruge denne information. Udfordringen for talknusere bliver at konstruere bedre modeller samt at levere nogle anvisninger på, hvordan brugerne bør forholde sig til målingerne. Tallene bør i almindelighed ses som ét blandt flere input, som man bør forholde sig kritisk til, men som kan anvendes til inspiration og læring. Det er i det lys, vi i KREVI gerne vil se de seneste initiativer på folkeskoleområdet.

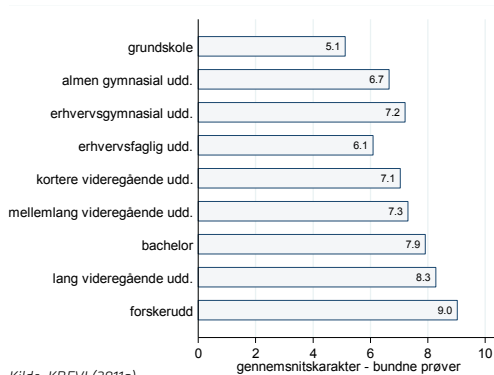
### Model der tager højde for elevernes sociale baggrund

En vigtig målsætning for KREVI's arbejde med analyse af faglig kvalitet og effektivitet i folkeskolen har været at producere bedre og mere relevante målinger end dem, der allerede forelå. Undervisningsministeriet har gennem flere år offentliggjort skolernes gennemsnitskarakterer ved afgangseksamen i 9. klasse, og disse gennemsnit er ofte i medierne blevet brugt til at udstille "gode" og "dårlige" skoler. Problemet med de rå gennemsnit er, at de ikke nødvendigvis siger så meget om skolernes præstationer.<sup>1</sup>

Elevernes karakterer afspejler nemlig i høj grad deres sociale baggrund. Nedenstående diagram viser gennemsnitskaraktererne for elever ved afgangseksamen i 9. klasse i 2008 fordelt ud fra deres mødres uddannelsesbaggrund. En elev, hvis mor kun har en grundskoleuddannelse, opnår i gennemsnit et resultat, der er omtrent 3 karakterer dårligere på 12-trin-skalaen end en elev, hvis mor har en længerevarende uddannelse. Tilsvarende klare sammenhænge mellem social baggrund og eksamens karakterer kan identificeres, når man ser på forældrenes indkomst, jobstatus, etniske baggrund m.v.

<sup>1</sup> Efter at KREVI offentliggjorde sin måling af skolernes faglige kvalitet, har Undervisningsministeriet besluttet at udarbejde sin egen "rangliste" over skolerens faglige resultater, korrigeret for sociale baggrunds faktorer. Undervisningsministeriets måling diskuteres ikke nærmere i denne artikel, men KREVI har udarbejdet et særskilt notat, der sammenligner de metoder, KREVI og Undervisningsministeriet har anvendt (KREVI, 2011c)

## Eksamenskarakterer i 2008 fordelt efter mødres uddannelsesniveau



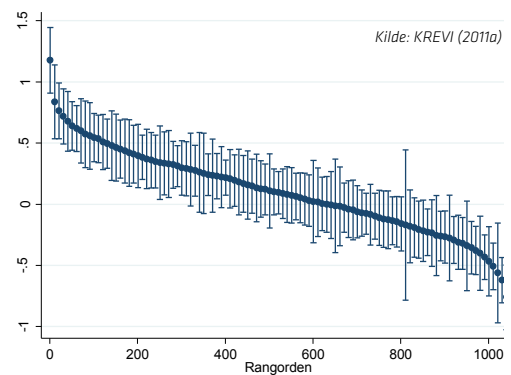
Samlet set kan KREVI's model, som er baseret på individdata fra Danmarks Statistik med diverse oplysninger om elevernes forældre, forklare hele 28 pct. af den variation, der er i eksamenskaraktererne. På den baggrund forekommer det ikke rimeligt at vurdere skolernes præstationer ud fra de rå karaktergennemsnit. En skole med et "lavt" karaktergennemsnit kan sagtens præstere bedre, end man skulle forvente ud fra elevernes sociale baggrund, og omvendt.

### Skepsis over for ranglister

Den liberale tænketank CEPOS har gennem en årrække produceret en rangliste over skolerne, der søger at tage højde for social baggrund. CEPOS har udført et banebrydende arbejde, men deres metode har ikke helt fulgt med den udvikling, der internationalt har været på området. Desuden kan der være grund til at have reservationer over for selve ideen med at præsentere resultaterne på en rangliste.

En betydelig fordel ved analysens anvendelse af multilevel-modeller til estimering af skolernes undervisningseffekt er, at vi får et mere realistisk billede af den usikkerhed, der knytter sig til estimatet på den enkelte skole. Figuren nedenfor viser et antal folkeskoler sorteret efter undervisningseffekt og med angivelse af konfidensintervaller. For overskuelighedens skyld er kun hver tiende folkeskole medtaget i figuren.

## Eksamenskarakterer i 2008 fordelt efter mødres uddannelsesniveau



Konfidensintervallerne er konstrueret således, at vi for to skoler, der ikke har overlappende intervaller, med omtrentlig 95 % sikkerhed kan konkludere, at deres reelle undervisningseffekt er forskellig.

Det fremgår af figuren, at vi ikke kan være helt sikre på, at den ud fra målingen bedst rangerende folkeskole rent faktisk har en bedre undervisningseffekt end f.eks. skolen, der rangerer som nummer 40. Dog er det meget usandsynligt, at skole nr. 1 samtidig også har en dårligere undervisningseffekt end alle de mellem-liggende skoler. Derfor er dens reelle placering med stor sandsynlighed bedre end nummer 40.

Eksemplet illustrer mere generelt, at det ikke giver megen mening at præsentere f.eks. en top ti-liste over folkeskoler med højest faglig kvalitet. Dertil er usikkerheden på målingen for stor.

### Fortsat forbedring mulig

I enkelte medier har der, særligt på baggrund af Undervisningsministeriets senere rangliste over skolerne, været fremsat kritik af den grundlæggende metode, som både Undervisningsministeriet og KREVI har benyttet sig af. Kort fortalt går metoden ud på at sammenligne de forudsagte (på baggrund af socio-økonomiske data) og de faktiske eksamensresultater. På baggrund af disse forskelle (residualer) beregnes

undervisningseffekten ud fra, hvordan eleverne på en given skole præsterer i forhold til det forventede. Kritikken af denne internationalt set meget udbredte metode har primært gået på, at man potentielt, ved inddragelse af andre relevante (men evt. umålelige) faktorer, kunne have fået anderledes residualer.

Det skal for det første bemærkes, at KREVI's analyse har benyttet sig af konservative beregningsmetoder, der ikke tilskriver skolen hele forskellen mellem de faktiske og de forventede eksamensresultater.

Dernæst er det væsentligt at hæfte sig ved modellens meget betydelige forklaringskraft. Mere end en fjerdedel af variationen i de eksamenskarakterer, skoleeleverne opnår, kan som nævnt forklares ud fra forskelle med hensyn til forældrenes uddannelse, indkomst og jobstatus mv. Hvis vi ønsker at have et nogenlunde sammenligneligt billede af skolernes faglige resultater, er det helt essentielt at korrigere for disse forhold, frem for blot at sammenligne "rå" karakterer.

Dette udelukker naturligvis ikke, at andre faktorer, der reelt er uden for skolernes kontrol, og som ikke er afspejlet i de anvendte variable, i et vist omfang kan bidrage til at forklare forskelle i de resultater, skolerne opnår. Eksempelvis har forskere fra AKF m.fl. netop påpeget, at den normering, eleverne oplevede, da de gik i børnehaver, kan have en vis (dog meget begrænset) indflydelse på de eksamenskarakterer, de opnår i 9. klasse (Bauchmüller m.fl., 2011). Det er således muligt, at en skole med en positiv undervisningseffekt til dels opnår denne, fordi den ligger i et område med gode børnehaver.

Da den centrale målsætning med analysen er læring og erfaringsudveksling, ser vi yderligere afdækning af forhold, der kan bidrage til at forklare forskelle i undervisningseffekt, som en naturlig del af den fremadrettede proces. KREVI opfordrer til, at skoler med et fagligt forbedringspotentiale sætter sig i kontakt med identificerede forbilleder, der har en beregnet højere faglig kvalitet. I mange tilfælde tror vi, det vil vise sig, at den skole, der er identificeret som et forbillede, har nogle værdifulde erfaringer, den kan dele ud af. Skulle det imidlertid ved nærmere eftersyn vise sig,

at der er andre forhold, som ikke kan tilskrives skolen, der kan forklare de gode resultater, så er det også en værdifuld information, der evt. kan bruges til at forbedre fremtidige modeller.

### Fokus på læringsperspektivet

Selv om det som følge af usikkerheden på målingen er problematisk at bruge den til opstilling af ranglister, så kan vi sagtens finde mange par af skoler, hvor vi med stor sikkerhed kan sige, at den ene skole fagligt har præsteret bedre end den anden. Det betyder også, at målingen sagtens kan danne udgangspunkt for erfaringsudveksling.

KREVI har netop ønsket at bruge beregningerne som udgangspunkt for erfaringsudveksling mellem skolerne, således at der for den enkelte skole identificeres relevante forbilleder, der ligner den pågældende skole på en række centrale punkter, men som med ganske stor sikkerhed gør det bedre.

Til det formål er der udviklet et softwareprogram, som kommuner og skoler vil kunne anvende interaktivt, og som vil blive lanceret sidst på efteråret 2011. En væsentlig fordel ved dette program, sammenlignet med den generelle måling, er, at kommuner og skoler selv har mulighed for at afprøve alternative modelvarianter, hvilket vil sige, at de selv definerer, efter hvilke kriterier de sammenlignes, og at det selv kan se, hvilke "forbilleder" der identificeres for en given skole ved forskellige modelforudsætninger. Med udgangspunkt i lokale præferencer, for hvordan en god skole kan se ud, kan der være god fornuft i at afprøve alternative modelvarianter i forsøget på at målrette benchmarkingen mod skoler, der lokalt vil blive opfattet som egnede forbilleder.

### Konklusion

Der er næppe tvivl om, at der fremover vil komme mere detaljerede data om folkeskolernes resultater. Hvis den øgede datamængde skal udnyttes konstruktivt, er der behov for stærke redskaber, der kan håndtere mange resultatdimensioner på én gang, og som kan understøtte erfaringsudveksling.



Hvis folkeskolen virkelig skal være blandt verdens bedste er der – med rejseholdets ord – ”beslutninger, der skal træffes. Penge, der skal omdirigeres. Funktioner og kompetencer og institutioner, der skal skæres, flyttes eller opbygges” (Skolens rejsehold, 2010: 8). KREVI's initiativer vedrørende Benchmarking og DEA sigter mod at hjælpe med til, at det er de rigtige beslutninger og tiltag, der vælges.

## Litteratur

Ayres, Ian (2007): *Super Crunchers: Why Thinking-by-Numbers Is the New Way to Be Smart*. New York, NY: Bantam.

Bauchmüller, Robert; Mette Gørtz og Astrid Würtz Rasmussen (2011): "Long-Run Benefits from Universal High-Quality Pre-Schooling". In: *AKF Working paper*, 2011(2), AKF.

Bevan, Gwyn & Christopher Hood (2006). "What's measured is what matters: Targets and Gaming in the English Public Health Care System". In: *Public Administration*, Volume 84, No. 3.

Bogetoft, Peter & Jesper Wittrup (2011): "Productivity and education. Benchmarking of elementary and lower secondary schools in Denmark". In: *Nordic Economic Policy Review*, No. 2, 2011.

KREVI (2011a): *Folkeskolens faglige kvalitet: analyse af folkeskolernes undervisningseffekt*. [www.krevi.dk](http://www.krevi.dk).

KREVI (2011b): *Effektivisering i folkeskolen: Muligheder og metoder*. [www.krevi.dk](http://www.krevi.dk).

KREVI (2011c): *Faglig skolekvalitet ifølge KREVI og Undervisningsministeriet*. [www.krevi.dk](http://www.krevi.dk).

OECD (2009): *Measuring Government Activity*.

Skolens rejsehold (2010): *Fremtidens folkeskole: En af verdens bedste – Anbefalingsrapport*. Udgivet af Styrelsen for Evaluering og Kvalitetsudvikling af Folkeskolen.

