

'Klart under middel, middel, over middel' – om datavisualiseringers betydning for synliggørelsen af folkeskolens kvalitet

Helene Friis Ratner

Grafiske visninger af data – datavisualiseringer – er ikke neutrale repræsentationer af data, men udgør i sig selv en fortolkning. Denne artikel undersøger Undervisningsministeriets visualiseringer af nationale test med henblik på at analysere, hvordan forskellige visualiseringer skaber folkeskolens kvalitet som et fænomen, der kan gøres til genstand for pædagogisk og styringsmæssig intervention. Teoretisk anvendes begreberne ”infrastrukturel inversion” og ”konfiguration” fra Science and Technology Studies (STS). Begreberne giver blik for, hvordan tekniske entiteter som visualiseringstype konfigurerer folkeskolens kvalitet ved enten at fremhæve elevdygtighed eller elevlæring som et udtryk herfor. Datavisualiseringer indebærer således en *infrastrukturel politik* og er ikke blot en neutral teknisk baggrund for politiske og pædagogiske kampe om ”folkeskolens kvalitet”. Dette har implikationer for lærerens rolle som dataanalytiker i en tid, hvor læring og databrug i stigende grad erstatter et historisk fokus på undervisning.

Datavisualiseringer, nationale test, infrastruktur, testskala, kvalitet, lærerrolle

Klart under middel, Middel og Over middel. Dette er et udsnit af de tilbagemeldinger, som elever, deres forældre og lærere i folkeskolen kan få, når de har gennemført nationale test. De nationale test er computerbaserede, adaptive og selvrettende tests i folkeskolens 2.-8. klasse. Det er samtidigt et af de mest omstridte store datasæt i den danske folkeskole, og i skrivende stund afventer vi Undervisningsministeriets udvikling af et nyt evaluering- og bedømmelsessystem.

Selvom nationale test har nydt stor opmærksomhed, så har betydningen af, hvordan testresultater visualiseres grafisk, ikke været genstand for diskussion. Hvordan data visualiseres er dog særdeles vigtigt, da lærere tilgår nationale testdata gennem visualiseringer, og visualiseringerne i sig selv udgør en fortolkning af data. Datavisualiseringer har altså stor betydning for, hvordan nationale test gør et fænomen som folkeskolens kvalitet synligt. Som er tilfældet med andre kvantitative indikatorer, er visualiseringer af nationale test ikke en neutral repræsentation af skolens kvalitet, men bidrager med at skabe den som genstand for pædagogiske og styringsmæssige interventioner (Dahler-Larsen, 2014). På den måde giver nationale test det ellers svært definerbare begreb

”kvalitet” et særligt udtryk og kan virke *konstituerende for*, hvordan vi forstår og intervernerer i denne. Spørgsmålet om, hvordan visualiseringer af data gør kvalitet synlig, er således både relevant for de aktuelle nationale test og for det fremtidige bedømmelses- og evalueringssystem.

I denne artikel undersøger jeg, hvordan skala- og visualiseringstyper skaber et fortolkningsrum for lærerens forståelse af kvalitet. Selvom kvantitative målinger ofte sigter mod at skabe gennemsigtighed, er deres visualiseringer aldrig neutrale, og forskere har karakteriseret datavisualiseringer som værdiladede fortællinger, der privilegerer nogle synspunkter på bekostning af andre¹ (Kennedy et al., 2016). Når vi har med kvantitative data at gøre, har skalaer ydermere stor betydning for, hvilke statistiske beregninger man kan udføre, og hvordan resultaterne kan vises (Berthelsen & Østergaard, 2017). Derfor er de vigtige at inddrage i analyse. Artiklen har ikke til formål at afskrække den pædagogiske brug af datavisualiseringer. Visualiseringer kan formidle mønstre i data og gøre data tilgængelige for fortolkning. De kan med fordel indgå *i samspil med* andre former for viden, erfaring og data. Ved at forholde os analytisk til datavisualiseringer kan vi begynde at diskutere de synspunkter, værdier og fortællinger, der er indlejret i visualiseringer. Det skaber mulighed for, at vi kan reflektere kritisk over, hvad de gør synligt og usynligt, og hvad dette betyder for lærerens fortolkningsrum.

Jeg indleder artiklen med en introduktion til analysens teoretiske ressourcer, som hentes fra de såkaldte Science and Technology Studies (STS). Med en analytisk sensitivitet over for materialitetens rolle i fænomeners tilblivelse er STS særligt velegnet til at undersøge, hvordan tekniske anordninger som testskalaer og visualiseringstyper medvirker til at konstruere et fænomen som kvalitet. Konkret anvender jeg begreberne ”infrastrukturel inversion” og ”konfiguration” til at få blik for materialitetens rolle heri. Disse begreber introduceres i næste afsnit, der også redegør for artiklens empiriske grundlag. Herefter følger to analytiske nedslag, som tager udgangspunkt i de nationale tests to skalaer og deres respektive visualiseringer: Normskalaen, der anvender benchmarking til at afrapportere data, og kriterieskalaen, der kan vise elevs progression. Analysen viser, hvordan de to visualiseringer konstruerer kvalitet som et spørgsmål om henholdsvis elevdygtighed (i form af høje testresultater) og elevlæring (i form af elevs progression). Afslutningsvist diskuteres, hvad dette betyder for aktuelle diskussioner om ”lærficering” af uddannelse (Biesta, 2009) og for lærerens rolle i en stadig mere digitaliseret skole.

¹ *Hvordan nationale test skaber kvalitet afhænger naturligvis ikke kun af visualiseringer, men også af fx hvilke færdigheder, elever testes (eller ikke testes) i. Nationale test er fx kritiseret for primært at teste tekniske færdigheder af læsning og dermed ikke dække alle de færdigheder, der er i Fælles Mål (Bundsgaard, 2018).*

Infrastrukturel inversion: Et analytisk blik for teknologiernes politik

Som analysestrategi er ”infrastrukturel inversion” særligt velegnet til at undersøge, hvordan informationsteknologi virker ind i vores sociale og politiske virkeligheder (Bowker og Star, 1994). Infrastrukturel inversion implicerer, at man gør op med en grundlæggende skelnen mellem ”det teknologiske” og ”det sociale” (Callon, 1986). I stedet for at antage, at (informations-)teknologi og dens sociale verden er adskilte størrelser, undersøger man med infrastrukturel inversion, hvordan mennesker og teknologi gensidigt påvirker hinanden. Hermed får vi som noget nyt øje på den agens, eller handlekraft, som teknologi har, og hvordan denne virker ind i den kontinuerlige skabelse af vores samfund.

Som ”inversion” angiver, så bytter begrebet ”infrastrukturel inversion” om på tekst og kontekst. I stedet for at lade ”konteksten” (fx ideologi, politiske reformer eller overordnede logikker) forklare infrastrukturen (som i dette tilfælde er nationale test), så undersøger man politiske, organisatoriske og sociale virkeligheder *med udgangspunkt* i infrastrukturen. Det betyder, at vi ikke kan reducere nationale test til en styringsfilosofi som New Public Management eller konkurrencestat. Infrastrukturel inversion åbner i stedet op for at forstå, hvordan klassificeringer, normer og standarder former vor tids digitaliserede og datainformerede praksisser. Dermed er politik og teknologi ikke adskilte men sammenvævede og gensidigt påvirkende størrelser.

Lucy Suchman (2007) har udviklet begrebet ”menneske-maskine konfiguration” til at studere sammenvævning af forskellige typer agens i socio-tekniske systemer. ”Konfiguration” betegner en relationel organisering, der er bygget ind i maskinens design. Denne organisering tilskriver og delegerer agens til henholdsvis menneske og maskine (Suchman, 2012, s. 49). At nationale test er computerbaserede, adaptive og selvrettende indebærer fx, at en række af lærerens tidligere opgaver er automatiserede. Læreren deltager fx hverken i udvælgelse af testopgaver (items) eller i bedømmelse af elevpræstationer, da testen automatisk beregner elevens dygtighed efter et endt testforløb. Til gengæld har læreren fået nye opgaver såsom afvikling af testforløb (Maguire, 2019) samt didaktisk fortolkning og anvendelse af testresultater.

I denne artikel anvender jeg konfigurationsbegrebet til at spørge, hvordan datavisualiseringer *definerer, muliggør og begrænser* lærerens mulige fortolkninger af nationale testdata. Hermed kan jeg undersøge, hvordan visualiseringer skaber (forskellige) genstande for pædagogisk intervention. Når jeg i analysen bruger begrebet ”infrastrukturel politik”, er det netop for at fremhæve, at nationale tests politik *også* produceres af infrastrukturen. Infrastruktur (databaser, standarder, algoritmer) er altså ikke et neutralt fundament, oven på hvilket politik foregår, men derimod et aktivt element, der virker ind i det levede liv. At nationale test også bruges som performance indikator i kommunernes mål- og

resultatstyring af folkeskoler understreger, at de ikke er teknikaliteter. I Peter Dahler-Larsens ord er performance indikatorer ”i sandhed politiske fordi de definerer kategorier, som er kollektivt betydningsfulde i samfundet” (Dahler-Larsen, 2014, s. 976, egen oversættelse). I dette tilfælde er der tale om definitionen af folkeskolens kvalitet. Heri spiller datavisualiseringer en vigtig rolle, idet de omsætter store mængder data til ”fortolkelige billeder af den sociale verden, som organiserer brugerens opmærksomhed og giver dem brugbare empiriske sensitiviteter” (Madsen, 2013, s. 11, egen oversættelse). Som jeg vil vise, har forskellige visualiseringer dog betydning for, hvordan kvalitet konstrueres som et iagttageligt og intervenserbart fænomen.

Visualiseringer er kun sidste led i en lang kæde af processer, der producerer, sorterer og laver beregninger på data. Nationale tests datavisualiseringer er betinget af andre elementer som item databaser, testens måleegenskaber, test software samt lærernes og elevernes testpraksisser (Bundsgaard og Kreiner, 2019; Kousholt, 2015, 2016; Maguire, 2019). Disse går forud for selve visualiseringen af data og ligger uden for denne artikels fokus. Hertil er det naturligvis vigtigt at indskyde, at visualiseringer aldrig *determinerer* lærernes fortolkninger, og at lærere kan misforstå eller bruge dem på andre måder end intenderet (Bundsgaard og Puck, 2016). Lærernes konkrete brug af datavisualiseringer ligger også uden for denne artikels fokus.

Dokumentstudie

Artiklens analyser er baseret på et dokumentstudie af nationale test fra 2006-2019. Dokumentarkivet omfatter Undervisningsministeriets hjemmesider omhandlende nationale test samt bekendtgørelser, brugervejledninger, rapporter (fx Den nationale præstationsprofil), folketingsdebatter og andre politiske dokumenter (fx politisk aftale vedr. folkeskolereform), debatter i tidsskriftet Folkeskolen, private virksomheders brug af nationale test (fx NordicMetrics) samt eksterne rapporter (fx Damvads PISA-relatering af de kriteriebaserede nationale tests). Jeg har endvidere anvendt ”wayback machine” (<http://web.archive.org/>), en internetarkiveringsside, til at fremsøge ministeriets tidligere vejledninger til testsystemet, herunder vejledninger til deres visualiseringer af nationale tests, for at kunne følge udviklingen i visualiseringer siden første vejledning i 2010.

Da analysen fokuserer på udvalgte aktuelle datavisualiseringer af nationale tests, danner størstedelen af dette materiale en kontekstuel baggrund for nærværende analyse (se dog Ratner *et al.*, 2019; Ratner, 2020 for relaterede forskningsresultater). I stedet for at afrapportere dokumentstudiet i bredden, fokuserer jeg i denne artikel på to forskellige visualiseringer med henblik på at vise, hvordan tekniske detaljer såsom skala og visualiseringstype har betydning for synliggørelse af kvalitet. Kriteriet for at vælge disse to eksempler er 1) at de anvender to forskellige skalaer (norm- og kriterieskala) og 2) anvender forskellige

visualiseringsformer (henholdsvis et søjlediagram, der benchmarker en klasse, og en tabel der visualiserer progression).² Ved at sammenligne forskellige visualiseringer bliver deres centrale rolle i skabelsen af kvalitet tydelig.

Hvad er nationale test?

Nationale test blev indført ved en ændring i folkeskoleloven i 2006 i kølvandet på en række internationale målinger af elevers læse- og regnefærdigheder, hvor danske elever klarede sig lavere end forventet. Det var dog først efter et OECD-review af det danske skolesystem i 2004, at nationale test kom på den politiske dagsorden (Egelund, 2008). Dette review anbefalede bl.a. ”færdighedsstandarder” og ”objektive evalueringer” (Ekholm et al., 2004, s. 129). Nationale test blev herefter introduceret med to formål: At udgøre 1) et pædagogisk redskab til lærere og 2) et nationalt monitoreringsværktøj til politikerne. Siden folkeskolereformen i 2014 har nationale test også været et centralt værktøj i mål- og resultatstyring af folkeskolen. Reformen indførte nationale resultatmål for folkeskolen, hvoraf flere måles via nationale test. Fx operationaliseres målsætningen om, at ”Folkeskolen skal udfordre alle elever, så de bliver så dygtige, de kan” gennem måltallet ”Mindst 80 procent af eleverne skal være gode til at læse og regne i de nationale test.” I forlængelse heraf opfordrer Undervisningsministeriet kommuner og skoler til at spørge, ”Hvordan (...) elevens og klassens resultater [er] sammenlignet med landets udvikling henimod de nationale resultatmål?” (Undervisningsministeriet, 2018b). Resultatmål medfører øget *accountability*, hvor skoler og kommuner holdes ansvarlige for deres resultater i nationale test. Dette gør det ekstra vigtigt at undersøge de nationale tests datavisualiseringer og det fortolkningsrum, de skaber for skolens lærere og ledere.

Nationale test har været obligatoriske siden 2010, og de obligatoriske test omfatter læsning (2., 4. 6. og 8. klasse), matematik (3., 6. og 8. klasse), engelsk (4. og 8. klasse) og fysik/kemi (8. klasse). Dertil kan kommuner og skoler anvende de obligatoriske test som frivillige nationale test en gang før og efter de obligatoriske test. Alle test måler tre profilmråder (fx sprogforståelse, afkodning og tekstforståelse i læsning). Nationale test måler elevers færdigheder via den psykometriske Rasch-model, også kendt som item response theory. Rasch er en sandsynlighedsbaseret testtilgang, der beregner sandsynligheden for, at elever med en given dygtighed vil besvare en opgave (item) med en given sværhedsgrad korrekt (Kreiner, 2009). At testen er adaptiv betyder, at elever gennemgår forskellige testforløb. En adaptiv algoritme udvælger løbende nye test items baseret på elevens beregnede dygtighed, således at højere dygtighed medfører sværere opgaver, mens en lavere dygtighed medfører lettere opgaver. Hvor dette i princippet skaber mere retvisende testresultater, fordi testen løbende tilpasses den enkelte elevs estimerede dygtighed, har det også medført, at elevers testforløb bliver sværere for lærere at gennemskue, da de skal klikke sig ind på hver enkelt elevs unikke testforløb for at få adgang til opgaver.

² Det er også muligt at lave en benchmark visning af de kriteriebaserede testresultater.

Rasch måler elevfærdigheder på en logit-skala, men af formidlingsmæssige hensyn har Undervisningsministeriet valgt at vise elevresultater ved hjælp af to andre skalaer: en normskala og en kriterieskala (se Ratner, 2020 for analyse af denne udvikling).³ Det er dem og deres tilhørende visualiseringer, som jeg vil fokusere på i de følgende analyser med det formål at undersøge, hvordan de konfigurerer kvalitet samt hvilke fortolkningsrum dette skaber.

Normvisualiseringens konfiguration af den forventelige og acceptable præstation

Da nationale test blev obligatoriske i 2010 valgte Undervisningsministeriet at afrapportere elevresultater på en normskala. Normskalaen består, som det også fremgår af figur 1, af to tilbagemeldinger. Kigger vi fx på den første (fiktive) elev Hampus Bachs resultat i profilområdet ”Tal og algebra”, bedømmes han både med et tal (98) og en kategori (klart over middel)⁴.

Navn	Tal og algebra Resultat 1-100	Geometri Resultat 1-100	Matematik i anvendelse Resultat 1-100	Samlet vurdering Resultat 1-100	Detaljer	Status
Hampus Bach	98 = Klart over middel	93 = Klart over middel	94 = Klart over middel	95 = Klart over middel	Se besvarelse	Gennemført
Kim Sørensen	47 = Middel	23 = Under middel	67 = Over middel	46 = Middel	Se besvarelse	Gennemført
Alger Dahl	---	---	---	---	Du mangler rettighed til at se resultatet	---
Alvin Dalgaard	---	---	---	---	Du mangler rettighed til at se resultatet	---

Figur 1: Grafisk visning på elevniveau fordelt på tre profilområder og som en samlet vurdering (Undervisningsministeriet, 2017a, s. 58).

Normskalaen er en selvreferentiel skala, hvor elever figurerer som et tal mellem 1-100 percentiler, og hvor deres placering refererer til deres rang i gruppen af elever, der tog testen i 2010. Det betyder, at hvis en elev får ”33”, så er denne mindst lige så dygtig som 33% af de elever, der tog testen i 2010. Normskalaen inddeler percentil-værdierne i fem kategorier: Klart under middel (1-10), Under middel (11-35), Middel (36-65), Over middel (66-90), Klart over middel (91-100) (Undervisningsministeriet, 2018b). Det er altså selve fordelingen og rangeringen af elever på en skala fra 1-100, der producerer normskalaen, og

³ Logit-skalaen er en intervallskala (dvs. en skala, hvor afstanden mellem dens enheder er konstant) og går fra minus til plus uendeligt. Normskalaen er en ordinalskala (dvs. en skala, hvor resultater kan rangordnes, men hvor afstanden mellem enheder ikke er konstant). Som fem-trinsskala (fra ”Klart under middel” til ”Klart over middel”) er det en forenklet percentilskala, der går fra 1-100 point, og hvor eleverne er rangordnet ift. deres præstation i 100 grupper. Kriterieskalaen vurderer en elevs resultater ift. fagligt fastsatte kriterier (her er det fælles mål) (Damvad, 2014; Ratner, 2020; Wandall, 2016).

⁴ Kun kategorien formidles til elev og forældre.

den enkelte elevs præstation får sin værdi i kraft af dens position i forhold til andre elevers præstationer. Hermed konfigurerer normskalaen kvalitet som et spørgsmål om den beregnede elevdygtighed, der afrapporteres som en rangeret position.

Hvordan konfigurerer den normbaserede visualisering så læreren til at fortolke elevens præstation? Og hvilken infrastrukturel politik implicerer en skala, hvor elever scores i forhold til deres distance fra skalaens middelværdier?

Kategoriernes numeriske værdier er fastlagt, så de matcher ECTS karakterskalaen (European Credit Transfer System grading scale⁵), der igen bygger på den statistiske normalfordeling. I en normalfordeling vil de fleste resultater ligge omkring midten (50) og de færreste i skalaens to yderpunkter, hvilket giver normalfordelingen den velkendte klokkeform, der er høj i midten og flader symmetrisk ud i ”enderne”. Normen fastlægges således af skalaens middelværdier (36.-65. percentil).⁶ Normskalaen gør elevpræstationer synlige i forhold til, hvor tætte de er på disse middelværdier, hvilket de fem kategorier også afspejler med deres reference til (klart) over/under middel. Denne skabelse af middelprestationen som det normale har således en ”normaliserende” virkning, idet den skaber forventninger til, hvad der udgør den normale og dermed acceptable præstation.

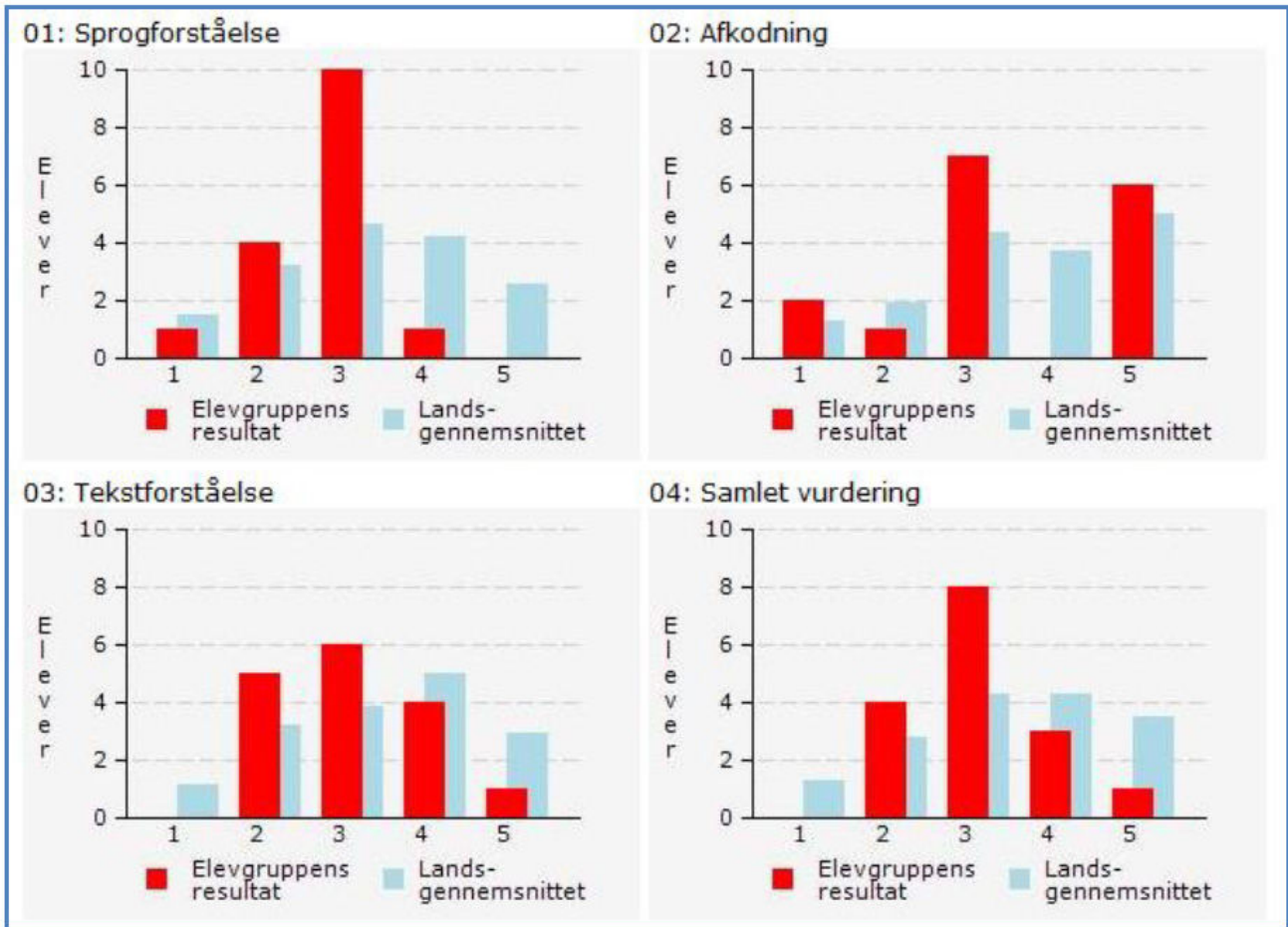
Den elevbaserede normvisning (figur 1) gør først og fremmest elevers præstationer synlige og sammenlignelige. Denne sammenlignelighed konfigurerer læreren til at undersøge, hvordan elever fordeler sig i de fem kategorier og måske endda rangordne elever i forhold til deres percentiler (den numeriske tilbagemelding), da elever både kan være i bunden og toppen af en kategori⁷. Visualiseringen konfigurerer yderligere læreren til at fortolke elevresultater omkring middelværdierne som ”normale” og andre resultater som positive eller negative afvigelser fra denne norm.

Det er dog ikke kun normskalaen, der har en normaliserende virkning. En overblikvisning (figur 2) giver læreren et supplerende sammenligningsgrundlag. Her vises elevernes fordeling på tværs af de fem kategorier i forhold til det aktuelle landsgennemsnit.

⁵ ECTS-skalaen, et ”overførbart kreditsystem”, blev udviklet som led i Bologna processen med henblik på at skabe sammenlignelighed og overførbare på tværs af europæiske karakterskalaer og dermed øge studentermobilitet.

⁶ Testsystemet anvender den første obligatoriske måling (2010) som normskala. Det betyder, at eleverne scores i forhold til 2010-fordelingen, selvom middelværdien er forbedret væsentligt siden 2010. Fx har middelværdien for 8. Klasses læseforståelse ændret sig fra 50. percentil til 61. percentil (Undervisningsministeriet, 2018a).

⁷ Dog vil den statistiske usikkerhed ofte være så stor, at variationer i få percentiler ikke er udtryk for en forskel i elevdygtighed.



Figur 2: Grafisk oversigtsvisning af normbaserede resultater. De røde søjler viser elevgruppens resultat på de tre profilområder og som en samlet vurdering. De blå søjler viser landsgennemsnittet fra seneste gennemførte obligatoriske test (Undervisningsministeriet, 2017a, s. 51).

I modsætning til den elevbaserede visning (figur 1) har vi ikke elevernes navne med her. I stedet fremtoner eleverne som en homogen "elevgruppe" (det vil typisk være en klasse) i røde søjler, sammenlignelige med det aktuelle landsgennemsnit, der figurerer som blå søjler. Denne visualisering skaber en ny form for normalisering: Det er ikke blot skalaens kategorier, men det nuværende landsgennemsnit, de blå søjler, der manifesterer, hvad der er den normale og acceptable elevdygtighed i en klasse. Denne type sammenligning kan tilskynde til at ville performe bedre end landsgennemsnittet (cf. Staunæs, 2018; Kousholt & Andreasen, 2015).

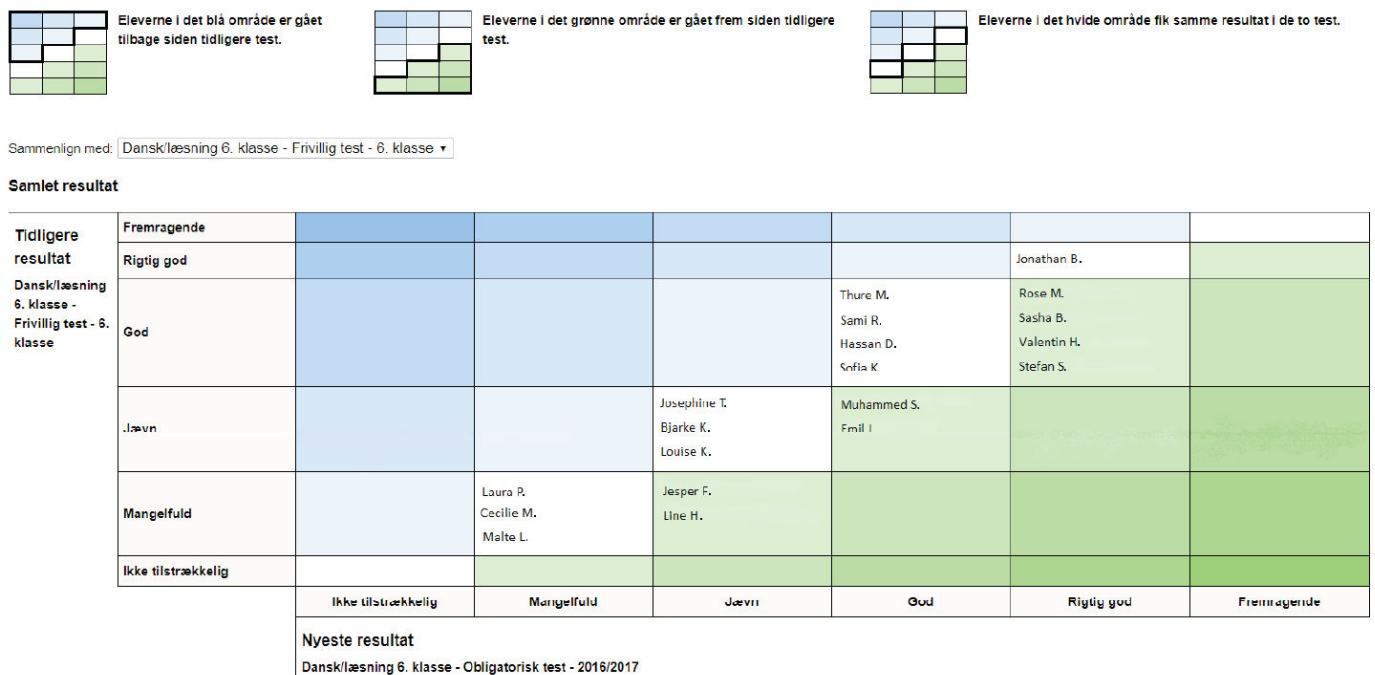
De normbaserede visualiseringers infrastrukturelle politik indebærer således en forståelse af kvalitet som elevdygtighed. Elevdygtighed visualiseres som værende på samme niveau som, over eller under landsgennemsnittet. Ved at forstå elever – og klassen – i forhold til en norm omkring elevdygtighed, producerer visualiseringen dog ikke kun en forståelse af status quo. Det fordrer samtidig overvejelser omkring, hvordan man kan forbedre præstationer hos den enkelte

elev eller klassen fremover. Som andre teknikker til at benchmarke organisationer kan den normbaserede visning tilskynde lærere til at forbedre testresultater, så man ligger på ”den rigtige” side af landsgennemsnittet (cf. Triantafillou, 2006).

Kriterievisualiseringens konfiguration af kvalitet som elevers progression

I forlængelse af folkeskolereformen (Regeringen *et al.*, 2013) introducerede Undervisningsministeriet en kriterieskala med nye resultatvisninger. Kriterieskalaen omfatter seks niveauer, rangerende fra ”Ikke tilstrækkelig præstation” til ”Fremragende præstation”, hvor hvert niveau er ”et udtryk for elevernes faglige niveau i de dele af fagene, som testes” (Undervisningsministeriet, 2017b). I modsætning til normskalaens niveauer, som er fastlagt i forhold til 2010-elevpopulationens spredning, udtrykker kriterieskalaen en forventning til, hvad elever skal kunne. Kriterierne er fastlagt med reference til Fælles Mål, PISA-skalaen og eksisterende testresultater på normskalaen (Ratner, 2020).

I det følgende vil jeg argumentere for, at en af kriterieskalaens visualiseringer konfigurerer kvalitet anderledes end normskalaen. Figur 3 er en kriteriebaseret resultatvisning med titlen ”Klassens progression”, og her kan læreren hurtigt danne sig et overblik over, hvilke elever der er gået frem eller tilbage siden sidste test.⁸



Figur 3: ”Klassens progression fra tidligere test” bruger farvekoder til at vise, hvilke elever, ”der fagligt er gået frem eller tilbage sammenlignet med tidligere test”, hvor blå farve indikerer et dårligere resultat på kriterieskalaen, hvid farve markerer samme resultat og grøn farve et bedre

8 Det er også muligt at få vist progression på individniveau.

resultat (Undervisningsministeriet, 2017c, s. 6).

”Klassens progression fra tidligere test” er en tabel, der formidler klassens samlede progression i en oversigtsvisning. Dens farvekoder konfigurerer læreren til at undersøge, hvordan eleverne har udviklet sig siden sidste test, fx hvilke elever har opnået et bedre eller dårligere resultat siden sidste test. Visualiseringen gør det først og fremmest muligt for læreren at forholde sig til en enkelt elevs udvikling. Hvor visualisering af elevdygtighed på normskalaen sker i relation til skalaens middelværdier, som udtrykker den gennemsnitlige præstation, så folder denne visualisering sammenligningen ind i eleven selv, hvor eleven splittes op i et tidligere og et nutidigt testresultat: Eleven *trackes* over tid og ”disaggreger[s] (...) i mønstre af adfærd eller præstation over tid” (Sellar & Thompson 2016, 493). Det vigtigste i visualiseringen er ikke skalaens niveauer, men farven, der indikerer, hvorvidt og hvordan de har flyttet sig på tværs af skalaens kategorier.

Visualiseringen konfigurerer således læreren til at undersøge, hvad der i undervisningen flytter elever mod et højere niveau, og dens farvekoder kan ”skabe og bevare en fornemmelse af progression” (Finn, 2016). Dette medfører nye idealer for både elev og klasse: Hvor normskalaen, med sin reference til den statistiske normalfordeling, skaber forventninger til at flest scorer i midten samt incitament til at score over middel, bliver *bevægelse mod et højere niveau* den nye markør for kvalitet med kriterieskalaens progressionsvisning. Progressionsvisningen konfigurerer med andre ord kvalitet som et spørgsmål om læring – og skolens løfteevne – frem for elevdygtighed. Vejledningen til den kriteriebaserede tilbagemelding anbefaler netop også læreren at bruge visualiseringen til at undersøge: ”Hvordan er elevens og klassens udvikling sammenlignet med tidligere test?” (Undervisningsministeriet, 2017c).

Den temporale sammenligning udgør ikke blot et ”før” (tidligere test) og ”efter” (aktuelle test), men konstituerer også et ”ikke endnu”. Som Matt Finn skriver: ”Data gør progression synlig (...) [ved at] indrullere og sammenholde elevers fortidige, nutidige og fremtidige selv som fortællinger om forbedringer og løfteevne” (2016, s. 32). Visualiseringen implicerer altså også en potentialitet (Hamre et al., 2015; Staunæs, 2011): Elever, der optræder i blå eller hvide kategorier, kan i fremtiden også blive grønne. Visualiseringen er således både individualiserende, i og med eleven får sin farvekodede værdi ift. et tidligere selv, og potentialiserende, da den installerer et håb om fremtidige grønne farver.

Synliggørelsen af elevers progression i forhold til hinanden betyder endvidere, at visualiseringen konfigurerer progression som et kollektivt fænomen: Hvor mange har undervisningen løftet mod det bedre? Igen er det vigtigste ikke deres præstation forstået som niveau eller rang, men netop deres bevægelse mod det bedre. Den grønne farve konfigurerer kvalitet som elevers læring mod det bedre – som et spørgsmål om løfteevne.

Datavisualiseringers konfigurationer af kvalitet og deres konsekvenser for lærerens rolle

Selve kvantificeringen og synliggørelsen af individers og institutioners præstationer kan motivere aktører til at forbedre deres præstationer (Brøgger, 2016). Derfor er det ikke ligegyldigt, hvordan data visualiseres, idet visualiseringen rammesætter brugerens opmærksomhed og dermed også hvad, der tæller som høj kvalitet. Dette har både konsekvenser for vores forståelse af samtidens skole, for vores forståelse kvalitet og for lærerens rolle.

For det første bidrager studiet af datavisualiseringer med en specificitet i den overordnede "Zeit-diagnose" om et større historisk skifte i uddannelse fra et fokus på undervisning til læring. Uddannelsesfilosoffen Gert Biesta (2009) taler om en "lærificering" ("learnification") af uddannelse, som indebærer en forskydning fra input (lærerens undervisning) til output (elevresultater). Biesta fokuserer på den sproglige transformation, men argumenterer for, at dette skifte er tæt forbundet med målinger og kvantificering af uddannelse. Denne iagttagelse peger mod vigtigheden af at studere disse målingers specificitet – og ikke mindst, hvordan de gør læring synlig. Finn ser fx brugen af data som udtryk for en ny lærerrolle: "Læreren bliver i mindre grad en informationskilde [der qua sin praksis kender til elevens dygtighed og læring] og i højere grad en dataproducent og –analytiker. (...) Læreren holdes ansvarlig for at producere læring" (2016, s. 44). Netop fordi datavisualiseringer allerede udgør en analyse og fortolkning af data, er det vigtigt ikke blot at forstå læreren som en data-analytiker – men netop at undersøge, hvordan selve fortolkningen og analysen af data er distribueret på tværs af lærer og testapparat i visualiseringernes konfigurationer. I denne artikel har jeg netop vist, at forskellige datavisualiseringer konfigurerer lærerens opmærksomhed på forskellige måder.

Der ligger altså en politik i infrastrukturelle komponenter som testskala og visualiseringstype, idet de gør kvalitet synlig på forskellig vis. Når visualiseringer konfigurerer kvalitet som henholdsvis elevdygtighed (normskalaens benchmarkvisning) og elevlæring (kriterieskalaens progressionsvisning), så understreger det, at skalaerne og visualiseringerne ikke er neutrale repræsentationer af data. Visualiseringernes politik ligger i, at de i sig selv udgør en fortolkning af data, og at fortolkningsarbejdet ikke er forbeholdt lærerens professionelle skøn. Datavisualiseringers fortolkning skabes bl.a. i designeres beslutninger om, hvilke data, der gøres sammenlignelige: Ved at sammenholde den individuelle elevs resultater med henholdsvis landsgennemsnittet (normvisningen) og egne tidligere testresultater (progressionsvisningen) skabes dermed også forskellige elevkonstruktioner: Eleven som højt præsterende versus eleven som lærende. Visualiseringer indebærer således en *infrastrukturel politik*. Testapparatet er ikke blot et teknisk bagtæppe eller et teknisk udtryk for sociale og politiske kampe om kvalitet. De går aktivt ind og intervernerer med læringsforståelser og evalueringspraksisser.

Konklusion

Datavisualiseringer – den grafiske formidling af data – er slutproduktet af en lang række aktiviteter, der omfatter produktionen af data (elevers interaktion med en computer, test items og algoritme) og beregningen af elevdygtighed. Disse processer er først og fremmest performative: De *skaber* elevdygtighed og elevlæring som fænomener, der kan sammenlignes og gøres til genstand for pædagogisk og styringsmæssig intervention. Således er de ikke neutrale *repræsentationer*, men har betydning for, hvad læring og dygtighed kan være, når folkeskolens kvalitet tematiseres med udgangspunkt i nationale test.

Ved at undersøge forskellige visualiseringer af data i nationale test har jeg demonstreret, at tekniske anordninger som testskala og visualiseringstype spiller en central rolle i testapparatets konstruktion af kvalitet. Hvor visualiseringer baseret på nationale tests normskala konfigurerer skolens kvalitet som et spørgsmål om høje versus lave præstationer, set ift. en elevpopulations middelværdier, værdisætter visualiseringer baseret på nationale tests kriterieskala elevers progression, dvs. deres bevægelse mod det bedre. Gennem begrebet *konfiguration* har jeg argumenteret for, at de forskellige visualiseringer organiserer lærerens opmærksomhed mod disse forskellige forståelser af kvalitet. Artiklens analyser viser hermed, at visualiseringer af store datasæt potentielt kan virke ind i pædagogiske forståelser af elevpræstationer ved at udgøre en selvstændig fortolkning af testdata. Den har også demonstreret, at samme data kan visualiseres på meget forskellige måder, med forskellige konfigurationer af lærerens opmærksomhed som konsekvens.

Jeg har i diskussionen udfoldet analysens implikationer. For det første har jeg argumenteret for, at de empiriske fund – som består i at samme data kan visualiseres på forskellige måder – bidrager med en vigtig specificitet i aktuelle diskussioner om uddannelsessystemets historiske skifte fra undervisning til læring. For det andet har jeg argumenteret for, at dette skifte ikke blot implicerer en ny lærerrolle, hvor læreren i højere grad er analytiker af data. Konfigurationsbegrebet og mine analyser af datavisualiseringer understreger netop, at den analyserende og fortolkende kompetence ikke er forbeholdt lærerens professionelle skøn, men i høj grad varetages af visualiseringssoftware, der gennem en æstetisk organisering af data tilbyder særlige fortolkninger af data.

Afslutningsvist er det vigtigt atter at understrege, at analysen kun har behandlet nationale tests datavisualiseringer, og derfor ikke kan konkludere på, hvordan de sætter sig igennem i praksis, når lærerne interagerer med testsystemets visualiseringer. Men da lærerne primært tilgår nationale testdata gennem dets visualiseringer, er det plausibelt at antage, at de spiller en væsentlig rolle i læreres fortolkningsprocesser. Videre empiriske studier kan med fordel undersøge, hvordan lærere og andre aktører (fx ledere, forældre og elever) i praksis

interagerer med de forskellige datavisualiseringer, og hvordan andre faktorer end de indlejrede fortolkningsmuligheder spiller ind (fx tid, ledelse, de nationale måltal, organisatoriske rutiner som klassekonferencer).

I en tid hvor skolens aktører i stigende grad konsumerer information om elevers læring og skolers kvalitet gennem datavisualiseringer, og hvor overbliksskabende datavisualiseringer er blevet almindelige anordninger i organisatoriske landskaber, er det af yderste vigtighed at skabe refleksion over datavisualiseringers indlejrede fortolkninger. Aktuelle diskussioner om læreres "data literacy" kan med fordel omfatte et fokus på datavisualiseringer, såvel deres muligheder som begrænsninger. Ved at skabe viden om datas begrænsninger, hvordan data produceres, hvordan de transformeres til grafiske visninger, samt hvilke fortællinger og synspunkter, der er indlejret i populære visualiseringstyper, kan vi begynde en vigtig kritisk diskussion om såvel konstruktiv som ukonstruktiv konsumtion af datavisualiseringer i hverdagens didaktiske og pædagogiske beslutninger.

Helene Friis Ratner, lektor (Ph.d.) på DPU - Danmarks Institut for Pædagogik og uddannelse, Aarhus Universitet.

Artiklen er skrevet med støtte fra en FKK-bevilling (Data Visions: Education in the age of digital data visualizations) fra Danmarks Frie Forskningsfond.

Litteratur

Berthelsen, U. D., & Østergaard, M. K. (2017). *Om learning analytics og datavisualisering: Introduktion og litteraturreview* (Data om Elevers Læring og Progression, s. 1–34). DPU, Aarhus Universitet.

Biesta, G. (2009). Good education in an age of measurement: On the need to reconnect with the question of purpose in education. *Educational Assessment, Evaluation and Accountability*, 21(1), 33–46.

Brøgger, K. (2016). The rule of mimetic desire in higher education: Governing through naming, shaming and faming. *British Journal of Sociology of Education*, 37(1), 72–91.

Bundsgaard, J. (2018). Pædagogisk brug af test. *Sakprosa*, 10(2), 1–40.

Bundsgaard, J. & Kreiner, S. (2019). *Undersøgelse af De Nationale Tests måleegenskaber*. Aarhus University Library. <https://doi.org/10.7146/aul.319.217>

Bundsgaard, J., & Puck, M. R. (2016). *Nationale test: Danske lærere og skolelederes brug, holdninger og viden*. DPU, Aarhus Universitet.

Callon, M. (1986). Some elements of a sociology of translation: Domestication of the scallops and the fishermen of St Brieuc Bay. In J. Law (Ed.), *Power, action and belief: A new sociology of knowledge?* (s. 196–223). Routledge.

Dahler-Larsen, P. (2014). Constitutive Effects of Performance Indicators: Getting beyond unintended consequences. *Public Management Review*, 16(7), 969–986.

Damvad (2014). PISA-relateret af de kriteriebaserede nationale test. Delrapport 1 – formidling af resultater. København: Damvad.

Egelund, N. (2008). The value of international comparative studies of achievement – a Danish perspective. *Assessment in Education: Principles, Policy & Practice*, 15(3), 245–251.

Ekholm, M., Mortimore, P., David-Evans, M., Laukkanen, R., & Valijarvi, J. (2004). *OECD-rapport om grundskolen i Danmark*. Undervisningsministeriet.

Finn, M. (2016). Atmospheres of progress in a data-based school. *Cultural Geographies*, 23(1), 29–49.

Hamre, B., Kousholt, K., Staunæs, D. & Krejsler, J. (2015). Eleverne, der skulle blive så dygtige, som de kunne – Kategorisering og potentialitet i skolens lærings- og testarrangementer. *Pædagogisk Psykologisk Tidsskrift*, 52(3), 21-43.

- Kennedy, H., Hill, R. L., Aiello, G., & Allen, W. (2016). The work that visualisation conventions do. *Information, Communication & Society*, 19(6), 715–735.
- Kousholt, K. (2015). Børns gætterier ved nationale test. CEPRAsriben. *Tidsskrift for Evaluering i Praksis*, 18, 46-57.
- Kousholt, K. (2016). Testing as social practice: Analysing testing in classes of young children from the children's perspective. *Theory & Psychology*, 26(3), 377–392.
- Kousholt, K. & Andreasen, K.E. (2015). De nationale test som ny praksis i den danske folkeskole – betydninger i klassens sociale fællesskab. I: K.E. Andreasen, M. Buchardt, A. Rasmussen & C. Ydesen (red.), *Test og prøvelser – Oprindelse, udvikling, aktualitet*. Aalborg: Aalborg Universitetsforlag. (s. 29-49).
- Kreiner, S. (2009). Om udvikling og afprøvning af pædagogiske test. In C. Bendixen & S. Kreiner (Eds.), *Test i folkeskolen* (s. 43–82). Hans Reitzels Forlag.
- Madsen, A. K. (2013). *Web-Visions: Repurposing digital traces to organize social attention* [Copenhagen Business School]. PhD Series 26.2013.
- Maguire, L. H. (2019). Adapting to the test: Performing algorithmic adaptivity in Danish schools. *Discourse: Studies in the Cultural Politics of Education*, 40(1), 78–92.
- Porter, T. M. (1994). Making Things Quantitative. *Science in Context*, 7(3), 389–407. <https://doi.org/10.1017/S0269889700001757>
- Power, M. (2015). How accounting begins: Object formation and the accretion of infrastructure. *Accounting, Organizations and Society*, 47, 43–55. <https://doi.org/10.1016/j.aos.2015.10.005>
- Ratner, H. (2020). Europeanizing the Danish School through National Testing: Standardized Assessment Scales and the Anticipation of Risky Populations. *Science, Technology, & Human Values*, 45(2), 212-234.
- Ratner, H., Andersen, B. L., & Madsen, S. R. (2019). Configuring the teacher as data user: Public-private sector mediations of national test data. *Learning, Media and Technology*, 44(1), 22–35.
- Regeringen, Venstre, & Dansk Folkeparti. (2013). *Aftale mellem regeringen (Socialdemokraterne, Radikale Venstre og Socialistisk Folkeparti), Venstre og Dansk Folkeparti om et fagligt løft af folkeskolen*.

Staunæs, D. (2011). Governing the potentials of life itself? Interrogating the promises in affective educational leadership. *Journal of Educational Administration and History*, 43(3), 227–247.

Staunæs, D. (2018). "Green with envy:" affects and gut feelings as an affirmative, immanent, and trans-corporeal critique of new motivational data visualizations. *International Journal of Qualitative Studies in Education*, 31(5), 409–421. <https://doi.org/10.1080/09518398.2018.1449983>

Suchman, L. (2007). *Human-Machine Reconfigurations: Plans and Situated Actions* (2nd ed.). Cambridge University Press.

Suchman, L. (2012). Configuration. In C. Lury & N. Wakeford (Eds.), *Inventive methods: The happening of the social* (s. 48–60). Routledge.

Triantafillou, P. (2006). Benchmarking som normaliserende styringsteknologi. *Politica*, 38(1), 22–39.

Undervisningsministeriet. (2017a). *Test- og prøvesystemet—De nationale test: Brugervejledning for skoler*. <https://www.stil.dk/support/test-og-proever/nationale-test>

Undervisningsministeriet. (2017b). *Vejledning om de nationale test—Til skoleledere*. Undervisningsministeriet. <https://www.uvm.dk/-/media/filer/uvm/udd/folke/pdf17/mar/170315-vejledning-til-skoleledere.pdf?la=da>

Undervisningsministeriet. (2017c). *Vejledning til nye resultatvisninger i de nationale test – til lærere i alle fag*. Undervisningsministeriet. <https://uvm.dk/folkeskolen/elevplaner-nationale-test-og-trivselsmaaling/nationale-test/vejledninger>

Undervisningsministeriet. (2018a). *Fælles Mål på læringsplatforme* (s. 1–12). Undervisningsministeriet. <https://uvm.dk/-/media/filer/uvm/udd/folke/pdf18/juni/180620-fm-laeringsplatforme-anbefalinger.pdf?la=da>

Undervisningsministeriet. (2018b, January 31). *Nationale test: Brugervejledninger til de nationale test*. <https://www.stil.dk/support/test-og-proever/nationale-test>

Undervisningsministeriet. (2018c, August 13). *Om Fælles Mål*. Om Fælles Mål. <https://uvm.dk/folkeskolen/fag-timetale-og-overgange/faelles-maal/om-faelles-maal>

Wandall, J. (2016). *Appendiks 2: Om normskalaer, kriterieskalaer og progression i de nationale test*. *NordicMetrics*. http://www.nordicmetrics.com/beregneren/vejl_mat_app2_v2-3-2/