

Fra KUP til kpi

- tilsyn med uddannelseskvalitet som styrings- og ledelsesproblem

Af: Peter Henrik Raae og Karl-Henrik Jørgensen

Artiklen analyserer den udvikling, der har fundet sted i det ministerielle tilsyn med gymnasierne siden 1990 med vægten lagt på de sidste to årtier og den virkning, det kan have på sektoren at tilsynet her er skiftet fra et tilsyn som en form for fagfællebaseret evaluering af praksis på udvalgte gymnasier til en praksis, hvor samtlige landets gymnasier screenes på baggrund af et antal indikatorer, og hvor kun de dårligst præsterende udtages til nærmere tilsyn. På baggrund af interview med rektorer angående deres erfaring hermed og informeret af evalueringforskning analyseres og diskuteres udviklingen: Har tilsynet den ønskede effekt eller risikerede det at virke kontraproduktivt? Hvordan positionerer tilsynet ledelse og med hvilke risici?

Den frisætning til selvstyre, som også blev de gymnasiale uddannelsers virkelighed det første årti efter årtusindskiftet, rejser spørgsmål om politisk kontrol på nye måder.

Tidligere undervisningsminister Bertel Haarder redegjorde allerede i 1988 i et selvforfattet nummer af det daværende undervisningsministerielle tidsskrift *Uddannelse* for den politiske vision, han og regeringen arbejdede for (Haarder, 1988). Her betonedes frisættningens muligheder. Det var en visionsbeskrivelse, der hvilede på en antagelse om, at institutioners øgede selvstyre kombineret med konkurrenceudsættelse ville fremme kvalitet. Det er dog ikke et synspunkt, der efterfølgende fik lov at stå uantastet. Senere regeringsinitiativer trak dele af ministeriets afgivne styringskapacitet tilbage, dog ikke som simpel tilbagerulning til den planstyring, der var kendt fra tidligere epokers opbygning af velfærdssamfundet. Nu indførtes i stedet lovbestemte krav om kvalitetssikringssystemer, præstationsmålinger, audit og monitoreringer, der skulle gøre kontrol i 'armslængdeafstand' mulig (Christensen & Lægred, 2007). En ny fortsat balance mellem selvstyrende enheder og den centrale styringsmyndighed blev dermed tematiseret.

Ikke mindst aktualiseredes og 'kompleksiseredes' denne balance i forbindelse med, hvad der er betegnet som konkurrencestatsudviklingen. Det økonomiske effektiviseringspres, der har hvilet over offentlig uddannelse fra og med 1970'erne, suppleredes med mål for effektivisering i relation til vidensøkonomi og innovation (Pedersen, 2011. Lindholm, 1995). Uddannelse blev i stigende grad set som et økonomisk-strategisk element og genstand for reformer, hvilket gjorde den ministerielle styringskapacitet vigtig at fastholde.

Præstationsmåling, benchmarking og monitorering skyder sig nu ind mellem centraladministrationen og de frisatte institutioner, og økonomisk effektivitet bindes i den politiske diskurs sammen med uddannelseskvalitet. Kvalitetsudvikling – skønt ikke som et

ganske modsætningsfrit begreb – bliver et forvaltningsmæssigt issue. Hvordan det gribes an i forhold til de gymnasiale uddannelser, er emnet for denne artikel.

Vi spørger til tilsynet med de gymnasiale uddannelser i lyset af de store forandringer, det er undergået de sidste tre årtier, både hvad angår måden, hvorpå det gennemføres, og i udbredelsen. Vi spørger til den i henseende til dets effekt. Sagt på anden måde: Fungerer tilsynet ønskeligt – og da i hvilken betydning af ordet? Er der forbundet en risiko med denne tilsynsvirksomhed, og hvad er implikationerne for skolens ledelse?

For at anskueliggøre den ændring, der er sket, beskriver vi i artiklens første del tilsynets historiske udvikling fra 1990 og frem, hvor det på en ny måde blev et emne i den ministerielle forvaltning. I artiklens anden del redegør vi for fund gjort ved en række aktuelle interview med ledere på gymnasier, der er udtaget til tilsyn. I tredje del analyserer vi informeret af evalueringsforskning den form for audit, tilsynet udgør. Dette fører frem til diskussionen, som vi fører i fjerde og afsluttende del, der angår, hvilken effekt det kan have på uddannelsesopgaven bredt set og hvilke modsætningsfyldte implikationer og risici, det indebærer for ledelse af opgaveløsningen. Og endelig overvejer vi, om der er måder at drive tilsyn på, der i et vist omfang imødegår de risici, analysen peger på.

Vi er i vores analyse især inspireret af diskursiv institutionel teori, der siger noget om, hvordan institutioner i bred forstand konstitueres, men også hvordan diskurser og oversættelser af diskurser sætter rammer om praksis ved at fremhæve praksiselementer og udelade andre (se fx Kjær & Pedersen, 2001). Kvalitet er i dette perspektiv et interessant begreb og betragtes som en form for bricolage, hvor forskellige betydninger løber sammen i og med, at forskellige interessenter søger indflydelse på, hvilke betydninger af begrebet, der skal dominere med hvilke konsekvenser for handling.

Historisk- fra KUP til KPI

Fra og med slutningen af 1900tallet blev kvalitet og hvordan kvalitet identificeres et emne, der fik megen forvaltningsmæssig opmærksomhed. Man kan se en bestemt linje i denne udvikling – fra den første bølge, hvor man statsligt var fokuseret på decentralisering, til de senere, hvor den politiske styringskapacitet søges styrket ved recentraliserende initiativer (Andersen, 2013). For at vise ændringerne i kvalitetsforståelsen og for derved at tydeliggøre, hvordan det nye tilsyn afgørende adskiller sig, beskriver vi i det følgende er par trin i tilsynets historie.

Et kvalitetsudviklingsprojekt

I 1989 tog det daværende Gymnasiedirektorat initiativ til at udvikle kriterier for kvalitet. Det skete i form af et projekt, navngivet Kvalitetsudviklingsprojektet eller KUP (Haue et al., 1998). Hvor Undervisningsministeriets faglige tilsyn tidligere var tilrettelagt og gennemført i forhold til de enkelte fag ved fagkonsulenters besøg på skoler ud over landet, arbejdede man med en ny

type, der skulle orientere sig mod en udvalgt institutions hele måde at drive undervisning på. Praktisk foregik det ved en eller flere undervisningsinspektørers og fagkonsulenters flere-dages besøg på en enkelt skole, hvor man observerede undervisningen og talte med ledere, lærere og elever. De indtryk, man fik, samledes og blev udgivet i en rapport, der også indeholdt anbefalinger. Rapporten blev efterfølgende sendt til alle gymnasieskoler til inspiration.

I 1993 udgav ministeriet efter 12 sådanne besøg *Tegn på kvalitet* (Undervisningsministeriet, 1993), der opsummerede og bearbejdede besøgserfaringerne og derigennem søgte at *indkredse kvalitet*. Daværende direktør for Gymnasieafdelingen Uffe Gravers Petersen anførte, at et svar på, hvad kvalitet er, "naturligvis aldrig kan blive udtømmende", og at "en levende skole hele tiden må tage sit eget grundlag op til overvejelse, må vælge og prioritere" (s. 3)

I 1997 udgav man første version af *Standarder og profiler* (Undervisningsministeriet, 1997). Standarder og profiler havde ganske vist baggrund i erfaringerne fra KUP, men markerede dog en bevægelse i tilgangen, idet den var et forsøg på at identificere *tværgående standarder* for gymnasial undervisning. Formålet var, skriver man, at give skolerne anledning til selvkritik, at borgerne får pålidelig information, samt at skaffe politisk grundlag for at konstatere, at uddannelsens mål nås. Teksten nævner selv to risici ved dette tiltag, nemlig at standardisering reducerer en vanskelig bestemmelig kvalitet og at standardisering betyder centralisering (s. 10 ff.). Man betoner heroverfor, at der fortsat skal være rum til, at skolen udvikler sin egen profil.

Man kan anskue Kvalitetsudviklingsprojektet og de erfaringsopsamlinger og videreudviklinger, der fulgte, som en proces, der søger at institutionalisere en ny kvalitetsdiskurs, som rammesætter nye praksisser og udpeger nye roller. Kvalitet handler nu ikke alene om de enkelte fags faglige praksis, men gøres til et organisationsanliggende og dermed et anliggende for ledelse, og samtidig markerer det, at kvalitet er et emne, der ikke uden videre kan overlades til de fagprofessionelle. Man kan se *Standarder og profiler* som et første udtryk for, at selvevaluering skal suppleres med ekstern evaluering. Indtrykket af den udvikling forstærkes efterfølgende i en række regulative initiativer, som kort skal nævnes.

Kvalitet og lovarbejde

I 1999 kom EVA-loven, *Lov om Danmarks Evalueringsinstitut* (Folketinget 1999). EVA skal samle nationale og internationale erfaringer og udvikle metoder til uddannelsesevaluering og kvalitetsudvikling og rådgive myndigheder og det politiske system inden for disse spørgsmål. EVA foretager politisk-administrativt bestilte evalueringer af institutioner, der modtager offentlige tilskud, og man gennemfører evaluering af egen drift. EVA er på den måde en ny institutionel frembringelse, sat i verden til at identificere kvalitet i bestræbelsen på at institutionalisere et kvalitetsbegreb på tværs af offentlige uddannelser.

I 2002 kom *Lov om gennemsigtighed og åbenhed i uddannelserne m.v.* (Folketinget 2002), der gjaldt samtlige lovregulerede uddannelser, og som betød, at fagudbud, værdigrundlag, men også karakterer og indsatser vedrørende undervisningsudvikling skal være tilgængelige for offentligheden og dermed for brug i evaluering.

I 2005 kom *Bekendtgørelse om Kvalitetsudvikling og resultatvurdering inden for de gymnasiale uddannelser*. Her stilles en række krav til skolernes kvalitetssystem. Loven er et led i standardiseringen af evaluering på området, men her er der tale om selvevaluering. Bekendtgørelsen vil fremme systematikken i indhentning og intern brug af data, og den stiller krav til ledelsens rolle i processen og om opfølgning på resultatet (Folketinget 2005). Ud over at udbygge gennemsigtighed og åbenhed (mhp. eksterne interessenter) tilsigtes herved også etableret et ledelsesinformationssystem.

Samlet set institutionaliseres evaluering derved som en offentlig rutine og samtidig markeres ledelse som *accountable*, dvs. vedrørende dens ansvarlighed i forhold til redegørelse for arbejdet med og resultaterne i forhold til de krav den offentlige opdragsgiver måtte stille.

KPI og det nye centrale tilsyn

Med et kvalitetssystem mere eller mindre på plads på de enkelte institutioner er det samtidig muligt at følge kvalitetsarbejdet eksternt. Fra og med 2006 fulgte en række planer for, hvordan man centralt vil angribe tilsynsopgaven. Formålet er, skriver Undervisningsministeriet på sin hjemmeside, at dele viden om undervisningens resultater og god praksis, men i et videre perspektiv kan det også ses som en aktivering af evaluering som led i den politiske styringskæde (Andersen 2013).

I 2011 oprettedes som endnu i ministeriet som en institutionel nyskabelse og dermed markering af fokus på kvalitet Styrelsen for Undervisning og Kvalitet (STUK). STUK er organisatorisk adskilt i et fagligt og et økonomisk tilsyn. Inden for det faglige tilsyn skelnes yderligere i fire former for tilsyn – et risikobaseret, et tematisk, et undersøgende og et enkeltsagstilsyn (se fx STUK 2024). Vi hæfter os i det følgende ved det risikobaserede tilsyn.

Det risikobaserede tilsyn udøves på baggrund af målinger af den enkelte skoles præstation. Det sker ved de tal, som skolerne forvejen indberetter til ministeriets eksamensdatabase og Danmarks Statistik, og som opsamles i Uddannelses- og Forskningsministeriets Datavarehus. Herudfra udvælges tal, der er besluttet som indikatorer på kvalitet, de såkaldte *Key Performance Indicators* (i daglig tale kpi'er). Tallene gør screening og sammenligning af gymnasier mulig. Skoler udtages til tilsyn, hvis de på en eller flere indikatorer over en periode på ligger lavt i forhold til de øvrige.

Antallet af indikatorer er blevet udvidet over årene. I 2024 screenes de gymnasiale uddannelser på følgende: Eksamensgennemsnit, frafald, overgang til videregående uddannelser inden for ét og to år, løfteevne (socioøkonomisk reference, dvs. evnen til at løfte eleverne i forhold til deres

sociale baggrund og faglige niveau fra grundskolen), elevtrivsel og risiko for udvikling af parallelsamfundsproblematikker (andel af elever fra udsatte boligområder og elever, der ikke består skriftlig dansk på A-niveau).

Udtages en skole til tilsyn, henvender styrelsen sig til skolens bestyrelse, der er den formelt ansvarlige. På baggrund af skolens redegørelse for det påpegede forhold afgøres, om skolen skal fortsætte i tilsyn. I det tilfælde skal skolen udarbejde en plan med beskrivelse af de initiativer, hvormed man vil møde problemet og rette præstationen op. Eventuelt tilbydes skolen rådgivning fra ministeriets læringskonsulenter. Der kan gives påbud, indgås indsatsaftaler eller der kan pålægges sanktioner.

I 2022 modtog 40 gymnasier (hvilket må siges at være relativt høj andel af samtlige 178 institutioner, der udbyder gymnasial undervisning) et såkaldt opmærksomhedsbrev og ifølge tilsynsberetningen for samme år fortsatte 31 i kortere eller længere tilsynsforløb. Fire har fået påbud (STUK 2023).

Fra proces til resultat

Over fire årtier kan man således tegne en bestemt udvikling. Det faglige tilsyn før KUP foregik som noget, der kan sammenlignes med fagfællebaseret evaluering, som foregår i en dialog mellem ministerielt ansatte med baggrund i sektoren (inspektører og fagkonsulenter) og den enkelte institutions lærere og i begrænset omfang også med institutionens ledere. Tilsynet byggede på mangeartede data, herunder ikke mindst kvalitative (data fra interview, observation og analyse af lokale dokumenter), og drejede sig om forståelse og forhandling af de enkelte faggruppers praksis. Formålet var at omfatte kvalitet i al dens sammensatte – herunder lokale – bredde og opfattedes som knyttet til opgaveløsningens proces.

I det kpi-baserede tilsyn er fokus på resultatet, og kvalitet er repræsenteret ved kvantitative størrelser, indikatorer. Skønt kpi'erne som sagt er indikatorer, udgør de alene baggrund for beslutninger om tilsyn og krav om indsatser for forbedring. Den standard, der etableres, sker ved en form for et omvendt *break even*, hvor de lavest-præsterende i forhold til populationens gennemsnit udtages. Det kpi-baserede tilsyn og den screening, der indgår, rummer derved en form for automatik, hvor det professionelle skøn i beslutningen reduceres. Ledelsens – rektors og bestyrelsens – opfyldelse af deres ansvar gøres så at sige måleligt. Som sådan kan systemet paralleliseres med, hvad der i generisk teori om organisationsstyring kaldes et ledelsesinformationssystem. Styrelsen fremhæver i den seneste tilsynsplan, at den derved tilsigter et virkningsfuldt, helhedsorienteret og effektivt tilsyn (STUK 2024, s. 7)

Det er denne udvikling og ambition for STUKs tilsynsvirke, vi diskuterer i det følgende – først gennem en mindre, empirisk undersøgelse, vi har foretaget blandt gymnasiets ledere, senere i lyset af teoretiske analyser vedrørende kvalitet og evaluering informeret af evalueringsforskning: Indfrir tilsynet på den måde, det drives, de politiske og administrative

ambitioner, STUK nævner, og hvilken effekt har det? Og i den forbindelse: Hvilket risici er knyttet til tilsynsmetoden – og hvilke implikationer for ledelse skaber det nye tilsyn?

Et blik på rektor erfaringerne

Vi har interviewet fem rektorer og en uddannelsesleder fordelt på store og små gymnasier, i storby og i mere tyndtbefolkede egne. Fælles er, at skolerne har været i tilsyn i kortere eller længere tid, et par af dem siden tilsynet i den nuværende form startede i 2014! Målet med undersøgelsen er at få et indtryk af, hvordan tilsynsaktiviteten opleves. Vi har spurgt til, hvordan man reagerede på, at gymnasiet blev udtaget til tilsyn, hvad rektorerne gjorde af tiltag som reaktion herpå og hvilket udbytte, man skønnede at have fået af tilsynet. Opleves det således blandt lederne som virkningsfuldt, helhedsorienteret og effektivt, jævnfør den hensigt, vi citerede STUK for ovenfor?

Gennemgående nævner rektorerne en ikke ubetydelig arbejdsmæssig indsats som følge af dels at skulle forholde sig til bestyrelsen, der modtager tilsynsskrivelse, dels at gennemføre diverse analyser og redegøre for handleplaner. De nævner også deres bekymring over, hvad et tilsyn kan betyde for gymnasiets lokale omdømme og dermed for dets strategiske position i konkurrencen om elever. Vores fokus er dog først og fremmest på, hvad der erfares om tilsynet som motor i kvalitetsudvikling, som ifølge STUK er den helt grundlæggende intentionen med tilsynet, og som man må se som baggrund for, at tilsynet skal kendes på at være helhedsorienteret, virkningsfuldt og effektivt. Og her forekommer svarmønstrene tvetydige. I det efterfølgende har vi rubriceret interviewresultaterne i forhold til STUKs formulerede, tredelte intention.

Helhedsorienteret?

Rektorerne anfægter på ingen måde tilsynets principielle legitimitet – som leder af en offentlig institution må man være indforstået med at stå til regnskab for institutionens ydelser (accountability). Tvetydigheden i svarmønstrene angår nærmere den nævnte intention.

Tværtimod at virke helhedsorienteret er det en oplevelse på skolerne, at indikator teknologien kan virke fragmenteret og fragmenterende og være blind for den systemsammenhæng, der kan være mellem kpi'erne. En indsats for højere gennemførelsesprocent kan således modvirke eksamensgennemsnittet. Vellykkethed med henblik løfteevne afhænger af eksamensgennemsnit for de elever, gymnasiet modtager: Er eksamensgennemsnittet højt, er løfteevnen erfaringsmæssig lav og vanskelig at forbedre. En specifik og stærk indsats for at fremme karaktergennemsnittet kan virke negativt på trivselskpi'en osv.

Det fragmenterende viser sig i overvejelser hos lederne over tilsynets mulige virkning på uddannelsens formål. Man nævner, at indikatorerne først og fremmest synes at have fokus på den gymnasiale undervisnings studieforberedende mål og nærmere betegnet den stabilt

arbejdende elevs præstationer og ikke på realiseringen af uddannelsens transfaglige mål og demokratiske dannelsesformål. Det er mål, der forudsætter, at forskellige og forskelligt præsterende elever lærer at omgås hinanden i faglige samarbejder. Et kvalitetsbegreb, der reagerer på enkelt-kpi'er, opleves på skolerne alt i alt som skævtrækkende i forhold til uddannelsens helhed og formål.

Virkningsfuldt?

Det fremgår mange steder i interviewmaterialet, at der er indikatorer som skolerne kun oplever at have ringe indflydelse på. Det gælder de konkrete, lokale optageområders særlige sociale og kulturelle forhold spændende fra områder med lave, henholdsvis høje forudsætninger, men det gælder også gymnasier i områder med tilsyneladende stabile normer for uddannelse. Vi hører om tilsyn, der trods skolens egne eksperimenterende tilgange og med brug af ekstern rådgivning i op til mere end 10 år ikke har oplevet, at den kan påvirke de indikatorer, der udløste tilsynet. Fordi de resultater, som indikatorerne efterspørger, er forårsaget af mangfoldige, men løstkoblede forudsætninger, vil indsatsers virkning ofte være usikre, til tider umulige at forudsige og eventuelt vise sig på helt andre måder end intenderet. Blindheden for systemsammenhænge synes således til tider at undergrave ambitionen om et tilsyn skal være virkningsfuldt set i forhold til skolernes handlemuligheder.



At kpi'er hver for sig kan udløse tilsyn, giver anledning til bizarre situationer. Der berettes om tilfælde, hvor kpi'en er fremkommet på basis af data fra så få elever, at udsigelseskraften statistisk kan anfægtes. I andre tilfælde ville mindre ændringer af den periode, hvor en måling er foretaget, ganske have ændret resultatet. Det medfører, at ledelserne oplever at have svært ved at motivere lærerne til at tage tilsynet alvorligt.

Effektivt?

Den ambition, STUK har om, at tilsynet skal være effektivt (forholdet mellem forbruget af ressourcer brugt i tilsynsprocessen og de ønskede virkninger, STUK 2024 s. 7), drages i interviewene med rektorerne stærkt i tvivl. Der nævnes som sagt, at der bruges meget tid på redegørelse og dokumentation, men at det til tider er svært at se, hvilken indflydelse den

didaktiske og pædagogiske indsats kan have på de påpegede lave resultater, endside hvilke konkrete indsatser, kpi'erne peger på. Det hævdes, at STUK ikke er til megen hjælp på det punkt, og det nævnes, at de indberetninger, man efter et større lokalt arbejde indsender, ofte blot tages til efterretning og ikke indgår i en efterfølgende dialog mellem STUK og den pågældende skole. Det medvirker til, at tilsynet opleves som kontrol og som mistillid og i mindre grad som sparring om effektivitet forstået som kvalitetsudvikling. Indsendte planer for kvalitetsudvikling opleves som køb af tid – nu vil der gå der tid, inden skolen igen stilles over for kpi-krav.

Selv om aktindsigt i en korrespondance med et par af gymnasierne viser, at problemet, der udløste tilsynet, ikke grundlæggende har ændret sig over flere år, vil vi dog ikke på baggrund af vores interview kunne afvise, at tilsynet kan have en effekt, der kan aflæses på de nævnte KPI'er. I det ene af de omtalte tilfælde har tilsynet givet anledning til indsatser, der aktuelt virker lovende og muligvis med tiden også vil kunne registreres på kpi-niveau. I et andet tilfælde betød tilsynet, at man har fået overvejet og anvendt ressourcer på måder, der vurderes at være en forbedring, også selv om man finder det usikkert, om forbedringen vil kunne registreres på de kvantitative opgørelser.

Alt i alt viser interviewundersøgelsen, at den intenderede virkning ikke entydigt synes at harmonere med intentionen om det helhedsorienterede, virkningsfulde og effektive. Vi peger på, at tilsynsteknologien opleves at skævtrække den samlede virkning ved at sætte fokus på enkeltaspekter og ved at være blind for eventuelle systemsammenhænge og grundlæggende ikke have den effekt, tilsynet selv annoncerer. Som helhed kan interviewene tolkes som en kritik af underliggende uudtalte årsags-virkningsantagelse, designet synes at bygge på. Endelig kan undersøgelsen først og fremmest tolkes som en efterlysning af et tilsyn, der tager sit afsæt i helhedsorienterede, faglige-pædagogiske vurderinger fremkommet i dialog med skolerne.

“VI PEGER PÅ, AT TILSYNSTEKNOLOGIEN OPLEVES AT SKÆVTRÆKKE DEN SAMLEDE VIRKNING VED AT SÆTTE FOKUS PÅ ENKELT-ASPEKTER OG VED AT VÆRE BLIND FOR EVENTUELLE SYSTEMSAMMENHÆNGE OG GRUNDLÆGGENDE IKKE HAVE DEN EFFEKT, TILSYNET SELV ANNONCERER.”



Samlet set giver interviewundersøgelsen ikke belæg for at hævde, at tilsynet i rektorperspektiv opleves at være den afgørende motor i den kvalitetsudvikling, STUK tilsigter i deres formulerede intentioner. Tvært imod giver denne mindre, empiriske undersøgelse af erfaringer med at være i tilsyn en række indikationer på forhold, der vanskeliggør den rektoropgave indadtil, der består i ledelse af den løbende kvalitetsudvikling: Den fragmenterende tendens, teknologiens manglende evne til at reflektere systemsammenhænge, eksemplerne på manglende baggrund

for statistisk udsigelseskraft undergraver den legitimitet, der er nødvendig i en meningsskabende lederrolle (Smirchich & Morgan, 1982).

Tilsynet i lys af evalueringsforskning

Den form, STUK udfører sit tilsyn i, er i evalueringsforskning kaldt præstationsmåling – en vurdering af en given indsats gennem et begrænset antal kriterier, oversat til en til en specifik score (Kroghstrup, 2006). Derigennem kan en given institutions opgavevaretagelse overvåges eller monitoreres (Albæk og Rieber, 2001). Det er vigtigt at gøre sig klart, at præstationsmåling er noget ganske andet end evalueringsformen effektmåling. Præstationsmåling opererer med kvantitative størrelser, der er valgt med henblik på at vise udviklinger eller forskydninger i det pågældende felt, men kan ikke gøre krav på at klarlægge indsatsens hele effekt, sådan som STUKs formulerede tilsynspolitik ellers antyder.

Målingens valgte målepunkter omtales som indikatorer. En indikator er ifølge Ordbog.dk et tegn eller "middel som bruges til at vise hvordan noget forholder sig" (jf. Den Danske Ordbog, <https://ordnet.dk/ddo/ordbog?query=indikator>). Præstations- eller performanceindikatorer – eller, som de (med inspiration fra regnskabsføring) i det konkrete tilfælde benævnes, *kpi'erne*, er således valgte tegn, der skal vise hen til forhold, der er mere komplicerede eller komplekse end repræsentationen. Der er med andre ord tale om en oversættelse, der reducerer og dekontekstualiserer en kompleks substans, for en alternativt overskuelig. Det er en pointe, at indikatoren udgør valg og som sådan har betydning for, hvad der bliver synligt, og hvad der unddrager sig synlighed. Meget af den kritiske granskning, der foregår omkring præstationsmåling, handler om dette forhold. I denne artikels sammenhæng gælder præstationsmålingen undervisning og den kompleksitet, der knytter sig til hertil, og det gør det opportunt at overveje, hvilke effekter selve præstationsmålings teknologi kan have på det evaluerede – *dens konstitutive effekt* (se fx Dahler-Larsen & Kroghstrup, 2001).

Den første del af det følgende omtaler kompleksiteten og de risici, præstationsmålings konstitutive effekt udgør. Anden del uddyber, hvordan de disse effekter endog kan virke kontraproduktivt på de visioner, der fra politisk hold gør sig gældende vedrørende gymnasial undervisning, især fra og med reformen i 2005. Vi peger hermed på, at ledelse af uddannelse i den henseende befinder sig i en modsætningfyldt felt.

Indikatormåling og kompleksitet

Styrken ved brugen af performanceindikatorer er overblikket. Svagheden er, at overblikket tilvejebringes ved at reducere kompleksitet. Især i forhold til velfærdsorganisationers komplekse ydelser er dette et bemærkelsesværdigt og risikabelt forhold (Norman, 2001).

At kompleksitet er et vilkår i velfærdsproduktion kan begrundes i, at den offentlige institutions opgaveløsning skal give mening i flere perspektiver, således i aftagerens – her elevens – og i

perspektiv af samfundet som helhed (Antonsen & Jørgensen, 1999, Rothstein, 1994). Denne dobbeltperspektiviske fordring om både at tage afsæt i aftageren og samfundet er ikke altid modsætningsfyldt, men kan være det, og er for eksempel nærværende, når det gælder uddannelsesmæssige spørgsmål: De samfundsbestemte krav kan være nok så tydelige, men kan eleven ikke individuelt se meningen med den undervisning, der skal møde kravene, går det ud over den faglig-pædagogiske effekt.

En anden kompleksitetsårsag vedrørende den offentlige velfærdsorganisations opgaver skyldes, at opgaven ofte hverken er præcist defineret, eller at viden er entydig om dens løsning. Undervisningens problem er således mere vildt end tamt (Harmon & Mayer, 1986): Hvad skal eleverne kunne for at kunne begå sig i en ukendt fremtid som borgere og som arbejdskraft, og hvordan skal lærere facilitere, at eleverne nærmer sig at mestre noget, der i øvrigt kun er utydeligt kendt? Vi har således at gøre med en række vilkår, som kolliderer med de forudsætninger, der ifølge Cave et al. (1997) skal være til stede ved konstruktion af performanceindikatorer, som gør, at de forekommer konsistente og veldefinerede i deres indbyrdes sammenhæng, at der er klare kriterier for, hvordan de anvendes i ledelsesmæssige beslutningsprocesser frem mod et entydigt mål (s. 39). Den rationalitet og de forudsætninger, som performanceindikatorer hviler på, synes at knytte sig til, hvad man med i Thompson og Tudens (1959) opstilling om alternerede beslutningsbaggrunde vil kalde det tekniske felt (klar viden om årsags-virkningssammenhænge, klare præferencer). Det er forudsætninger, der gør sig gældende for de færreste af de beslutninger, der indgår såvel i den pragmatiske lærerhverdag (Arfwedson, 2001).

I det konkrete tilfælde synes kpi'erne dog ikke konstrueret ud fra en samlet og sammenhængende analyse, men er kommet til hen over årene som afspejlinger af aktuelle politiske debatter (fx løfteevne, trivsel og sidst vedr. risiko for parallelsamfundsdannelse, se fx STUK 2024).

En sammenligning af den rationalitet, der knytter sig til præstationsmålinger, og rationaliteten, der knytter sig til den komplekse uddannelsesmæssige praksis, åbner for en modsætning, der gør overvejelser over præstationsmålings anvendelse opportune. Vi vil med reference til Hanne Kathrine Kroghstrup og Burt Perrin fremhæve tre risici, som vi i relation til STUKs tilsyn finder væsentlige. Det gælder for det præstationsmålingernes abstraktion fra kontekst, dernæst risikoen for målforskydning og endelig målingernes eventuelle blindhed for opgaveløsningens samlede effekt (Kroghstrup, 2006 og Perrin, 1998).

Præstationsmålingernes abstraktion fra kontekst er en grundlæggende forhold også i denne tilsynsteknologi. Fordi resultaterne skal indgå i sammenligninger, må målingerne baseres på tværgående og derved kontekstuaafhængige standarder. I denne artikels tilfælde er gymnasiernes statistiske indberetninger data, men det er data, der repræsenterer forskellige kontekster og de udfordringer, som sådanne lokale kontekster udgør. På baggrund af vores interview synes det for eksempel at give mening, at løfteevnekpi'en såvel som for eksempel karakterkpi'en og kpi'en for overgangsfrekvens lige så vel kan sige noget om det givet

gymnasiums lokale kontekst, end den er et udtryk for en undervisningsmæssig præstation. Er det tilfældet, betyder det en afgørende forskel for det enkelte gymnasiums handlemuligheder.

Målforskydning ved præstationsmåling er et andet forskningsmæssigt beskrevet problem. Præstationsmåling er alene fokuseret på resultatet og siger intet om processen eller den praksis, der fører frem til dette (se fx Newcomer, 1997). Et fokus på en specifik praksisindsats, besluttet med øje på at rette op på en givet talværdien (og således blind for en eventuel systemsammenhæng), kan have væsentlige ikke-intenderede resultater og fremkalde ikke-intenderede og i den større sammenhæng u hensigtsmæssige rutiner. Dette er et velbeskrevet fænomen inden for almen studier af såkaldt frontpersonales ageren i forhold til indikatormåling (Lipsky, 1980), og inden for uddannelse er det klassiske og stadig vigtige eksempel den såkaldte wash back-effekt *teaching to the test*. I rektorinterviewene kan spor af samme fænomen være bekymringen for, at planer for forbedring af visse kpi'er snarere bliver betragtet som 'køb af tid' frem for seriøse forbedringsforsøg. Det snævre fokus på specifikke præstationer overser eventuelt lige så vigtige, men svært målelige effekter af den hidtidige praksis og kan betyde, at effekten af indsatsen får uheldige virkninger på andre effekter. Dette fænomen er veldokumenteret i international litteratur (se fx Tienken & Canton, 2009) og diskuteres også i undersøgelser af virkninger af de nationale test indført i den danske grundskole (se Bundsgaard & Puck, 2016). Forholdet giver også anledning til bekymring blandt de interviewede ledere og omtales som en risiko for skrævrækning af gymnasiets almindelige effekt.

Som et tredje og sidste forhold skal nævnes spørgsmålet vedrørende præstationsmåling og en given institutions samlede effekt. Perrin nævner, at en institution kan have og oftest har en virkning, der ikke umiddelbar er målelig (og muligvis endog ikke formuleret), og som præstationsmåling derfor vil være blind for, til trods for, at virkningen kan være aldrig så vigtig. Et gymnasium kan således indtage en symbolsk rolle i et ressourcetsvagt område, fordi dets eksistens demonstrerer, at almene interesser også gælder det pågældende sted. En sådan effekt kunne således være et bidrag til den regionale eller nationale sammenhængskraft, skønt den næppe er kvantitativt målelig. Foreløbige og upublicerede resultater af undersøgelser af lokale fortællinger om et af de gymnasier, der på grund af 'lave tal' fik optage-stop i 2021, kunne tyde herpå.

Perrin (og Kroghstrup med ham) vil ikke generelt afvise præstationsmålingers brugbarhed, men påpeger, at risikoen ved at anvende dem bør medreflekteres. Den magtfulde diskurs, som består i effektivisering gennem denne teknologi, risikerer at udblænde virkninger, som den allerede udviklede uddannelsespraksis har og som er væsentlige i lokalområdet.

Præstationsmåling og risikoen for kontraproduktivitet?

Cave et al. (1997) nævner, at på trods af ønsker om forandring kan præstationsmålingsteknologiens virkning lige så vel blive status quo. De begrundet det med, at teknologien forudsætter veldefinerede og dermed allerede kendte standarder, og derved i

princippet anlægger et 'mere af det samme-syn' på effektivitet. Det uddyber vi i det følgende med henblik på gymnasieområdet, hvor vi finder to forhold, hvor kontraproduktivitet kan komme til at dominere overordnede politisk-strategiske intentioner. Det ene omhandler politiske ønsker vedrørende et nyt faglighedsbegreb, det andet om samfundsudvikling og dannelseskraft.

Det nye faglighedsbegreb og præstationsmåling indgår i en paradoksale sammenhæng. Cave et al. (1997) ser baggrunden for udviklingen af præstationsmåling og monitoreringen i sammenhæng med, at diskurser vedrørende uddannelse forandres fra at fremstille uddannelse som et alment gode til en samfundsmæssig økonomisk-strategisk investering. Her finder de grunden til den øgede politiske interesse for kontrol med institutionernes performance (Ibid. s. 103ff). Men samtidig ser vi nationale og transnationale diskurser fra tunge, transnationale institutioner, der med afsæt i samme økonomisk-strategiske sammenhæng vil styrke faglighedens anvendelsesaspekt, og som argumenterer for, at dette skal gribe ind i undervisningens procesformer (Kereluik et al., 2013). Undervisningen skal frembringe 'langtidsholdbar arbejdskraft' (OECD, 1996), der løbende af egen kraft kan tilpasse sig og bidrage til den hastige udvikling. Det er argumentationen for det kompetencebegreb og den kompetenceudvikling, der sættes på dagsordenen i gymnasiereformeringer fra og med 2005 (se fx Hjort & Raae, 2014). Nu betones det transfaglige moment i faglighed, det vil sige evnen til på tværs af fagene at se og definere problemer i det foreliggende på nye og overraskende måder. Paradokset opstår i og med, at de specifikke målepunkter, som præstationsmåling hviler på, trækker i modsat retning (se også Perrin 1998, s.375). Den styrkelse af det formale eller procesmestrende element (evnen til at definere problemer, evnen til nyskabende løsninger), som kompetencebegrebet betegner, reflekteres ikke i den målingsteknologi, der hviler på standardiseringer.



Der til kommer et forhold, der også fremhæver behovet for et transfagligt fokus, idet også det 'at leve moderniteten' stiller spørgsmål til uddannelse på andre måder. En kulturel udvikling, der er præget af frisættelse fra kulturelle overleveringer, udgør samtidigt et pres på individets evne til at orientere sig i den kollektive sammehæng, man indgår i (se fx Ziehe, 1989 og Ziehe & Stubenrauch, 1990). I undomsuddannelserne synes det at betone spørgsmålet om almindannelse på en ny måde, sammenlignet med det klassiske gymnasiums

dannelsesbegreb: Behovet for dannelsens formale og etiske virkninger øge – at kunne anvende, hvad man lærer, i et engagement vedrørende vedligeholdelse af det fælles, for derved også selv at styrkes (Schmidt, 1999). Sådanne virkninger er, som vi har nævnt, netop præstationsmålingens svaghed, hvilket også er en kritik, vi ser fremhævet i interviewene med skolernes ledere.

Både i forhold til den ny faglighedsopgave ('kompetencefaglighed' i 2005 og senere reformer) og i forhold til den stadig mere påtrængende dannelsesopgave i senmoderniteten er det således en oplagt risiko, at målingsteknologien snarere end at være motor i udviklingen fungerer kontraproduktivt ved at understøttende fragmenterende forbedringsindsatser.

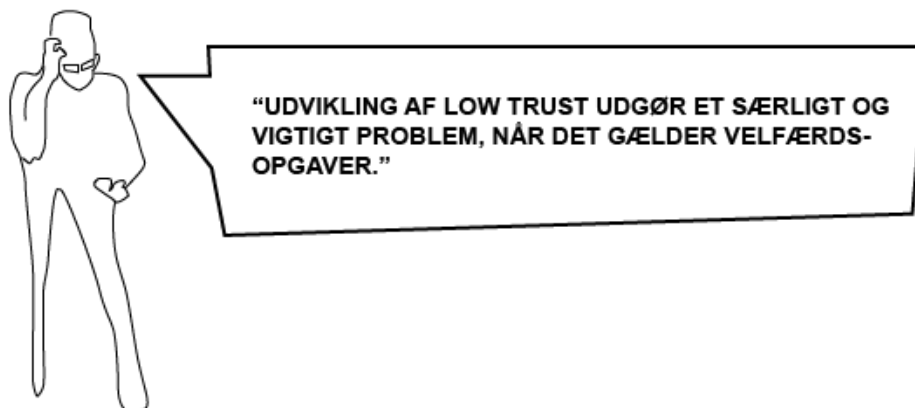
Præstationsmåling i styringsrelationen og low trust-cultures

At de gymnasiale uddannelser er overgået til selveje, gør den statslige styringsrelation mere kompleks. På den ene side er rektor (sammen med bestyrelsen) gjort økonomisk og strategisk ansvarlig parallelt med frisættelsen. På den anden side (og som omtalt i forbindelse med de styringsmæssige dobbeltbevægelser) er frisættelsen af institutionen udfordret i mødet med den politisk diskurs om uddannelse som økonomisk-strategisk element. Vi har således tidligere nævnt, at det nye tilsyn med gymnasierne kan forstås på baggrund af et ønske om at tage styringskapacitet tilbage.

STUK beskriver selv formålet med tilsynet som kontrol og læring. Det er to funktioner blandt en række, beskrevet i evalueringsforskning, men det er funktioner, som også indbyrdes kan være på kant med hinanden (Dahler-Larsen & Larsen, 2001). I artiklens empiriske materiale kan vi som nævnt finde eksempler på læring igangsat af tilsynet, men den tætte kommunikation, som læring ifølge evalueringsforskning forudsætter (ibid) oplever rektorerne i vores undersøgelse fraværende, hvilket betyder, at kontrolfunktionen står tilbage. Dette bekræftes af de gennemførte interview.

Evaluering som kontrol kan da også ses som en nærliggende håndteringsløsning af kompleksiteten vedrørende decentralisering og centralisering. Dette findes beskrevet agentteoretisk (se fx Ravn, Nygaard, & Kristensen, 2006). Lederen – rektor – er agent for ministeriet og skal på vegne af ministeriet (principalen) forestå løsning af den allerede definerede opgave. Men principalen må tage skridt til at sikre, at opdraget udføres i forlængelse af principalens intentioner, for det asymmetrisk informationsforhold og de to parter forskellige interesser er en risiko: Agenten er tættere på realiseringsprocessen og ved mere om den end principalen, og agenten kan have mål, der ikke er sammenfaldende med principalens. Monitorering og tydeliggørelse af sanktionering er mekanismer, der kan tages i brug for at fremme principalens interesser (hvilket i øvrigt er tilfælde i STUKs beskrivelser af tilsynsplanen, STUK 2023 og 2024). Ulempen herved er imidlertid, at den gensidige positionering at parterne ofte bliver omdrejningspunkt i udvikling af såkaldt *low trust culture*. Netop oplevelsen af mistillid fremstår også som et stærkt markeret mønster i vores interviewempiri.

Udvikling af *low trust* udgør et særligt og vigtigt problem, når det gælder velfærdsopgaver. Professionsteori nævner, at løsning af svært præciserbare opgaver forudsætter et professionelt skolet skøn og en vis *automi* (se fx Freidson, 1986), og videre forudsætter autonomi den tillid, som principal-agentrelationens rationalitet tendentielt underminerer.



Den styringsrelation, som kpi-målingen udformer, bærer således efter alt at dømme konstitutive effekter med sig, der risikerer at virke kontraproduktivt også på læreres og lederes professionelle engagement. Også på den baggrund synes tilsynsteknologien nærmere at besværliggøre end at understøtte lederopgaven forbundet med et løbende kvalitetudviklingsarbejde.

Tilsynets symbolske effekt?

I en artikel, der perspektiverer den omfattende brug af evaluering, som har afløst plansamfundet, gør en række danske evalueringsforskere en form for status over det offentlige brug af evaluering (Albæk et al. 2001). Overordnet ser de evaluering som en efterhånden befæstet og institutionaliseret aktivitet, begrundet i ikke mindst den stigende økonomiske interesse for, hvad der kommer ud af offentlig politik. Men de anfører også, at evalueringstræthed efterhånden synes at være en faktor. Den utilfredshed, som vi også konstaterer i interviewene, er for dem at se forbundet med, at evaluering generelt ikke synes at have den ønskede effekt (s. 260 f). Alligevel spår Albæk et al., at målopfyldelsesevaluering og måling af institutionernes præstationer vil få en stigende ubredelse. Det gør spørgsmålet, hvorvidt sådanne kvantitativt baserede evalueringer også følges kvalitativt op med henblik på effekt, centralt.

Vores empiri kan ikke tages til indtægt for, at kvalitativ opfølgning er tilfældet, når det gælder tilsynet med gymnasiet, mens det modsatte til gengæld hævdes. Det kan være én blandt de mange mulige grunde til oplevelsen af ringe effekt. Men måske man skal lede efter effekt i andre betydninger? Måske som et tiltag først og fremmest af symbolsk karakter, der (set fra styrelsesniveau) rækker såvel op som ned?

Den symbolske funktion opadtil – mod det politisk lag – ville i så fald være, at sætte tilsynet i stand til at demonstrere sin kapacitet med henblik på at handle responsivt i forhold til ekstern og politisk kritik. Vi kan se, at denne symbolske funktion faktisk ind imellem aktiveres. Eksempelvis sker det i forbindelse med en beretning fra Rigsrevisionen til Folketinget vedrørende studenters overgang til videregående uddannelser (Rigsrevisionen 2021). Her påpeger Rigsrevisionen, at overgangsfrekvensen til videregående uddannelser er faldende. STUK reagerer med 'mere af det samme' ved derefter at foretage målinger hvert år frem for hver 3. år og ved at udtage flere gymnasier inden for kategorierne. Til trods for, at effekten af tilsynet synes tvivlsom, kan man således demonstrere (symbolsk) handlekraft.

I og med det kontinuerte kontrolpres mod skolerne kan man tillige tale om en symbolsk funktion nedadtil. En grundlæggende pointe i governmentality-teori om moderne styring er, at frie individer må formes for at kunne handle ansvarligt, og at et nærvær af normer og standarder skaber bestemte subjektive orienteringer, sådan at en givet opmærksomhed bliver naturlig, ja, taget for givet (Dean, 2006, fx s. 262). Måske kan man i forlængelse af en sådan teori overordnet læse kpi-tilsynet som en sådan symbolsk 'opdragelse', der skal skabe og forme lederes ansvarlighed? Med perspektivet den symbolske funktion op- og nedadtil udvider vi det funktionsbegreb, som tilsynet selv anfører, som henviser til de funktionelt registrebare resultater (se STUK 2024, jf. tidligere). Og med den symbolske funktion nedadtil taler vi om en særlig positionering af (gymnasie-)ledelse – en positionering, der – når den skal oversættes til funktionel ledelse – trækker væk fra eller virker eroderende i forhold til, hvad er blevet betegnet substantiel ledelse. Dette er emnet for det følgende afsnit, hvor vi afsluttende deler et par overvejelser om, hvordan tilsynet alternativt kunne drives.

"[HER] TALER VI OM EN SÆRLIG POSITIONERING AF (GYMNASIE-)LEDELSE – EN POSITIONERING, DER – NÅR DEN SKAL OVERSÆTTES TIL FUNKTIONEL LEDELSE – TRÆKKER VÆK FRA ELLER VIRKER ERODERENDE I FORHOLD TIL, HVAD ER BLEVET BETEGNET SUBSTANTIEL LEDELSE."



Substantiel ledelse og tilsynets forandring?

Ifølge Ove K. Pedersen fører moderniseringsprogrammerne for det offentlige fra og med 1980'erne til, hvad han betegner som 'den omringede institution' – de mange statslige enheders selvstændighed lagt i en ring af styring, kontrol og tilskyndelser: Lederne har fået ansvar for drift også i udvidet betydning, men også for ændringer og tilpasninger til de krav, der strømmer imod institutionen fra nævnte styrings-, kontrol- og incitamenttiltag (Pedersen, 2011, s. 216-217). Nye tilsynsområder og øgede tilsyn kan ses som et led i dette – ligesom den modsætningefulde positionering af ledelse, der følger med:

KUP var et forsøg på at konstruere et begreb om kvalitet under indtryk af den kompleksitet, der kendetegner uddannelse, ikke som nagelfast og endeligt, men som et emergerende begreb, der udvikles i dialog med den enkelt skole, med sektoren og opdragsgiveren, insisterende på at inddrage dets kvalitative aspekt. Den opsamling, den daværende Gymnasieafdelingen foretog på baggrund af de gennemførte besøg, var da også 'åben', hvilket både titel på opsamlingen og forordet – *Tegn på kvalitet* – antyder (Undervisningsministeriet, 1993).^[1]

Kpi-tilsynet adskiller sig afgørende herfra. Det er kvantitativt registrerende, ud fra generelle, standardiserede og centralt besluttede målepunkter, hvor målestokken tilvejebringes gennem sammenligning af skolernes præstationer. Selv om kvalitet afgøres gennem benchmarking og skalaen derfor ikke er absolut, er rationaliteten, at kvalitet kan fikseres og en kvalitetsstandard fastlægges gennem et indikatorniveau. Teknologien fremstiller tilsynet som handleparat og i modsætning til, hvad der tidligere gjaldt for tilsynet, tillige som suverænt handleanvisende. Det er den rationalitet og sikkerhed, der kommer til udtryk, når styrelsen på hjemmesiden nævner, at "Vi omsætter viden og data om, hvad god kvalitet er, til handlingsanvisende vejledning" (STUK 2024).

Rektor har (sammen med bestyrelsen) til enhver tid ansvaret for opretholdelse af uddannelsernes kvalitet, men rollefunktionen positioneres afgørende forskelligt i den tidlige og den aktuelle tilsynsdiskurs. I det første tilfælde (KUP) specificeres kvalitet kun omtrentligt og fastholdes som noget, der emergerer i mødet mellem det centrale og lokale niveau. Fordi kvalitet i KUP-tilgangen er en ikke-fikseret størrelse, vil ledelsesansvaret – i dialog med tilsynet – bestå i en løbende og principielt aldrig afsluttelig afgrænsning af en praksis, der forhandlende skal forholde sig til mangehånde konkurrerende outcomeforventninger. Det er en ledelsesaktivitet, der kan sidestilles med ledelse af, hvad Nils Brunsson kalder den politiske eller multideologiske institution (Brunsson, 2003). I det andet tilfælde (kpi-tilgangen) tages afsæt i en fiksering af kvalitet, hvad der gør eksekveringsfordringer entydige og tydelige på baggrund af præstationsscreeningen. I dette tilfælde positioneres rektor som leder af en handleorganisation (ibid.). Brunsson vil hævde, at disse to organisationsaspekter og deres til tider konfliktuerende krav altid er til stede i den enkelte enhed, dog i forskellig indbyrdes vægtning. Ikke mindst i lyset af offentlige, moderniserede velfærdsorganisationers styringspositionering rejser det ofte en tydelig diskrepans og konflikt mellem ledelsesopgaven opadtil (ledelse i styringsperspektivet) og opgaven indadtil, det vil sige ledelse af den praktiske opgaveløsning.

Opadtil – det vil sige i styringsperspektivet – betyder præstationskontrol i sig selv en afgørende reduktion af ledelsesopgavens kompleksitet og en potensering af kontrolfunktionen på baggrund af den indikatorfrembragte standard. Det er dog samtidig en udblænding af så væsentlige elementer i såkaldt human processing, at det – med Kaspersen og Nørgaard (2015) – giver anledning til at tale om erosion af den substantielle ledelsesopgave. Det fører til problemstillingen ledelse indadtil:

Ledelsesopgaven indadtil – i forhold til den praktiske opgaveløsning – synes derimod omvendt 'kompleksiceret'. I forlængelse af de risici, vi har påpeget tidligere i artiklen, stilles der for os at se krav til ledelsens evne til buffering (se fx . Hatch, 2001, s. 108 ff) – det vil sige til en vis grad at kunne afskærme organisationens kerneprocesser for den usikkerhed, som kpi-teknologiens turbolenser forårsager. Nils Brunssons arbejde med måder, hvorpå konfliktende krav organisatorisk mødes, vil her have relevans. I stedet for at mindske afstanden mellem handleorganisationens og den politisk organisations logikker taler Brunsson om løskobling eller afkobling som (nødvendig) strategi (Brunsson, 2003). I vores tilfælde gælder det løskobling med henblik på at undgå den skade, som en beslutning taget i den umiddelbare hensigt at rette op på en dårlig kpi vil kunne gøre. En vis løskobling er et genkommende tema i rektorinterviewene.

Det fører os frem til to afsluttende overvejelser: Én overvejelse, der er rettet mod institutionsstyringsniveauet og én overvejelse rettet mod ledelse indadtil.

Den første overvejelse, vedrørende styringen, gælder den konflikt og det pres, der kan opstå, når accountability møder responsibility. Dette dobbelte ansvar er som nævnt et vilkår, der knytter sig uadskilleligt til professionelles opgaveløsning, når det gælder offentlige velfærdsopgaver. Men den dominans, accountability får gennem præstationsevaluering, er nu voldsomt forøget, og med det før citerede fra Albæk et al. (2001) in mente er der ingen grund til at tro, at et sådant pres vil forsvinde. Det får os til at overveje, om man ikke i stedet kunne se præstationsmålingen som et input til en fælles undersøgende diskussion af, hvordan kvalitet udfolder sig i praksis i spændet mellem den nationale og lokale mission? Det ville gøre det muligt at trække de aspekter af gymnasiernes samlede uddannelsesopgave ind, sådan som rent faktisk beskrives i styringsdokumenterne, men som bliver usynlig i kpi'erne. I stedet for en ordre-adlydekommunikation vil det fremhæve et organisatorisk læringselement, der stimulerer skolerens kvalitetsarbejde, men også give styringsniveauet indsigt i kvalitetsens kompleksitet.

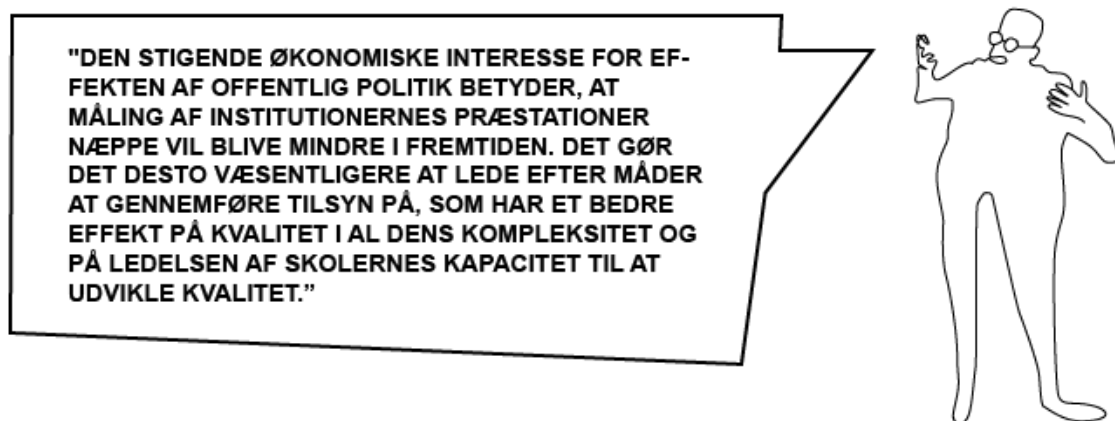
Den anden overvejelse gælder ledelse indadtil. Hvis nu præstationsmålinger i stedet for endgyldige resultater opfattes som forstyrrelser af etablerede praksisser og som invitation til undersøgelser og forbedringer af samme, vil man kunne inddrage det aspekt med fra KUP-forløbet, hvor kvalitet betragtes som en proces – et begreb i stadig udvikling i spændingen mellem den politiske vilje, det lokale behov og den forekommende praksis. Det blev tilgodeset i det dialogiske KUP-sammenhæng og må antages at være en forudsætning for at undgå den mistillid, der truer kpi-systemet og dermed som forudsætning for meningsfuld ledelse indadtil. Samtidig diskvalificeres kpi'er som elementer i et ledelsesinformationssystem ikke uden videre, med afgørende er, at de kritiske data bruges som afsæt til projekter om didaktisk udvikling – for først i og med, at der handles, skabes den kontekst, der gør, at man kan få øje på egen praksis (Huberman, 1997, se også Weick, 1979). For organisatorisk at kunne udnytte data som invitation til didaktisk udforskning, må rektor styrke det lag i organisationer, der (formaliseret eller uformaliseret) er i stand til at lede en løbende didaktiske udviklingsaktivitet. Det bør ikke ske i form af projekter, hvis succeskriterie alene er at rette op på 'dårlige tal' – målene er forbedret undervisning, som knytter an til en lokal forestilling om god didaktisk kvalitet, men

selvfølgelig med et sideblik til, at projektet virkning også kan aflæses på de problematiserede kpi'er.

Konklusion

I artiklen har vi dels ved en mindre, empirisk undersøgelse blandt gymnasierektorer undersøgt oplevelsen af, hvorvidt den ministerielle tilsyn virker som den motor i kvalitetsudvikling, tilsynet har intention om at være. Selv om det tilsyneladende ikke er feltets oplevelse at det generelt er tilfældet, kan vi ikke afvise, at det kan forekomme. Dels har vi undersøgt teknologien, der anvendes, den såkaldte præstationsevaluering. I den forbindelse har vi peget på en række risici, der i et vist omfang også bekræftes af de oplevelser, empirien fremstiller. I fortsættelse heraf har vi set på den lederpositionering, tilsynet implicerer, og peget på risikoen for, at væsentlige elementer af den uddannelsesopgave, som uddannelsens styringsdokumenter udpeger, udblændes. Det får os slutteligt at foreslå en ændret tilgang til kvalitetsarbejde i relationen opadtil (styrelse og gymnasieledelse) og indadtil (i forhold til den praktiske, didaktiske kvalitetsudvikling).

Vi citerer Albæk et al. (2001) for en vurdering af, at den stigende økonomiske interesse for effekten af offentlig politik betyder, at måling af institutionernes præstationer næppe vil blive mindre i fremtiden. Det gør det desto væsentligere at lede efter måder at gennemføre tilsyn på, som har en bedre effekt på kvalitet i al dens kompleksitet og på ledelsen af skolernes kapacitet til at udvikle kvalitet.



Litteratur

Albæk, E., & Rieper, O. (2001). Evaluering i Danmark: Effektevaluering, monitorering og formativ evaluering. I P. Dahler-Larsen, & H. Kroghstrup, *Tendenser i evaluering*. Odense: Odense Universitetsforlag.

Albæk, E., Dahler-Larsen, P., Hansen, H., Kroghstrup, H., Larsen, F., Lassen, M., & Winter, S. (2001). Evalueringers fremtid- fremtidens evalueringer. I P. Dahler-Larsen, & H. Kroghstrup, *Tendenser i Evaluering*. Odense: Odense Universitetsforlag.

- Andersen, V. (2013). Evaluering som omdrejningspunkt. I K. Hjort, Q. Ane, & P. Raae, *Der styres for vildt. Om paradokser i styringen af pædagogik*. Aarhus: Klim.
- Antonsen, M., & Jørgensen, T. (1999). Den offentlige kontekst: Begreber og Modeller. I T. Jørgensen, & P. Melander, *Livet i offentlige organisationer. Institutionsdrift i spændingsfeltet mellem stat, profession og marked*. København: Jurist- og Økonomforbundets Forlag.
- Arfwedson, G. (2001). *Nyare forskning om lärare. Presentation och kritisk analys av huvudlinjer i de senaste decenniernas engelskspråkiga lärarforskning*. Stockholm: HLS Förlag.
- Brunsson, N. (2003). Organized Hypocrisy. I B. Czarniawska, & G. Sevón, *The Northern Lights. Organization Theory in Scandinavia*. Copenhagen Business Schools: Copenhagen.
- Bundsgaard, J., & Puck, M. (2016). *Nationale test: Danske lærere og skolelederes brug, holdninger og viden*. Århus: DPU, Århus Universitet, UCL.
- Cave, M., Hanney, S., Henkel, M., & Kogan, M. (3. udg 1997). *The Use of Performance Indicators in Higher Education*. London: Jessica Kingsley Publishers.
- Christensen, T., & Lægreid, P. (2007). Introduction- Theoretical Approach and Research Question. I T. Christensen, & P. Lægreid, *Transcending New Public Management: The transformation of public reforms*. Surrey: Ashgate.
- Dahler-Larsen, P., & Kroghstrup, H. (2001). Evalueringens konstitutive virkninger. I P. Dahler-Larsen, & H. Kroghstrup, *Tendenser i evaluering*. Odense: Odense Universitetsforlag.
- Dahler-Larsen, P. (2001). Monitorering. I P. Dahler-Larsen, & H. Kroghstrup, *Tendenser i evaluering*. Odense: Odense Univesitetsforlag.
- Dahler-Larsen, P., & Larsen, F. (2001). Anvendelse af evaluering. Historien om et begreb, der udvikler sig. I P. Dahler-Larsen, & H. Kroghstrup, *Tendenser i evaluering*. Odense: Odense Universitetsforlag.
- Dean, M. (2006). *Governmentality. Magt og styring i det moderne samfund*. Frederiksberg C: Forlaget Sociologi.
- Freidson, E. (1986). *Professional Powers. A Study of the Institutionalization of Formal Knowledge*. Chicago: The University of Chicago Press.
- Folketinget 1999: Lov om Danmarks
Evalueringinstitut, <https://www.retsinformation.dk/eli/lta/2019/600>
- Folketinget 2002: Lov om åbenhed og gennemsigtighed i uddannelserne
m.v., <https://www.retsinformation.dk/eli/lta/2002/414>
- Folketinget 2005: Bekendtgørelse om kvalitetsudvikling og resultatvurdering
inden for de gymnasiale uddannelser <https://www.retsinformation.dk/eli/lta/2005/23>

- Harmon, G., & Mayer, R. (1986). *Organization Theory for Public Administration*. Glenview & London: Scott Foresman and Company.
- Haue, H., Nørr, E., & Skovgaard-Petersen, V. (1998). *Kvalitetens vogter. Statens tilsyn med gymnasieskolerne 1848-1998*. København: Undervisningsministeriet.
- Hjort, K., & Raae, P. (2014). *Velfærdsledelse i gymnasiet. Hvorfor og hvordan*. København: Gymnasieskolernes Rektorforening. http://issuu.com/rektorforeningen/docs/velf_rdsledelse_web.
- Haarder, B. (1988). Perestrojka i Uddannelsessystemet. *Uddannelse nr. 8*.
- Huberman, M. (1997). Critical Introduction. I M. Fullan (Red.), *Successful School Improvement*. Buckingham: Open University Press.
- Jensen, I., & Prahl, A. (2000). Competence som intersubjektivt fænomen. I T. Andersen, I. Jensen, & A. Prahl (Red.), *Kompetence som organisatorisk fænomen*. Roskilde: Roskilde Universitetsforlag.
- Kaspersen, L., & Nørgaard, J. (2015). *Ledelseskriser i konkurrencestaten*. København: Hans Reitzels Forlag.
- Kereluik, K., Mishra, P., Fahnoe, C., & Terry, L. (2013). What Knowledge is Most Worth: Teacher Knowledge for 21st Century Skills. *Journal of Digital Learning in Teacher Education Vol. 29 no. 4*.
- Kjær, P., & Pedersen, O. (2001). Translating Liberalization: Neoliberalism in the Danish Negotiated Economy. I J. Campbell, & O. Pedersen (Red.), *The Rise of Neoliberalism and Institutional Analysis*. Princeton: Princeton University Press.
- Krogstrup, H. K. (2006). *Evalueringsmodeller*. Århus: Academica.
- Larsen, F., & Lassen, M. (2001). Evaluering set i et policy-perpektiv. I P. Dahler-Lasern, & H. Krogstrup, *Tendenser i evaluering*. Odense: Odense Universitetsforlag.
- Lindholm, M. (. (1995). *Det skal Danmark leve af. En nyvurdering af Danmarks udviklingsevne*. København: Mandag Morgen Strategisk Forum.
- Lipsky, M. (1980). *Street-Level Bureaucracy. Dilemmas of the Individual in Public Services*. New York: Russel Stage Foundation.
- Newcomer, K. (1997). Using Performance Measurement to Improve Programs. *New Directions for Evaluation No. 75*.
- Norman, R. (2001). Letting and making managers manage: the effect of control systems on management action in New Zealand's central government. *International Journal of Public Management* .

OECD. (1996). *Lifelong Learning for All. Meeting of the Education Committee of Ministerial Level 16.-17. January 1996*. OECD Pedersen, O. (2011). *Konkurrencestaten*. København: Hans Reitzel.

Pedersen, O. (2011). *Konkurrencestaten*. København: Hans Reitzel.

Perrin, B. (1998). Effective Use and Misuse of Performance Measurement. *American Journal of Evaluation*. Vol. 19 no. 3.

Ravn, J., Nygaard, C., & Kristensen, P. (2006). Agentteori til analyse af strategizing. I C. Nygaard (Red.), *Strategizing - kontekstuel virksomhedsteori*. Frederiksberg: Samfundslitteratur.

Rigsrevisionen (2021): Rigsrevisionens beretning afgivet til Folketinget med Statsrevisorernes bemærkninger Studenters overgang fra gymnasium til videregående uddannelse

Rothstein, B. (1994). *Vad bör staten göra? Om välfärdsstatens moraliska og politiska logik*. Stockholm: SNS Förlag.

Schmidt, L. H. (1999). Dannelse på ny. Om det socialanalytiske perspektiv på velfærdssamfundets dannelsesformer. *Dansk pædagogisk tidsskrift*, 1999, nr. 1.

STUK (2023): Tilsynsberetning 2022. <https://www.stukvum.dk/-/media/filer/stuk/pdf23/230704-tilsynsberetning-2022.pdf>

STUK (2024): Tilsynsplan 2024 <https://www.uvm.dk/-/media/filer/uvm/aktuelt/pdf24/feb/240223-tilsynsplan-2024.pdf>

Smirchich, L., & Morgan, G. (18/3 1982). Leadership: The Management of Meaning. *Journal of Applied Behavioural Science*.

Thompson, J., & Tuden, A. (1959). Strategies, structures and processes of organizational decision. I J. Thompson, P. Hammond, R. Hawkes, B. Junker, & A. Tuden, *Comparative Studies in Administration*. Pittsburg: University of Pittsburg.

Tienken, C., & Canton, D. (2009). National Curriculum Standards: Let's think it over. *AASA Journal of Scholarship and Practice* 6/3.

Undervisningsministeriet. (1993). *Tegn på kvalitet i gymnasier, på studenterkurser og på hf*. Undervisningsministeriet.

Undervisningsministeriet. (1997). *Standarder og Profiler. Kvalitetsudvikling og institutionsbeskrivelse for gymnasiet og hf*. Undervisningsministeriet.

Weick, K. (1979). *The Social Psychology of Organizing*. Reading, MA: Addison-Wesley.

Ziehe, T. (1989). *Ambivalenser og mangfoldighed*. København: Politisk Revy.

Ziehe, T., & Stubenrauch, H. (1990). *Ny ungdom og usædvanlige læreprocesser*. København: Unge Pædagoger.

[1] Haue et al. citerer også daværende direktør for Undervisningsministeriets gymnasieafdeling Uffe Gravers Pedersen for i et interview i anledning af udgivelsen at have sagt, at man "har forsøgt at undgå det, som regnedrengene rundt om i systemet har forsøgt, nemlig at reducere skolen og uddannelse til fem nemt kvantificerbare indikatorer" (Haue et al, 1998 s, 384).