



Data om elevernes læring og progression

En forskningskortlægning og syntese

JENS DIETRICHSON
KIRSTINE KARMSTEEN
BJØRN CHRISTIAN VIINHOLT NIELSEN
IDA SCHEEL RASMUSSEN
METTE AGERSKOV SMITH

KØBENHAVN 2016

DATA OM ELEVERNES LÆRING OG PROGRESSION
EN FORSKNINGSKORTLÆGNING OG SYNTSE
Afdelingsleder: Mette Deding
Afdelingen for Skole og Uddannelse

© 2016 SFI – Det Nationale Forskningscenter for Velfærd

SFI – Det Nationale Forskningscenter for Velfærd
Herluf Trolles Gade 11
1052 København K
Tlf. 33 48 08 00
sfi@sfi.dk
www.sfi.dk

SFI's publikationer kan frit citeres med tydelig angivelse af kilden.

SFI-notater skal danne grundlag for en faglig diskussion. SFI-notater er foreløbige resultater, og læseren bør derfor være opmærksom på, at de endelige resultater og fortolkninger fra projektet vil kunne afvige fra notatet.

INDHOLD

1	SAMMENFATNING	5
	Resultater	6
	Afsluttende kommentarer	9
2	INDLEDNING	11
	Formål og forskningsspørgsmål	11
3	FORSTÅELSER OG DEFINITIONER AF LÆRING OG PROGRESSION	15
	Perspektiver og holdninger	16
4	MÅLING AF ELEVENES LÆRING OG PROGRESSION	21
	Datakilder, skalaer og metodikker	21
	Hyppigheden af målinger	27
	Måling af tværgående kompetencer	30
	Måling af trivsel	36
	Hvad måles progression op imod?	39
	Visualisering	44

5	DET FAGPROFESSIONELLE ARBEJDE MED ELEVENS PROGRESSION	49
	Progressionsmåling som en del af det fagprofessionelle arbejde	49
	IT-understøttelse af det faglige arbejde med data	56
	Fagprofessionelles kompetencer til at fortolke data	59
	Kendetegn ved skoler og kommuner, som benytter progressionsmåling	64
6	DISKUSSION OG VIDERE PERSPEKTIVER	67
	LITTERATUR	73

SAMMENFATNING

Denne forskningskortlægning og -syntese indgår som en del af et større udviklings- og inspirationsprojekt, Data om elevernes læring og progression. Formålet med udviklings- og inspirationsprojektet er at give videns- og praksisbaserede anbefalinger til, hvordan data om elevernes progression kan indsamles, sammenstilles, vises og understøtte, at danske skoleelever bliver så dygtige, som de kan. Projektet skal illustrere og give anbefalinger til, hvordan skoler og kommuner kan bruge data til at beskrive og visualisere elevers progression, og hvordan pædagogisk personale kan anvende data i deres didaktiske arbejde for at styrke elevernes læring. Projektet skal have et bredt datafokus og bringe såvel centrale som de mange lokale data i anvendelse. Projektet skal vise muligheder for udveksling af data om progression mellem it-systemer, som anvendes i folkeskolens arbejde.

Denne kortlægning udgør, sammen med en erfaringsopsamling af praksis i danske skoler, første del af projektet. Forskningskortlægningen omhandler tre hovedtemaer:

1. Forståelser og definitioner af læring og progression
2. Måling af elevens læring og progression
3. Det fagprofessionelle arbejde med elevens progression.

Formålet med kortlægningen er at finde frem til relevante danske og internationale studier og beskrive de svar, som den tidligere litteratur har afgivet på forskningsspørgsmålene. Vi diskuterer her også, hvilke muligheder og udfordringer der kan være forbundet med at overføre de resultater, vi har fundet fra internationale studier, til en dansk kontekst, og hvilke typer af læring der er vanskelige at fremskaffe data omkring.

RESULTATER

Nedenfor følger en kort sammenfatning af de vigtigste resultater, opdelt efter kortlægningens tre hovedtemaer. Overordnet er det dog vigtigt at påpege to gennemgående aspekter ved de studier som vi har fundet, da det også påvirker, hvordan vores resultater skal fortolkes. For det første må vi konkludere, at der generelt mangler effektstudier på området, det vil sige, studier der – med eksperimentelle eller kvasi-eksperimentelle metoder – forsøger at afdække årsagssammenhænge. Vi kan således beskrive de erfaringer, holdninger og perspektiver, der er i litteraturen, men ofte ikke sige, hvad der virker bedst, eller hvad der har størst effekt. I resultatafsnittet nedenfor vil vi fremhæve det i teksten, hvis litteraturen, vi henviser til, indeholder effektstudier. For det andet er der en overrepræsentation af studier fra USA, som ikke nødvendigvis kan overføres til en dansk kontekst.

FORSTÅELSER OG DEFINITIONER AF LÆRING OG PROGRESSION

I forhold til forståelser af læring og progression ser vi flere, delvist modsatte, strømninger. I de fleste OECD-lande er der en tendens mod et større fokus på udviklingen af tværgående kompetencer, frem for et mere snævert fokus på isoleret viden og færdigheder. Det øgede fokus på socio-emotionelle kompetencer og inkluderingen i mange nationale pensum af de typer af færdigheder og kompetencer, som samles under begrebet det 21. århundredes kompetencer (*21st century skills*), kan ses som evidens for denne strømning. En anden strømning, vi ser, er et øget fokus på datainformeret skoleudvikling og herunder et gennemgående fokus på mere formative målinger. Modsat denne tendens til måling af bredere kompetencer og mere formative målinger er der dog samtidig også en tendens til, at der laves flere målinger, der er – eller opfattes – som kun summative og et øget fokus på *accountability* for skoler i forhold til fag-faglige resultater.

MÅLING AF ELEVERS LÆRING OG PROGRESSION

Dette tema omhandler spørgsmål lige fra datakilder, skalaer, hyppigheden af målinger, til tværgående kompetencer, visualisering af data og spørgsmålet omkring, hvad elevers progression skal måles op imod.

I spørgsmålet omkring datakilder og skalaer er en vigtig pointe, at man først og fremmest må gøre sig klart, hvad formålet er med en given måling, før man vælger, hvilken type data og skalaer, man vil anvende. Man kan overordnet skelne mellem eksterne standardiserede vurderinger (ofte mere summative) og interne lærerbaserede vurderinger (mere formative). Forskellige målinger er her ofte udviklet og designet med et bestemt formål, og anvendes de til andre formål, er der en risiko for, at

de vurderinger, man foretager på baggrund af målingen, ikke er valide. I dette udviklingsprojekt er fokus på lærere og pædagogers formative arbejde med data, hvorfor det her vil være vigtigt med et måledesign, der giver detaljerede informationer, som gør det muligt at identificere de enkelte elevers udfordringer.

Samlet set virker der til at være en sammenhæng mellem hyppigere *formative* målinger og bedre resultater. Desuden har flere effektstudier vist, at det at foretage enkle tests, som ligger tæt op ad det materiale, der undervises i, kan være en mere effektiv måde at lære på end det at repetere studiet af et givent undervisningsstof. Men selvom nogle studier indikerer, at flere målinger er bedre, er der naturligvis en grænse for, hvor ofte man bør teste. Hvor den grænse går, vides dog endnu ikke. Det er desuden svært at svare på, hvordan disse resultater om hyppigere målinger, som stammer fra amerikanske studier, kan overføres til en dansk kontekst, idet vi ikke har fundet nogen danske studier, der reelt afprøver effekterne af at måle hyppigere, end man plejer at gøre i danske skoler.

Hvad angår tværgående kompetencer – kompetencer, der er vigtige i mange eller alle fag – er der i litteraturen mange forslag til, hvordan disse kan måles. For måling af det 21. århundredes kompetencer, som fx problemløsning i samarbejde og computer- og informationsforståelse, har vi dog ikke fundet noget eksempel på måleinstrumenter, hvor man har testet, om instrumentet har en selvstændig og signifikant sammenhæng med skole- og arbejdsmarkedsudfald. Derimod er evidensen for, at socio-emotionelle kompetencer har en selvstændig betydning for skoleresultater og senere arbejdsmarkedstilknytning og viden omkring gode måleinstrumenter, meget mere omfattende.

I forhold til elevernes progression og forventningen til, hvordan de bør gøre fremskridt, bliver dette ofte evalueret ved, at klasser bliver sammenlignet på tværs af trin eller årgange, uden hensyn til sammensætning af eleverne. En anden almindelig metode er at holde elevernes individuelle præstationer op imod et fælles standardiseret ideal. Mere retvisende metoder, der tager udgangspunkt i elevernes individuelle standpunkt og muligheder for progression, er beskrevet i litteraturen, men de kræver som oftest mere data og mere avancerede metoder. Det kan derfor være svært for den enkelte skole at bruge denne type af progressionsmål.

I spørgsmålet om visualisering af data er der generelt i litteraturen en mangel på evidens-baserede guidelines til, hvordan data bedst visualiseres både i forhold til lærere, skoleledere og andre beslutningstagere. På baggrund af den eksisterende litteratur er en gennemgående pointe dog, at visualisering af data skal være så simpel og overskuelig som muligt.

DET FAGPROFESSIONELLE ARBEJDE MED ELEVERS PROGRESSION

Dette tema omhandler, hvordan fagprofessionelle bør arbejde med data-drevet udvikling, hvilke kompetencer det kræver, hvordan arbejdet kan it-støttes, og hvad der kendetegner skoler og kommuner, der arbejder godt med progressionsmåling. Selvom der ikke er mange effektmålinger på dette område, er der dog en stor litteratur, der samler op på erfaringer fra vellykkede og mindre vellykkede projekter. Denne litteratur fremhæver flere aspekter, der hænger sammen med vellykket brug af data:

- Skabelsen af en datakultur ude på skolerne er vigtig – herunder klare forventninger om og visioner og mål for, hvordan data skal bruges, og herunder skoleledelsens aktive rolle i den forbindelse.
- Fokus på læreres og pædagogers medinddragelse i udvikling og udvælgelse af data for at skabe en bedre implementering, opbygge ejerskab og sikre, at de data, som er tilgængelige, også opleves som relevante for lærere og pædagogers arbejde.
- Implementering og opbygning af datakapacitet kræver generelt tid og ressourcer, fx i forhold til opbygning af it-systemer og udvikling af fagprofessionelles datakompetencer.
- Der er behov for, at man i relativt høj grad understøtter fagprofessionelle i deres arbejde med fortolkning og særligt anvendelse af data til tilpasning og justering af undervisning (fx via professionel kompetenceudvikling og faglig sparring). Herunder er der et behov for IT-systemer, som kan koble mange typer af elevdata samt skabe mulighed for at koble undervisningsaktiviteter til elevpræstationer.

Relateret til det sidste punkt er et af budskaberne fra de få effektstudier af data-drevet udvikling, som vi har fundet, at kendskab til standpunkt og progression, eller visualisering af progressionsdata, ikke altid er nok til at løfte eleverne. Ofte er fagprofessionelle og skoler bevidste omkring, hvilke elever der fx har det fagligt svært, men kan være i tvivl om, hvad de skal gøre, for at det skal gå bedre for eleverne, eller mangler ressourcer til at gennemføre indsatser. I den forbindelse kan det netop tyde på, at man også skal understøtte med information og ressourcer i forhold til, hvordan man kan *handle* på baggrund af informationer om den enkelte elevs standpunkt og progression.

Der advares endvidere i litteraturen mod det at træffe beslutninger på baggrund af resultaterne af enkeltstående tests samt at rette et for stort fokus på målbare kompetencer. I arbejdet med progressionsmåling fremhæves nødvendigheden af, at lærere trænes i at opstille målbare mål for elevernes læring og progression, men det kan diskuteres, om man derved tillægger bestemte typer af kompetencer for stor vægt og overser andre. Selvom der er et øget fokus på at forsøge at måle andre kompe-

tencer end de traditionelt fag-faglige, er der få eksempler blandt de fundne studier på skoler, kommuner eller lande, der systematisk måler fx det 21. århundredes kompetencer.

AFSLUTTENDE KOMMENTARER

For det fortsatte projekt og for den videre anvendelse af resultaterne fra denne kortlægning er det vigtigt at diskutere, hvordan vi kan overføre international viden og praksis til en dansk kontekst. Selvom mange studier er fra USA, er der generelt få emner og områder i den litteratur, vi har fundet, som ikke også er aktuelle i Danmark. Mange gange gør lignende tendenser sig gældende på tværs af lande, fx i forhold til det øgede fokus på tværgående kompetencer. Når det gælder specifikke resultater, såsom spørgsmålene om, hvor hyppigt man skal måle, eller effekterne af data-drevet udvikling, er der dog ikke foretaget nogen danske studier, og det er derfor uklart, hvor vidt disse resultater kan overføres til danske forhold.

Nogle resultater kan være særligt vigtige at tage højde for i det fortsatte projekt. Studierne indikerer blandt andet, at udviklingen af den datainformerede skole og af fagprofessionelles datakompetencer kræver, at brugerne investerer mere tid i starten. Dette kan være en udfordring nu og her, men indeholder på sigt også muligheder, idet en bedre adgang, anvendelse og visualisering af eksisterende data, og øget brug af digitale læremidler mere generelt, har potentiale til at forenkle fagprofessionelles arbejde.

Vi fandt på baggrund af kortlægningen generelt ikke nogen typer af data om elevers læring og progression, hvor der helt mangler viden om, hvordan data kan måles. Men det er værd at bemærke, at det ikke er alle typer af data, som er tilgængelige på individniveau. Eftersom trivselsmålinger og målinger af socio-emotionelle kompetencer kan indeholde følsomme data og anonyme besvarelser, er de ikke altid velegnede at bruge til at følge den enkelte elevs progression. Der er også data, som ikke måles via tests eller mere formelle målinger, men stammer fra lærernes egne vurderinger og observationer, og her mangler der viden om, hvordan man bedst systematiserer og digitaliserer denne type af data.

På baggrund af litteraturen skal det desuden fremhæves, at det ofte delvist afhænger af øjet, der betragter, hvorvidt målinger vurderes som formative eller summative. Selvom projektet Data om elevernes læring og progression har et tydeligt formativt formål, er det derfor vigtigt at gennemtænke, hvilke forskellige incitament der kan være på spil i samme type af måling og data – afhængigt af det niveau, målingen bruges på.

INDLEDNING

Projektet ”Data om elevernes læring og progression” er en del af det fællesoffentlige initiativ ’It i folkeskolen’ og er en del af et samarbejde mellem Styrelsen for It og Læring, Digitaliseringsstyrelsen og KL. Formålet med udviklings- og inspirationsprojektet er at give videns- og praksisbaserede anbefalinger til, hvordan data om elevernes progression kan indsamles, sammenstilles, vises og understøtte, at danske skoleelever bliver så dygtige, som de kan. Projektet skal illustrere og give anbefalinger til, hvordan skoler og kommuner kan bruge data til at beskrive og visualisere elevens progression, og hvordan pædagogisk personale kan anvende data i deres didaktiske arbejde for at styrke elevernes læring. Projektet skal have et bredt datafokus og bringe såvel centrale som de mange lokale data i anvendelse. Projektet skal vise muligheder for udveksling af data om progression mellem it-systemer, som anvendes i folkeskolens arbejde.

Projektet udføres af et konsortium bestående af SFI – Det Nationale Forskningscenter for Velfærd, Danmarks Institut for Pædagogik og uddannelse (DPU), Aarhus Universitet, og Danmarks Evalueringsinstitut (EVA). Denne kortlægning er, sammen med en erfaringsopsamling af praksis i danske folkeskoler, projektets første del.

FORMÅL OG FORSKNINGSSPØRGSMÅL

Formålet med forskningskortlægningen og syntesen er at finde dansk og international viden, der sammen med erfaringsopsamlingen bidrager med grundlaget for det videre projektføreløb. En systematisk forskningskortlægning baserer sig principielt på alle studier inden for et givent felt. I dette projekt, hvor grundlaget for henholdsvis erfaringsopsamling og

forskningskortlægning og syntese skal bruges i udviklingen og afprøvnin- gen af indsatserne efterfølgende, prioriterer vi at sammenfatte et så re- præsentativt udvalg af studier i kortlægningen, som tidsrammen har til- ladt. Kortlægningen er stadig lavet med systematiske og eksplicite meto- der til at identificere, udvælge og vurdere relevante studier, men alle in- kluderede studier bliver ikke beskrevet i rapporten. Kortlægningen om- handler tre hovedtemaer:

1. Forståelser og definitioner af læring og progression
2. Måling af elevens læring og progression
3. Det fagprofessionelle arbejde med elevens progression.

I tabel 2.1 vises de tre hovedtemaer og de undertemaer og specifikke forskningsspørgsmål, der vejleder kortlægningen.

TABEL 2.1

Hovedtemaer, undertemaer og forskningsspørgsmål.

Tema	Undertema	Forskningsspørgsmål
Forståelser og definitioner af læring og progression		Hvilke forståelser, der er om- kring elevens læring og pro- gression?
Måling af elevens læring og progression	Datakilder, skalaer og metodikker	Hvilke data indgår/bør indgå i målinger af læring og pro- gression? Hvilke skalaer skal benyttes? Hvor hyppigt skal målinger foretages?
	Tværgående kompetencer, udvik- ling og trivsel	Hvorledes måles tværgående kompetencer? Hvordan måles trivsel?
	Referencegrupper	Hvad måles progression op imod?
	Visualisering	Hvordan visualiseres data, så de giver det rigtige grundlag for beslutningstagen?
Det fagprofessionelle arbejde med elevens progression	Fra progressionsmåling til tilpas- ning og udvikling af under- visningen	Hvordan kan progressionsmå- ling indgå som en del af det fagpædagogiske arbejde med at fremme elevens faglige og alsidige kompetencer samt trivsel? Hvilke praktiske erfaringer er der?
	It-understøttelse	Hvordan it-understøttes det faglige arbejde med progres- sionsmåling?
	Data literacy og kompetencer	Hvilke kompetencer har fagpro- fessionelle i forhold til for- tolkning af data? Hvilke kompetencer kræves der, og hvordan udvikles det?
	Institutionel kapacitet	Hvad kendetegner skoler og kommuner, hvor progressi- onsmåling er blevet en natur- lig del af det faglige arbejde på skolen?

Rapportens disposition følger denne opdeling. Ud over et metodekapitel (Bilag 1 <http://www.sfi.dk/elev-bilag1/>) indeholder kortlægningen et resultatkapitel for hvert hovedtema, som er opdelt efter de enkelte undertemaer og forskningsspørgsmål. Syntesen har også til formål at diskutere, på tværs af forskningsspørgsmålene, hvilke muligheder og udfordringer der kan være forbundet med at overføre international forskningsviden og international praksis til en dansk kontekst, samt hvorvidt der er typer af læring, som er vanskelige at fremskaffe data om. Vi diskuterer dette løbende i resultatkapitlerne og vender også tilbage til disse to spørgsmål i det sidste kapitel, Diskussion og perspektivering.

FORSTÅELSER OG DEFINITIONER AF LÆRING OG PROGRESSION

I dette afsnit tager vi udgangspunkt i forskningsspørgsmålet omhandlende, *hvilke forståelser der er omkring læring og progression*. Generelt tager vi i kortlægningen ikke afsæt i en bestemt definition af begreberne data, læring og progression, idet rapporten er styret af de studier, vi inddrager. Vi definerer her, hvilke typer af henholdsvis data, læring og progression, vi taler om, afhængigt af, hvad vi finder i litteraturen, og belyser i stedet de forskellige måder, hvorpå begreberne anvendes.

I afsnittet her sætter vi fokus på forskellige definitioner og forståelser omkring målinger af elever og deres formål. Vi har altså først og fremmest fokus på, hvilke typer af vurderinger der inkluderes i målinger af læring og progression, frem for hvordan man konkret kan foretage målinger af fx elevers progression. Dette aspekt tager vi i stedet op i afsnittene omhandlende henholdsvis måling af progression, trivsel og tværgående kompetencer.

I litteraturen om evalueringer og vurderinger på skoleområdet anvendes en række forskellige ord og begreber, der henviser til bagvedliggende forståelser af læring og progression. Generelt har der i vores søgning været få studier, der behandler de bagvedliggende overvejelser om måling af elevers læring og progression, og vi har derfor først og fremmest udvalgt to rapporter, som dybdegående behandler forståelser og definitioner bag disse begreber. Begge rapporter har et tværnationalt perspektiv og giver derfor svar på forskningsmålet, der går på tværs af kontekstuelle forskelle. Disse suppleres med eksempler fra litteraturen, der belyser de tendenser, som rapporterne fortæller om.

PERSPEKTIVER OG HOLDNINGER

Målinger af læring og progression skal overordnet forstås inden for en bredere samfundsudvikling med et øget fokus på 'value for money' og i forlængelse heraf fokus på effekt og evidens. Målinger af læring og progression på skoleområdet kan derfor forstås som et led i at sikre undervisning af høj kvalitet (OECD, 2013).

TRE CENTRALE BEGREBER

Tre begreber står centralt i litteraturen om målinger på skoleområdet, nemlig: 'evaluation', 'appraisal' og 'assessment'. De tre begreber fokuserer på tre forskellige niveauer. *Evaluation* omhandler bedømmelser af skoler, skolesystemer og politikker på området. *Appraisal* omhandler bedømmelser af de fagprofessionelle på skolerne, såsom lærere og skoledere, mens *assessment* omhandler vurderinger af den enkelte elevs præstationer (Nusche m.fl., 2012). At skelne mellem disse tre niveauer er vigtigt – en dansk oversættelse af begreberne er dog ikke lige til, idet både 'appraisal' og 'assessment' oversættes til *vurdering*. Dette notat har fokus på målinger af elevers læring og progression, hvorfor betegnelsen 'vurdering' udelukkende henviser til det engelske begreb 'assessment' i notatet her.

Vurderinger skal ikke forveksles med tests, der er et mere snævert begreb og kan betragtes som en metode til at indsamle data med det formål at foretage en vurdering. Vurderinger af elever indebærer, ifølge Harlen (2014), frembringelse, analyse, formidling og brug af data med et givent formål for øje. Vurderinger kan derfor tage mange former med det tilfælde, at de alle involverer fire trin: 1) elever involveres i en aktivitet, 2) der indsamles data om aktiviteten, typisk af skoler eller fra nationalt hold, 3) data vurderes ved at måle dem op mod en given standard, og 4) vurderingen beskrives og kommunikeres gennem forskellige værktøjer og medier (Harlen, 2014). Trin 2-4 beskrives nærmere i kapitel 5 og 6.

FORMÅL MED MÅLINGER

Det ultimative mål for evalueringer og vurderinger af elever og skolesystemer er at styrke elevernes udbytte af undervisningen og deres skoleforløb som helhed (OECD, 2013). Ifølge den seneste samlede OECD-rapport (2013), med fokus på landes praksis omkring evaluering og vurdering, kan vurderinger dog samtidig have en række andre formål, der kan indgå som delmål i denne målsætning, men som ikke nødvendigvis udfylder denne funktion. Vurderinger kan fx indgå som beslutningsgrundlag for at træffe beslutninger på skoleområdet. De kan også indgå i et lærings- og udviklingsperspektiv med henblik på at skabe bedre læring for eleverne. Overordnet skelner OECD-rapporten mellem tre formål, nemlig *accountability*, *udvikling* og i forlængelse af udvikling en *diagnostisk funktion*.

'Accountability' henviser til muligheden for, at man kan stille fx lærere og skoleledere til ansvar for undervisningen på den konkrete skole. Det foregår på flere måder, fx som et led i beslutninger om forfremmelse, højere løn, sanktioner eller som information til forældre, som de kan handle på baggrund af. Ved at holde skoleledere og lærere ansvarlige på denne måde er målet at skabe incitament til, at man forbedrer sine præstationer eller som et værktøj til at identificere skoler eller lærere, der underpræsterer sammenlignet med andre lærere eller skoler (OECD, 2013). 'Udvikling' henviser til brugen af vurderinger med understøttelse af undervisningen af eleverne som formål, mens den 'diagnostiske funktion' henviser til anvendelse af vurdering med det formål at afdække tilstanden på et givent område, fx i hvilket omfang målene for elevens læring indfris på tværs af skoler. Det er essentielt at finde en god balance mellem vurderinger, der foretages med henblik på 'accountability', og vurderinger, der skabes med henblik på 'udvikling'. På den ene side er det vigtigt, at både skoleledere og lærere kan stilles til ansvar for undervisning og skolemiljø, mens det på den anden side er vigtigt, at de data, der indsamles, også kan anvendes til at skabe udvikling og forbedring (OECD, 2013).

Der er her eksempler på empiriske studier fra USA, hvor et stærkt fokus på accountability har ført til skolepraksisser, der risikerer at underminere validiteten af tests. I den forbindelse er der også eksempler på langt mere problematisk adfærd som strategisk omklassificering af elevens diagnoser, så de ikke behøver at tage en given test, suspendering af elever, der forventes at score lavt og decideret lærersnyd (fx Deming & Figlio, 2016; Jacob & Levitt, 2003). Hvordan man helst præcis skal finde denne balance mellem målinger for henholdsvis accountability og udvikling, er der dog få eksempler på i litteraturen.

SUMMATIVE OG FORMATIVE MÅLINGER

Vurderinger af elever kan endvidere opdeles i *summative* og *formative* vurderinger. Summative vurderinger har til formål at opsummere eller måle, hvor meget en elev har lært af et givent undervisningsforløb eller på et givent tidspunkt. Det kaldes også 'vurdering af læring'. Formative evalueringer har til formål løbende at skabe viden om en elevs læringsforløb for derved at blive i stand til at sætte ind på de områder, hvor eleven har behov for at udvikle sine kompetencer yderligere. Det kaldes derfor 'vurdering for læring'. Inden for de formative vurderinger findes desuden *diagnostiske* vurderinger, der har til formål at identificere elevens udgangspunkt og derpå udvikle et passende undervisningsforløb for den enkelte elev (Harlen, 2014; OECD, 2013).

Summative og formative vurderinger har forskellige formål. De formative har ét primært formål, nemlig at tilvejebringe information, der kan anvendes til at understøtte elevernes fremadrettede læring, mens de summative har en række formål. De kan være målrettet dels den enkelte

elev, dels mere aggregerede niveauer såsom den samlede klasse, den samlede skole eller det samlede land.

I mange lande ser man her generelt en bevægelse mod større politisk og uddannelsesmæssigt fokus på brug af formative tests. Dette fokus på at anvende vurderinger som et led i udvikling af elevers læring er en nyere tendens inden for måling af elevers læring og progression. Tidligere er vurderinger blevet opfattet som uafhængige af undervisning og læringsprocesser, fx i form af prøver ved afslutningen af et undervisningsforløb, der i stedet opsummerer elevers færdigheder på et givent tidspunkt. Denne bevægelse mod formative målinger skyldes blandt andet, at større nationale standardiserede tests ofte ikke er detaljerede nok i deres afdækning af emner, til at lærere kan benytte disse tests til at forstå, hvor eleverne har brug for hjælp, og hvilke undervisningsstrategier der vil være bedst at anvende (Looney, 2011; OECD, 2013). Samtidig med dette fokus på mere formative målinger er der dog også blevet skabt et øget politisk fokus på skolers testscore og rankings. Der er derfor også en tendens til, at der laves flere målinger, der er – eller opfattes – som kun summative og herunder et øget fokus på *accountability* for skoler i forhold til særligt fag-faglige resultater (OECD, 2013).

NYE FORSTÅELSER AF LÆRING

Generelt er der inden for OECD-landene sket en reformering af uddannelsessystemer. Et mere snævert fokus på isoleret viden og færdigheder er i dag blevet udvidet med inddragelsen af fokus på mere komplekse kompetencer. Definitionerne af, hvilke kernekompetencer der karakteriserer komplekse kompetencer, varierer mellem lande. Fælles for dem alle er dog, at de udfordrer traditionelle opfattelser af, hvilke kernekompetencer elever skal udvikle i løbet af deres skolegang og i stedet anvender en bredere læringsmodel. Denne læringsmodel kan fx indeholde en kompleks kombination af viden, færdigheder, holdninger og handlinger, der er nødvendige for at bære sig i det 21. århundrede (*21st century skills*). Det vil sige, at der er fokus på, at eleven skal opnå læring, der kan bruges uden for skolen eller ”i den virkelige verden”. Disse komplekse kompetencer inkluderer dimensioner såsom kritisk tænkning, kreativitet, problemløsning, kommunikation, informations- og kommunikationsteknologi (*ICT literacy*) samt samarbejdsevner, sociale kompetencer og det at indgå som borger i et demokratisk samfund (*citizen skills*) (OECD, 2013). Det er dog i mange tilfælde svært at adskille komplekse kompetencer fra fag-faglige kompetencer, idet de på mange områder er delvist sammenhængende.

Generelt har det øgede fokus på at måle tværgående kompetencer en klar kobling til den teknologiske udvikling, som har flyttet fokus fra læring af fakta mod et større fokus på kompetencerne til fx at indhente, forstå og anvende informationer samt forholde sig kritisk hertil (Greenstein, 2012). Denne udvikling har altså også en betydning for,

hvordan vi gennemgående forstår og går til målinger af elevers læring og progression i dag.

Der er eksempler i litteraturen på forsøg på at måle andet end fag-faglige kompetencer i stor skala. West (2016) beskriver den amerikanske debat omkring, hvorvidt såkaldte ikke-kognitive kompetencer (*non-cognitive skills*), et samlebegreb, der i litteraturen inkluderer fx karaktertræk, motivation og sociale og emotionelle færdigheder, skal være en del af opfølgningssystemet for skoler (*school accountability systems*). Han inddrager her resultater fra Californien, som er den delstat i USA, der er kommet længst i udviklingen med at bruge målinger af ikke-kognitive færdigheder i opfølgning og evaluering af skoler. PISA-undersøgelsen i 2015 indeholdt ligeledes målinger af problemløsning i samarbejde og spørgsmål om ikke-kognitive kompetencer som fx motivation, holdninger til skolen og tro på egen formåen (*self-efficacy*) (OECD, 2016). Den danske nationale trivselsmåling er også et eksempel på målinger i stor skala af andre aspekter og kompetencer end de traditionelt fag-faglige (Keilow m.fl., 2014). I afsnittet omkring Måling af tværgående kompetencer vil vi mere detaljeret beskrive udviklingen af målinger af det 21. århundredes kompetencer og socio-emotionelle kompetencer.

MÅLING AF ELEVENES LÆRING OG PROGRESSION

I dette kapitel sætter vi fokus på, hvordan elevens læring og progression kan måles. Vi inddrager også studier om målinger og data, der kan forklare progression eller mangel på progression. Temaet berører derfor mere end bare målinger af læring og progression. Kapitlet er opdelt på følgende afsnit: Datakilder, skalaer og metodikker, Tværgående kompetencer, Udvikling og trivsel, Hvad måles progression op imod og Visualisering af data.

DATAKILDER, SKALAER OG METODIKKER

Dette afsnit omhandler de data, skalaer og metodikker, der anvendes i målinger af elevernes læring og progression. En ordbogsforståelse af data er en 'nøjagtig oplysning, ofte udtrykt i en objektiv eller målbar form' (www.ordnet.dk). Den type af data, som vi interesserer os for her, produceres af og anvendes i forskellige typer af tests, målinger, lærerobservationer og lignende. Det vil sige, at vi er interesserede i data, der produceres eller anvendes i en eller anden form for vurdering, defineret som i det foregående kapitel. Skalaer og metodikker er tæt forbundet med de data, der produceres af eller anvendes i en given vurdering, og vi behandler dem derfor samlet. Afsnittet afdækker litteraturens input til forskningsspørgsmålene omkring, *hvilke data der indgår eller bør indgå i målinger af læring og progression*, og *hvilke skalaer der skal benyttes i målinger af læring og progression*.

EKSEMPLER PÅ TYPER AF DATA

Der findes en række typer af data i form af forskellige tests og målinger, som kan anvendes i vurderinger af elevernes læring og progression. Data kan både efterspørges og udformes af interne og eksterne aktører på en

skole. Interne vurderinger er udformet af fagprofessionelle, ofte i samarbejde med elever, og implementeres som en del af undervisningen, enten i løbet af et undervisningsforløb (formativ funktion) eller ved dets afslutning (summativ funktion). Eksterne vurderinger er udformet af aktører uden for den enkelte skole. De er ofte konstrueret således, at elevers præstationer kan sammenlignes, ligesom der kan sammenlignes på tværs af skoler eller lande (OECD, 2013).

I litteraturen skelnes der, som fremhævet i kapitel 4, mellem summative og formative evalueringer. De formative evalueringers fokus på at inddrage elever i udviklingen af, hvordan vurderinger gennemføres, understøtter det stigende fokus på elevernes komplekse kompetencer og individuelle fremskridt. På den måde kan vurderinger udvikle og ikke blot opsummere elevernes udbytte af undervisningsplanerne (OECD, 2013).

En tredje type vurdering er innovative vurderinger eller 'performance-baserede' vurderinger. De kan være baseret på opgaver uden et bestemt facit, fx mundtlige præsentationer, essays, eksperimenter, projekter, samarbejdsopgaver, cases fra det virkelige liv eller porteføljemapper. Det primære i denne type vurdering er, at den vurderer elevens viden og kompetencer på baggrund af, hvordan vedkommende løser en opgave og ikke ved at efterspørge ét korrekt svar (Looney, 2011; OECD, 2013).

EN SAMLET RAMME FOR VURDERINGER

En central pointe i OECD (2013) er, at det er vigtigt at skabe en samlet ramme for vurderinger af elever. Det er der en række årsager til. Som vi så i forrige kapitel, kan målinger af elevers læring og progression have forskellige formål, som den samlede ramme skal tage højde for. Derudover er der fordele og ulemper forbundet med alle målinger, hvorfor evaluører mindsker bias ved at anvende flere typer af vurderinger med forskellige design til at afdække forskellige perspektiver af elevers læring og progression. Med de nye forståelser af læring, som beskrevet i kapitel 4 (om forståelser), følger desuden et behov for målinger, der indfanger komplekse læringsmål, fx det 21. århundredes kompetencer (OECD, 2013).

Blandt OECD-landene har den innovative vurderingsform kun fundet systematisk og sammenlignende anvendelse i begrænset omfang. Der er få eksempler på en systematisk brug af de samme typer af innovative vurderinger på tværs af skoler og kommuner, og innovative vurderinger er i høj grad interne. I stedet er det primære fokus for sammenligninger og evalueringer fortsat på at vurdere fag-faglige færdigheder blandt elever. Det indikerer, at selvom der i nationale undervisningsplaner er fokus på elevers komplekse kompetencer, så indfanger data og målinger ofte ikke disse kompetencer. Det kan betyde, at de komplekse kompetencer tillægges mindre værdi end de fag-faglige, idet de ikke synliggøres i mange af de eksisterende evalueringer, hvilket igen kan skabe motivation for lærere til at undervise i de mere fag-faglige færdigheder, hvor der findes ét korrekt

svar (OECD, 2013). Der er dog, som vi så i det tidligere kapitel, eksempler på lande og regioner, der måler ikke-kognitive kompetencer, fx i PISA-undersøgelsen og i den danske trivselsmåling.

Vurderinger af målinger, tests, et projekt eller andre former for data om en elev, baseres, ifølge OECD, ofte på vejledninger, der beskriver forskellige niveauer af elevers færdigheder. Derved har den, der vurderer, fx en lærer, et grundlag for at afgøre, om en præstation er under middel, middel eller over middel. Der findes flere forskellige værktøjer til at understøtte vurderinger: For det første er der præstationskrav (*performance criteria*), som er principper, som en elevs færdigheder eller opgaveløsninger kan bedømmes ud fra. For det andet findes der vurderingsskemaer (*rubrics*), som er et scoringsværktøj med kriterier og skalaer, hvor alle point er beskrevet og defineret. For det tredje findes typiske eksempler (*exemplars*), der viser virkelige eksempler på elevers opgaveløsninger og præstationer med henblik på at illustrere forskellige point på en skala (OECD, 2013).

OECD pegede for Danmarks vedkommende på, at der fandtes normer eller mål for, hvilken viden og færdigheder, undervisningen skal føre til, men der fandtes ingen standarder, der beskriver konkrete læringsudfald, som eleverne skal opnå, eller kriterier for, hvordan disse udfald skal måles og vurderes (OECD, 2013; Shewbridge m.fl., 2011). De Forenklede Fælles Mål, implementeret i 2015, er blevet udviklet med det formål at være mere anvendelige for lærerne. De er læringsmål, der tager udgangspunkt i elevernes læringsudbytte og er formuleret som kompetencemål og færdigheds- og vidensmål (Ministeriet for børn, undervisning og ligestilling, 2016c).

Det er vigtigt at have formålet med en vurdering for øje, når man vælger, hvilken type af data og skalaer, man vil basere sin måling på. Forskellige vurderingsformer er ofte udviklet og designet med et bestemt formål. Anvender man dem til andre formål, er der risiko for, at de vurderinger, man foretager på baggrund af målingen, ikke er valide eller er ukorrekte. Ønsker en skole fx at lave en diagnostisk vurdering af elever, skal testen designes, så den giver detaljerede informationer, der muliggør identifikation af udfordringer blandt de enkelte elever. Er formålet i stedet at blive i stand til at sammenligne på tværs af skoler eller kommuner, skal testen designes således, at den skaber pålidelige summative resultater med udgangspunkt i brede sammenlignelige kategorier (OECD, 2013).

En måling kan have flere formål samtidigt. Har den det, skal det primære formål med vurderingen nøje overvejes, og afsenderen af målingen skal foretage en eksplicit prioritering mellem de forskellige formål. Ifølge Newton (2007) er det ikke et problem, at en måling udfylder flere formål, hvis formålene ikke er logisk uforenelige, og hvis det fremgår tydeligt, hvad det valgte design, herunder data og skalaer, kan sige noget om.

RELIABILITET OG VALIDITET

I overvejelser af, hvornår og hvordan interne eller eksterne målinger skal anvendes, anbefaler OECD (2013), at man finder en balance mellem de to vurderingstyper. En fordel ved eksterne standardiserede vurderinger er, at de er forbundet med høj reliabilitet. De sikrer, at alle elever stilles de samme opgaver, og at resultaterne opgøres efter de samme standarder. Samtidig foretages de ofte under superviserede forhold, hvor man er sikker på, at eleven selv har lavet besvarelsen, og den, der vurderer elevens præstation, kender ikke eleven og er derfor uvildig, hvilket sikrer mod bias (OECD, 2013). Til gengæld kritiseres eksterne standardiserede målinger for at have lavere validitet end interne lærerbaserede vurderinger. De standardiserede tests muliggør ofte kun målinger af en begrænset del af undervisningsplanerne, og de har svært ved at indfange komplekse kompetencer, som fx det 21. århundredes kompetencer. Desuden kan evaluerende tests, med et mere summativt formål, skabe incitament blandt lærerne til at undervise eleverne i at blive testet i stedet for at fokusere på den dybere læring, såkaldt *teaching to the test*.

Interne – eller lærerbaserede – vurderinger er også forbundet med både fordele og ulemper. På den ene side giver de mulighed for, at vigtige færdigheder og kompetencer, herunder fx de komplekse kompetencer, der er svære at indfange med standardiserede tests, kan måles og dermed kommer til at indgå i en vurdering af eleven. De har således en højere sandsynlighed for at vurdere elevens præstationer i forhold til alle mål i en undervisningsplan (Crooks, 2004; Harlen, 2007). Når de interne vurderinger er tænkt ind i undervisningen, viser de desuden et mere autentisk billede af elevens kompetencer end vurderinger, som udelukkende er baseret på eksterne standardiserede tests (OECD, 2013).

På den anden side afhænger validiteten i de lærerbaserede vurderinger i høj grad af de muligheder, som den enkelte lærer har, og det er svært at sikre, at alle lærere formår at udnytte potentialet i de interne vurderinger og fx ikke blot måler på fag-faglige kompetencer (OECD, 2013). Derudover kan der være stor variation mellem lærere på, hvordan de vurderer deres elever, og de interne vurderinger kan derfor ikke bruges til sammenligninger på tværs af skoler.

Ifølge Crooks (2004) er det derfor nødvendigt at kombinere eksterne og interne vurderinger for at sikre størst mulig validitet og reliabilitet. Læringsmål, der kan vurderes på baggrund af eksterne vurderinger, skal vurderes ud fra denne type af data, mens mere komplekse kompetencer skal vurderes på baggrund af løbende interne lærerbaserede vurderinger (OECD, 2013).

Med dette udgangspunkt skulle komplekse kompetencer, som fx det 21. århundredes kompetencer, ikke være muligt at sammenligne på tværs af skoler. Der er dog ikke konsensus om dette i litteraturen, og der er forsøg på at måle og sammenligne denne type af kompetencer, som

tidligere nævnt. Vi vender tilbage til spørgsmålet om, hvordan det 21. århundredes kompetencer kan måles i afsnittet omhandlende Tværgående kompetencer.

BIAS I MÅLINGER

Der er også fordele og ulemper forbundet med forskellige tilgange og typer af vurderingsværktøjer. Looney (2011) beskriver fire grupper af vurderingsværktøjer med hver deres potentialer og ulemper (Looney, 2011; OECD, 2013):

Multiple-choice vurderinger kan anvendes til at vurdere fag-faglige kompetencer, men kan ikke indfange mere komplekse kompetencer. Derudover er der – hvis kvaliteten af designet er utilstrækkeligt – risiko for målefejl, for eksempel hvis eleverne misforstår bestemte spørgsmål eller svarer tilfældigt.

Adaptive vurderinger er computerbaserede vurderinger, der løbende tilpasser spørgsmålene i testen efter de svar, der bliver givet af den, der tager testen. Elever, der svarer rigtigt, får sværere spørgsmål og omvendt. Sådanne tests kan give mere detaljerede data om elevens færdigheder end traditionelle tests, fordi opgaverne tilpasses den enkelte elevs niveau undervejs, og eleverne dermed kan svare på flere opgaver, der er tilstrækkelige, men ikke for udfordrende i løbet af den samme tid. Af samme årsag fungerer adaptive tests også lige så godt for stærke og mindre stærke elevgrupper (Kreiner, 2009). Til gengæld kan det være problematisk at sammenligne elever på tværs, da de ikke har besvaret samme spørgsmål (Jacob & Rothstein, 2016). Det behøver dog ikke at være et problem, hvis en adaptiv test fx opfylder kravene for en såkaldt Rasch-skala, hvor man netop kan sammenligne elever på tværs (Kreiner, 2009).¹

Adaptive tests kræver desuden et stort spørgsmålsbatteri for at opnå en høj reliabilitet (Jacob & Rothstein, 2016; OECD, 2013). I Danmark anvender vi de nationale tests, der er adaptive tests. Ministeriet for Børn, Undervisning og Ligestilling har beregnet testenes reliabilitet og fundt overordnet en relativt høj grad af reliabilitet, mens enkelte profilområder havde en noget lavere reliabilitet (Ministeriet for børn, undervisning og ligestilling, 2016a).

Performance vurderinger er karakteriseret ved, at deres formål er at skabe grundlag for at vurdere en række kompetencer og holdninger blandt eleverne ved at bede dem om at løse en opgave i stedet for at give

1. Kravene for en Rasch-skala er, at opgaverne i en test skal være: 1) endimensionelle: opgaverne måler kun én færdighed; 2) monotone: chancerne for at svare korrekt på opgaverne bliver større, jo dygtigere eleven er; 3) homogene: rangordningen af opgaverne i forhold til sværhedsgrad skal være den samme for alle elever, uanset hvor dygtige de er; 4) lokalt uafhængige: svaret på en opgave må ikke indeholde noget, som gør det nemmere eller sværere at svare på en anden opgave; og 5) uden differential item-funktion: hvis to elever er lige dygtige, må der ikke være opgaver, der forfordeler en elev frem for en anden; fx opgaver, som typisk er nemmere for drenge at svare på end for piger (Kreiner, 2009).

ét korrekt svar. Sådanne opgaver kan fx være essays, mundtlige præsentationer, porteføljeopgaver, eksperimenter eller gruppeopgaver. Denne type vurdering er dog behæftet med bekymringer omkring deres reliabilitet. I modsætning til multiple-choice og adaptive tests, hvor en computer scorer testene, er det mennesker, typisk elevens lærer og evt. en censor, der bedømmer opgaveløsningen, og der kan være variation mellem forskellige læreres bedømmelse af den samme løsning.

Computerbaserede performance vurderinger har potentiale til at vurdere komplekse kompetencer ved at anvende informations- og kommunikationsteknologi. Denne vurderingsform er under udvikling, og især i USA arbejder man med at udvikle computerbaserede vurderingsværktøjer, der kan score løsninger af opgaver, der ikke bare har ét korrekt svar. Disse værktøjer vil både kunne bidrage til at imødekomme den reliabilitetsudfordring, der er forbundet med performancevurderinger, og validitetsudfordringerne, der er forbundet med multiple-choice tests.

Resultater af målinger er desuden afhængige af deres kontekst. Når der indsamles data til en vurdering af elever, er det vigtigt at tage højde for denne kontekst og undgå bias i vurderingen (OECD, 2013). På en skole kan der fx være flere elevgrupper, og udformningen af en test kan betyde, at bestemte grupper opnår en fordel frem for andre. Det er fx vigtigt at overveje den lingvistiske kompleksitet i sproget i en test. Høj lingvistisk kompleksitet kan betyde, at elever, der har testens sprog som modersmål, har lettere ved at opnå gode resultater i testen end elever, der ikke har testens sprog som modersmål. Det skyldes blandt andet, at de kan have behov for at bruge længere tid på at forstå og besvare spørgsmålene, ligesom de har højere risiko for at misforstå spørgsmålene (OECD, 2013).

Proportionen mellem forskellige trin på en skala har ligeledes betydning for de konklusioner, man kan drage. Ifølge Jacob og Rothstein (2016) er det almindeligt, at evaluatorene behandler målinger som intervalskalaer. Det vil sige, at hvis en elev flytter sig ét point på en skala, afspejler det den samme ændring, uanset om eleven går fra fx 1 til 2 eller fra 2 til 3. Denne antagelse er dog ikke holdbar, idet målinger af elevers læring og progression næsten altid skal forstås ud fra en ordinal tilgang, hvor der ikke nødvendigvis er proportionalitet mellem at flytte sig mellem forskellige trin på den samme skala. Fx kan et korrekt svar på to opgaver i en matematiktest begge give et point, men opgaverne kan være af forskellig sværhedsgrad og kan derfor ikke sammenlignes i en vurdering af elevernes matematikfærdigheder. Denne pointe har betydning for de analyser, der kan laves på baggrund af de målinger, der foretages på skoleområdet, samt hvilke kompetencer det kræver at analysere data fra diverse tests af elevers læring og progression (Jacob & Rothstein, 2016).

Der kan derudover skabes bias i validiteten af testens indhold, idet valget af en bestemt type viden kan betyde, at nogle elever opnår en

fordel frem for andre. Endeligt kan der være bias forbundet med de enkelte spørgsmål, man stiller, og i den forbindelse de spørgsmål, man ikke stiller (*bias in item selection*). Denne type bias betyder, at der kan være del-elementer i en test, der er behæftet med bias, og det kun at bruge dele af tests giver dermed ikke et retvisende billede af elevens læring og progression (Jacob & Rothstein, 2016).

Overordnet afhænger valget af data, skalaer og metodikker til vurdering af elevers læring og progression altså i høj grad også om vurderingens formål. Det er her vigtigt at holde sig for øje, om vurderingen fx er til formativt eller summativt brug, og hvorvidt den foretages internt eller eksternt. Der er i den forbindelse forskellige forhold omkring reliabilitet, validitet og bias, som er vigtige at være opmærksom på, afhængigt af den type af måling og vurdering, der foretages.

HYPPIGHEDEN AF MÅLINGER

I forlængelse af de foregående afsnit, som var centreret omkring datatyper og skalaer, vil vi i dette afsnit sætte fokus på forskningsspørgsmålet omhandlende, *hvor hyppigt målinger skal foretages*. For at belyse dette spørgsmål vil vi inddrage forskningsoversigter, rapporter fra en både dansk og international kontekst samt flere studier med effektmålinger og meta-analyser.

Overordnet er spørgsmålet om, hvor hyppigt man bør foretage målinger, dog svært at give et overordnet svar på. Den ”rette” hyppighed af målinger afhænger af mange forskellige forhold – herunder først og fremmest, hvilken type af test, man benytter, og hvad man ønsker at måle med testen. Samtidig kan målingernes effekt også være påvirket af elevernes alder og klassetrin, samt hvilke fag, målingerne beskæftiger sig med. Selvom der inden for dette område er studier, der undersøger årsagssammenhæng med troværdige metoder, bruger de forskellige typer af tests med forskellige formål, fag og elever, hvilket gør det svært at sammenligne på tværs.

FORSKELLIGE TYPER AF MÅLINGER

I spørgsmålet om hyppigheden af målinger har det blandt andet afgørende betydning, hvorvidt målingen skal anvendes formativt eller summativt. Som belyst i kapitel 4 er summative tests centreret omkring målinger af læring med fokus på at samle op på den læring, som har fundet sted, for at kunne vurdere elevers præstationer med henblik på fx karaktergivning (OECD, 2013).

Omvendt har man med formative tests fokus på målinger for læring. Formålet med denne typer måling er at skabe viden om elevens læringsproces og læringsbehov for på baggrund heraf at differentiere og

tilpasse den fremadrettede undervisning. Et vigtigt element i den formative måling består desuden af feedback til eleverne, umiddelbart efter at målingen er foretaget, hvilket også giver mulighed for at engagere eleverne selv i deres egen læringsproces. Denne type af test kræver, at der foretages løbende målinger og må derfor integreres i den daglige undervisning og læringsproces. Da formålet med formative og summative målinger er meget forskelligt, er der således også stor forskel på, hvor ofte de to typer af målinger bør foretages (OECD, 2013).

FEEDBACK SOM ET VIGTIGT ELEMENT I FORMATIVE MÅLINGER
Inden for uddannelsesforskningen har mange studier beskæftiget sig med effekterne og evidensen for brugen af formative målinger, hvor flere studier viser positive effekter på elevers læring (Van der Kleij, Feskens & Eggen, 2015). Udfordringerne ved nogle af disse studier er dog, at fokus af indsatsen ofte er meget bredt, hvorfor en sammenligning på tværs af studier kan være problematisk (ibid.). Andre studier har i stedet fokuseret direkte på brug af *feedback* til både elever og lærere, som er et helt centralt element i den formative tilgang til læring. Feedback anses her generelt for at være et effektivt redskab til at forbedre undervisning og læring, idet det er forbundet med lave omkostninger og høje effektstørrelser (se fx Education Endowment Foundation, 2016). Feedback kan dog have mange forskellige udformninger og kan være både verbal, skriftlig eller leveret gennem tests eller it-systemer og kan komme fra både lærere eller andre elever.

I en forskningskortlægning af Hattie og Timperley (2007) baseret på 12 tidligere meta-analyser konkluderer forfatterne, at feedback er imellem de 5-10 komponenter, som har størst indflydelse på elevers præstation (Hattie & Timperley, 2007). I studiet skelnes der mellem 4 niveauer af feedback; henholdsvis feedback omkring selve opgaven; omkring opgaveprocessen; omkring selvregulering og omkring eleven som person. Mest effektiv feedback viste sig her at være den type af feedback, som er fokuseret på procesniveau frem for selve opgavens løsning. Omvendt viste analysen lavere effektstørrelser ved feedback forbundet med ros, belønning eller straf. I forlængelse heraf fremhæver Van der Kleij, Feskens og Eggen (2015), baseret på en meta-analyse om effekterne af computerbaseret feedback, at udvidet feedback, der tilbyder eleven en forklaring, har større effekt, end feedback, der kun fokuserer på opgavebetsvarelsens korrekthed eller giver eleven det korrekte svar.

Hattie og Timperley (2007) konkluderer på baggrund af deres fund, at tests for ofte bruges til at vurdere elevers standpunkt frem for at blive brugt som et redskab til at videreudvikle læring. Dette på trods af, at omkostningerne ved *accountability*-tests er høje, mens feedback-udbyttet til lærere og elever er lavt.

EFFEKTERNE AF AT FORETAGE MÅLINGER

Studier om feedback fremholder ofte to mulige positive effekter. Dels får læreren information, som kan bruges til at tilpasse undervisningen til den enkelte elevs behov og give eleverne bedre hjælp. Dels får eleverne viden om deres standpunkt (og måske også progression), og hvad de mangler at lære. En stor litteratur inden for kognitiv psykologi indikerer også en tredje kilde til positive effekter af formative målinger: At lave tests kan i sig selv være en effektiv metode til læring. Rowland (2014) viser i en systematisk forskningskortlægning og meta-analyse, at den såkaldte *testing effect* er et robust resultat inden for den psykologiske litteratur; det vil sige, at lave en test af tidligere studeret information giver bedre fastholdelse end det at gentage studiet af information. Langt færre undersøgelser er dog gennemført i et skolemiljø med faglige tests som grundlag, og endnu færre af dem er gennemført i grundskolen. De studier, der er, peger dog generelt på samme type af resultat, det vil sige, at tests giver bedre resultater end det at repetere studiet af information. Effekterne på fastholdelse af information er større, når testene understøttes med korrigerende feedback (*corrective feedback*), eller når testene kombineres med repetition af information (se fx Karpicke & Grimaldi, 2012; McDaniel, Roediger & McDermott, 2007 og Rawson & Dunlosky, 2012 for oversigter). De tests, der bruges i disse artikler, er relativt enkle og ligger tæt på det materiale, som der undervises om. Man kan altså i høj grad sammenligne disse med de typer af tests, der som oftest bruges til formative målinger. Rawson og Dunlosky (2012) argumenterer, på grundlag af disse resultater, at lærere og elever underudnytter brugen af tests som læringsstrategi.

Der findes også andre undersøgelser, som indikerer, at skoler, klasser og lærere, der måler oftere eller overhovedet foretager målinger, opnår bedre resultater (se fx Black & Willian, 2009 og Fuchs & Fuchs, 2002 for oversigter). Konstantopoulos m.fl. (2015) beskriver resultater fra et stort lodtrækningsforsøg. 59 skoler er blevet randomiserede til at være enten kontrolskole eller bruge et af to *interim assessment programs*, hvor skoler og lærere gennemfører diagnostiske tests i løbet af skoleåret og får støtte til, hvordan de kan vurderes. De finder positive og signifikante effekter, først og fremmest for de elever, der havde de laveste resultater før interventionen. Dobbie og Fryer (2013) finder, at de skoler, i et udvalg af amerikanske charter-skoler (*charter schools*),² der er gode til at løfte elevernes faglige resultater, også, relativt set, foretager flere tests i løbet af skoleåret. Denne sammenhæng er dog ikke signifikant i et nyere, lignende studie med et større udvalg af charter-skoler (Chabrier, Cohodes & Oreopoulos, 2016). At måle hyppigere indgår som et komponent i en pakke af ”best practices”, der fordeles ud mellem skoler ved

2. Charter-skoler er finansieret med offentlige midler, men nyder en større grad af selvstyre end almindelige skoler. De kan blandt andet fastsætte deres eget pensum og har større frihed ved ansættelse af personale.

hjælp af lodtrækning i Fryer (2014). Studiet finder positive og rimelig store effekter af den samlede pakke, men studiet kan ikke skelne mellem effekterne fra de forskellige komponenter. Studier, der tester tilgange til data-drevet udvikling i skoler og skoledistrikter, hvor hyppigere målinger er inkluderet som komponent, finder oftest positive, men små og heriblandt ikke signifikante effekter (se fx Slavin m.fl., 2013).

Samlet set virker der til at være en sammenhæng mellem hyppigere målinger og bedre resultater. At foretage tests kan i sig selv være en måde at lære på, men det er, som tidligere nævnt, af flere grunde svært at give et mere præcist svar på spørgsmålet omkring, hvor hyppigt målinger skal foretages. Selvom nogle studier indikerer, at flere målinger er bedre, er der naturligvis en grænse for, hvor ofte man bør teste. Men hvor den grænse går, vides endnu ikke.³ Det er desuden svært at svare på, hvordan ”mere er bedre”-resultater, som først og fremmest stammer fra amerikanske studier, kan generaliseres til en dansk kontekst, idet vi ikke har fundet nogen danske studier, der reelt afprøver effekterne af at måle hyppigere, end man plejer at gøre i danske skoler.

MÅLING AF TVÆRGÅENDE KOMPETENCER

I dette afsnit vil vi tage spørgsmålet op omkring, *hvordan man kan måle tværgående kompetencer*. Tværgående kompetencer skal i denne rapport forstås som kompetencer, der ikke er koblet til et særligt fag, men som er vigtige i mange eller alle fag. I litteraturen omkring tværgående kompetencer og hvordan de kan måles, ser vi to, delvist relaterede, hovedstrømninger. Den ene strømning interesserer sig for emotionelle og sociale kompetencer, som, sammen med fx karaktertræk og motivation, i litteraturen også betegnes som ikke-kognitive kompetencer (*non-cognitive skills*) (fx Cunha, Heckman & Schennach, 2010; OECD, 2016).⁴

Den anden strømning omhandler en mængde kompetencer og færdigheder, der ofte kaldes det 21. århundredes kompetencer (*21st century skills*) (fx Greenstein, 2012; OECD, 2013). Det er dog et meget bredt begreb, og der mangler konsensus i litteraturen om, hvad der skal indgå

3. Det studie, der kommer tættest på at teste spørgsmålet om, hvor hyppigt målinger skal foretages, er en meta-analyse fra 1991, der først og fremmest inkluderer studier gennemført på amerikanske universiteter (Bangert-Drowns, Kulik & Kulik, 1991). Det er derfor kun blevet inkluderet som baggrundstudie. Elever, der tog mindst én test inden for en 15-ugers periode, scorer højere end elever, der ikke tog nogen test. Studiet indikerer også, at der er en relativt hurtigt aftagende effekt af flere målinger, dvs. den største effekt fås, når man går fra slet ikke at måle til at måle mindst én gang.

4. Begrebet bunder i det forhold, at disse færdigheder ofte kontrasteres med kognitive færdigheder, hvormed der menes fag-faglige færdigheder, som matematik- og læsefærdigheder, eller IQ-relaterede koncepter. Navnet er, som Borghans m.fl. (2008) påpeger, ikke helt rammende. Det er svært at finde et begreb for menneskelig adfærd, der ikke involverer kognitive processer i en eller anden grad.

(se fx Geisinger, 2016 for en diskussion). Men mange inkluderer fx samarbejde, kommunikation, problemløsning, kritisk tænkning, computer- og informationsforståelse (*computer and information literacy*), innovationskapacitet og kreativitet. Den samlede begrundelse for de færdigheder, der inkluderes i begrebet, plejer at være, at de alle anses for at være blevet vigtigere i det 21. århundrede, ofte på grund af udviklingen af informations- og kommunikationsteknologi.⁵ Relationen mellem det 21. århundredes kompetencer og mere traditionelle skolefærdigheder, som læse- og matematikfærdigheder, er, ifølge Greenstein (2012), at der er tale om kompletterende færdigheder, som alle er nødvendige for børns og unges udvikling.⁶ Det 21. århundredes kompetencer har også mange fælles dimensioner med kognitive og socio-emotionelle kompetencer, fx kan man forestille sig, at det kræver sociale kompetencer at være god til at arbejde sammen i hold, og kognitive kompetencer påvirker sandsynligvis problemløsningsfærdigheder. Det er derfor generelt svært at adskille tværgående kompetencer fra informations- og faktaviden.

For at give eksempler på, hvordan tværgående kompetencer kan måles i en skolekontekst, bruger vi dels et antal forskningsoversigter, og når det gælder det 21. århundredes kompetencer i høj grad også en ny udgave af tidskriftet *Applied Measurement in Education*, som er helt dedikeret til dette emne. Der er dog ret stor forskel på, hvor langt udviklingen af måleinstrumenter er kommet for de to typer af tværgående kompetencer. Når det gælder det 21. århundredes kompetencer, er der dels tale om et nyere begreb, dels er de forskellige dimensioner i begrebet i sig selv ofte sværere at indfange i målinger. Vi har fx ikke fundet noget eksempel på et måleinstrument i kategorien det 21. århundredes kompetencer, hvor man har testet, om måleinstrumentet har en selvstændig og signifikant sammenhæng med skoleresultater og senere arbejdsmarkedstilknytning.

I modsætning hertil har fx socio-emotionelle kompetencer en lang historie for at blive målt i samfundsvidenskabelige undersøgelser, særligt inden for psykologi, men interessen for skolens rolle i udviklingen af disse færdigheder og dermed for at måle socio-emotionelle kompetencer i skolen er steget i de seneste år. En forklaring her kan være de relativt nye resultater, der viser en betydelig og selvstændig effekt af socio-emotionelle kompetencer på skole- og arbejdsmarkedsudfald (se Almlund m.fl., 2011 og Borghans m.fl., 2008, for oversigter). Der er også resultater, der peger på, at ikke-kognitive kompetencer, inklusive socio-

5. Se Greenstein (2012, s. 22-23), for en sammenfatning af, hvad nogle forskellige definitioner af begrebet indeholder.

6. I nogle definitioner af det 21. århundredes kompetencer indgår både kognitive og ikke-kognitive kompetencer (Geisinger, 2016). Vi behandler dem separat her, dels fordi det er tendensen i de fleste studier, dels for at understrege, hvad der adskiller sig i målingen af det 21. århundredes kompetencer.

emotionelle kompetencer, bidrager til udviklingen af kognitive kompetencer (fx Cunha, Heckman & Schennach, 2010).

Evidensen for, at socio-emotionelle kompetencer har en selvstændig betydning for skole- og arbejdsmarkedsudfald og viden omkring gode måleinstrumenter, er dermed meget mere omfattende end for det 21. århundredes kompetencer. De næste to afsnit gennemgår eksempler på, hvordan det 21. århundredes kompetencer og socio-emotionelle kompetencer kan måles.

MÅLING AF DET 21. ÅRHUNDREDES KOMPETENCER

Mens kritisk tænkning og appliceret læring nok altid har været i fokus i skolen, og i den forstand ikke er noget, der gælder særligt for det 21. århundrede, er selve målingen af det 21. århundredes kompetencer et nyere fænomen (Greenstein, 2012). Fokus på at måle det 21. århundredes kompetencer har en klar kobling til den teknologiske udvikling. *Digital literacy* er fx meget mere relevant i dag, end det var for 20 år siden. En yderligere strukturel motivering er, at den stadig øgede informationsmængde og de tilgængelige instrumenter for at finde information flytter fokus fra læring af fakta til læring af, hvordan man indhenter, forstår og anvender information (Greenstein, 2012). I forlængelse af disse tendenser er der også, som tidligere nævnt, et øget fokus i mange lande på, at skolen skal udvikle bredere kompetencer end kun de fag-faglige (Ainley m.fl., 2016; OECD, 2013).

Der er en stor mængde begreber, som kan indgå i det 21. århundredes kompetencer, og listen på instrumenter, der er blevet brugt til at måle disse færdigheder, er ligeledes lang. Greenstein (2012) deler det 21. århundredes kompetencer op i tre kategorier, nemlig; tænke, handle og leve (*thinking, acting, living*), og nævner følgende eksempler på typer af instrumenter: skemaer (*rubrics*), tjeklister, selv-vurdering og -reflektion, *peer-review*, observationer, logbøger, anekdotiske optegnelser (*anecdotal records*), konceptkort (*concept maps*), dagbøger, *questioning*, konferencer og *portfolio reviews* (s. 53ff). Vi har desværre ikke mulighed for at beskrive måleinstrumenter for alle begreber, der kan indgå i definitioner af det 21. århundredes kompetencer i denne kortlægning. I stedet har vi valgt at fokusere på eksempler på, hvordan henholdsvis computer- og informationsforståelse (*computer and information literacy*), kreativitet og problemløsning i samarbejde (*collaborative problem solving*) er blevet målt i en række nye studier. Vi har udvalgt disse kompetencer, fordi de er eksempler på færdigheder, der er komplekse og dermed er gode til at illustrere både den appel, der er i at kunne måle komplekse færdigheder og samtidig også de problemer, der kan opstå.

Ainley m.fl. (2016) forsøger i en større tværnational undersøgelse at måle elevens computer- og informationsforståelse, der defineres som “an individual’s ability to use computers to investigate, create, and com-

municate in order to participate effectively at home, in school, in the workplace, and in society” (s. 292). Undersøgelsen inkluderer over 60.000 8.-klasses-elever, fordelt på mere end 3.300 skoler i 21 forskellige lande. Det anvendte måleinstrument er designet til at indfange to overordnede kompetencer, nemlig: indsamling og forvaltning af information (*collecting and managing information*) og produktion og udveksling af information (*producing and exchanging information*) (Ainley m.fl., 2016).

Selve målingen bestod af en it-baseret prøve baseret på 4 moduler med i alt 62 forskellige kortere spørgsmål og længere opgaver. Som et eksempel drejede et af modulerne sig om at planlægge et efterskoletræningsprogram, hvor der indgik en serie af kortere opgaver som at allokerer redigeringsrettigheder til andre studenter. I den længere opgave skulle eleverne designe en plakat for programmet. Ved at koble den enkelte elevs besvarelser med opgavernes sværhedsgrad blev eleverne vurderet på en hierarkisk skala med 4 niveauer. Målingen viste her en moderat sammenhæng mellem elevernes egen vurdering af it-kompetencer og deres score på den samlede skala, og forfatterne konkluderer dermed, at testen reelt indfanger og måler elevernes computer- og informationsforståelse (Ainley m.fl., 2016).

Et andet eksempel på måling af det 21. århundredes kompetencer er Lucas (2016), som udvikler et eksempel på, hvordan man kan måle elevens kreativitet. Baseret på eksisterende forskning og teori på feltet definerer Lucas her kreativitet ud fra 5 centrale aspekter, nemlig; nysgerrighed (*inquisitiveness*), forestillingsevne (*imagination*), vedholdenhed (*persistence*), samarbejdsevne (*collaborative*) samt disciplin. Hver af de 5 færdigheder er uddybet med 3 forskellige underaspekter, såsom det at undre sig og stille spørgsmålstegn, at udvikle ideer og bruge intuition, at kunne tolerere usikkerheder, give og modtage feedback, udvikle nye færdigheder samt gøre sig kritiske refleksioner (Lucas, 2016). 12 skoler i England blev udvalgt til at deltage i studiet på baggrund af deres engagement og erfaringer med at udvikle elevens kreative kompetencer. I samarbejde med skolens lærere og elever udviklede man et online redskab til at vurdere de 5 kreative dimensioner og subdimensioner, ud fra henholdsvis deres styrke, bredde og dybde. En udfordring ved lærernes brug af måleredskabet var dog, at de mange dimensioner betød, at der for hver af de 5 kompetencer var potentielt 12 felter at forholde sig til, hvilket kunne gøre vurderingen uoverskuelig. Med studiet ønskede man at demonstrere, hvordan kreativitet kan operationaliseres og udgøre et redskab til at tilgå og udvikle elevens kreativitet. Studiet havde dog også visse begrænsninger, såsom elevernes alder, lærernes villighed til at deltage og selve spørgsmålet om måleredskabets validitet og reliabilitet (Lucas, 2016).

Et tredje og sidste eksempel, som vi vil fremhæve, omhandler færdigheder til at løse problemer i fællesskab. Care, Scoular og Griffin (2016) skelner i deres definition mellem to forskellige domæner – hen-

holdsvis sociale og kognitive kompetencer. Under det sociale domæne fremhæves kompetencerne til at deltage (*participation*), perspektivere (*perspective taking*) og det at kunne foretage sociale reguleringer (*social regulation*), såsom at forhandle og tage ansvar. Fra det kognitive domæne er der fokus på kompetencen til at foretage opgavemæssige reguleringer (*task regulation*), såsom at skabe systematik og opsætte mål samt evnen til at opbygge viden (*knowledge building*). Der blev foretaget en pilotmåling i 6 forskellige lande, baseret på hele klassers deltagelse. Hvert land involverede som minimum 60 elever i 11-års-alderen, 300 elever i 13-års-alderen og 300 elever i 15-års-alderen. Selve målingen bestod i en online baseret prøve, indeholdende en praksisopgave samt et spørgeskema. De forskellige praksisopgaver blev udviklet med fokus på, at deres løsning skulle være afhængig af online samarbejde på tværs af elever, samt at opgaverne skulle aktivere de ressourcer, som var en del af definitionen på at løse problemer i fællesskab.

Care, Scoular og Griffin (2016) konkluderer, at vores evne til at indfange og måle komplekse sociale og kognitive kompetencer i en online kontekst stadig er på sit indledende stadie. En grundlæggende problematik er, hvordan man ikke kun rammer faglige kompetencer, men også indfanger de sociale. I målingen her blev det sociale aspekt tilgæet via en log over elevernes online chat omkring opgaveløsningen, men der ligger her en udfordring i, hvordan man systematisk kan kode og vurdere sådanne interaktioner. Herunder, hvordan man præcist skelner mellem de forskellige typer af kompetencer i en vurdering af elevernes præstation (Care, Scoular & Griffin, 2016).

Eksemplerne på målinger af komplekse problemløsningsfærdigheder illustrerer overordnet forskellige problemer i forhold til måling af det 21. århundredes kompetencer. Som nævnt i afsnittets indledning er det i mange tilfælde svært at adskille disse færdigheder fra informations- og faktaviden og evaluere dem separat. Den type af faktoranalyser,⁷ der ligger til grund for de fleste målinger af karaktertræk, er dermed nødvendige, men ikke tilstrækkelige, til at sikre disse måleinstrumenters validitet (se Ercikan & Oliveri, 2016, for en diskussion omkring dette).

MÅLING AF SOCIO-EMOTIONELLE KOMPETENCER

I en systematisk forskningsoversigt af instrumenter til måling af sociale og emotionelle kompetencer for børn og unge inkluderer Humphrey m.fl. (2012) koncepterne selvbevidsthed (*self-awareness*), selvkontrol (*self-management*) og social bevidsthed (*social awareness*) i emotionelle færdigheder. Derudover inkluderes koncepterne social problemløsning (*social pro-*

7. En faktoranalyse er en teknik, som kan belyse, hvorvidt bestemte spørgsmål spørger ind til den samme underliggende skala, og analysen kan bidrage til en indledningsvis vurdering af, hvordan de mange spørgsmål i spørgeskemaet grupperer sig på underliggende skalaer. Endelig kan analysen identificere spørgsmål, der ikke egner sig til at indgå i skalaen, fordi de ikke bidrager til at beskrive en given faktor (Keilow m.fl., 2014).

blem solving) og relationsdannelse (*relationship skills*) i (pro)sociale færdigheder. De finder 12 instrumenter, som er blevet valideret for børn eller unge med en relativ stor aldersspredning (fx ikke kun for daginstitutionsbørn), som er tilgængelig på engelsk, og som også er blevet brugt i tidligere forskning. Kiil, Arendt og Rosholm (2016) fokuserer på spørgsmålet om, hvordan trivsel kan måles, men giver også eksempler på måleinstrumenter, der kan bruges til at måle karaktertræk og socio-emotionelle kompetencer og er blevet brugt i Danmark. Ingen af de 12 instrumenter, som Humphrey m.fl. (2012) nævner, indgår dog i deres oversigt.

I en helt ny oversigtsartikel går Kiil, Sjøe og Pontoppidan (2016) igennem måling af børns trivsel i interventionsstudier med udgangspunkt i en psykologisk forståelse af trivsel. Forfatterne mener, at den psykologiske forståelse af børns trivsel ofte defineres som deres socio-emotionelle funktion og udvikling. Socio-emotionelle kompetencer for børn i skolealderen er, ifølge Kiil, Sjøe og Pontoppidan (2016), centreret omkring følelsesregulering og interaktion med andre. Deres definition af socio-emotionelle kompetencer er tæt på den tidligere beskrevne i Humphrey m.fl. (2011), men tilføjer nogle aspekter i forhold til beslutningstagning (fx at kunne tage beslutninger på et reflekteret grundlag).

Ud fra disse definitioner står det klart, at målinger af trivsel og socio-emotionelle kompetencer ofte har overlappende elementer. I en dansk kontekst er Strengths and Difficulties Questionnaire (SDQ) fx blevet brugt i relativt stor skala, som en måling af ikke-kognitive kompetencer (fx Datta Gupta & Simonsen, 2010), men er også ofte nævnt i litteraturen som et måleinstrument for psykologisk trivsel (fx Goodman & Goodman, 2009). Vi beskriver SDQ mere indgående i afsnittet, som omhandler trivselsmåling.

I forhold til at måle progression i socio-emotionelle kompetencer er det, ifølge Kiil, Sjøe og Pontoppidan (2016), en fordel at bruge måleinstrumenter med forholdsvis brede aldersintervaller, fordi det giver bedst mulighed for at bruge det samme spørgeskema over tid. Det kan også være en fordel at bruge skemaer med flere potentielle respondenter: At kunne spørge både forældre, lærere og børn giver forskellige aspekter af et børns socio-emotionelle kompetencer. Svarene er på den anden side forskellige, afhængigt af hvem der har besvaret et skema. De advarer også omkring påvirkning af respondenter, idet spørgsmål om private forhold kan starte en følelsesmæssig proces hos børn og unge, og dermed bør man være opmærksom på, at der er muligheder for at følge systematisk og professionelt op på bekymrende besvarelser. Forfatterne påpeger også, at spørgeskemaer, der er blevet validerede på gruppeniveau, ikke nødvendigvis er velegnede til screening eller diagnosticering af individer.

Blandt skemaer med maksimal besvarelsestid på 20 minutter for lærere og 30 minutter for elever, som er valideret, og som findes i en dansk udgave til måling af børn mellem 8 og 12 år, der er inden for eller

på grænsen til normalområdet, anbefaler Kiil, Sjøe og Pontoppidan (2016) først og fremmest tre måleinstrumenter. De tre instrumenter er Social Skills Rating System (SSRS), Behavior Assessment System for Children, Second Edition (BASC-2) og SDQ. Bedømmelsen er lavet efter kriterierne: aldersmålgruppen, skalaernes indhold, psykometriske egenskaber, udbredelse, respondenttyper, gruppebesvarelse, besvarelsestid og normer.

MÅLING AF TRIVSEL

I forlængelse af foregående afsnit omhandlende målingen af tværgående kompetencer vil vi i dette afsnit sætte fokus på spørgsmålet omkring, *hvordan man måler trivsel*. Forskning peger generelt på, at skoleelevers trivsel har afgørende betydning for elevers faglige udvikling. Ikke alene har god trivsel i barndommen en positiv indvirkning på sundhed og trivsel senere i livet – trivsel udgør også en væsentlig faktor i arbejdet med at styrke børns læring. Der er således gode grunde til at interessere sig for skoleelevers trivsel, og det er da også et område, som modtager stigende opmærksomhed – både inden for uddannelsesforskningen og i politisk regi (Keilow m.fl., 2014). For at belyse dette spørgsmål inddrager vi flere forskningsoversigter samt rapporter fra en både dansk og international kontekst. Vi benytter her begrebet trivsel som samlebegreb for både trivsel og det engelske begreb well-being.

DEFINITIONER OG VÆRKTØJER TIL MÅLING AF TRIVSEL

En rapport udgivet af KORA i 2016 syntetiserer eksisterende forskning omhandlende børn og unges trivsel og målinger heraf (Kiil, Arendt & Rosholm, 2016). I studiet defineres flere forskellige videnskabelige tilgange til undersøgelsen af børn og unges selvrapporterede trivsel. I litteraturen skelnes der her generelt mellem *subjektiv trivsel* (eller hedonistisk), der fokuserer på glæde og tilfredshed med livet, og *psykologisk trivsel* (eller eudaimonisk), der inddrager dynamiske processer såsom personlig udvikling og vækst. Subjektiv trivsel måles ofte ved at spørge direkte, hvordan barnet eller den unge har det, fx ved hjælp af måleinstrumenterne Cantrils stige, Huebners livstilfredshedsskala eller det positive og negative affektskema. Psykologisk trivsel dækker mere komplekse koncepter såsom selvværd og følelsen af kontrol og depression og måles typisk ved hjælp af batterier af spørgsmål (Kiil, Arendt & Rosholm, 2016). Inden for forskningen om selvrapporteret trivsel er der dog diskussioner omkring usikkerheden ved, hvad de præcist måler, samt undersøgelseernes reliabilitet og validitet. Forfatterne fremhæver dog et studie, som konkluderer, at det inden for forskningen om trivsel er bredt anerkendt, at børn fra mindst 10 år og opefter er i stand til at svare på spørgsmål om deres egen trivsel (ibid.).

På baggrund af erfaringsindsamlingen omkring definitioner og måleredskaber omhandlende trivsel anbefaler man fra KORA to forskellige redskaber til måling af subjektiv trivsel, nemlig *WHO-5* og *KID-SCREEN*. *WHO-5* fokuserer på positiv psykologisk trivsel, mens *KID-SCREEN* fokuserer mere bredt på trivsel defineret som livskvalitet på tværs af livssfærer, herunder fysisk og psykologisk trivsel og trivsel relateret til forældre, jævnaldrende og skolemiljø. En risiko ved det brede fokus i *KID-SCREEN* er dog, at det er svært at indfange effekterne af en specifik intervention, hvorfor man anbefaler, at målingen benyttes som et instrument til at skabe viden om gruppen af børn, som et givent studie omhandler (Kiil, Arendt & Rosholm, 2016). I forhold til målingen af sociale og emotionelle funktioner, som et aspekt af psykologisk trivsel, anbefales *SDQ* (Strengths and Difficulties Questionnaire). *SDQ* er en valideret psykometrisk skala, der blandt andet kan bruges til at måle emotionelle problemer, adfærdsproblemer, hyperaktivitet og kammeratskabsrelationer, og instrumentet er billigt og findes tilgængeligt på dansk. Et af kritikpunkterne ved brug af *SDQ* kan dog være en manglende evne til at belyse forandring (ibid).

Som udgangspunkt for udviklingen af den danske trivselsmåling fik Ministeriet for Børn, Undervisning og Ligestilling ligeledes foretaget en systematisk kortlægning med fokus på indsatser, der har en effekt på undervisningsmiljø og/eller trivsel (Keilow m.fl., 2014; Rambøll, 2014a). I forlængelse af rapporten fra KORA viste kortlægningen, at der blandt de inkluderede studier var meget stor spredning i forhold til, hvilke indikatorer der anvendes til måling af elevers trivsel og undervisningsmiljø. Nogle studier tog udgangspunkt i en meget bred definition af trivsel og børns relationer til voksne, mens mange studier omvendt satte fokus på meget specifikke elementer af trivsel (fx selvtillid og selvværd). Endvidere fandt man i kortlægningen, at enkelte af de inkluderede studier anvendte psykometriske instrumenter såsom fx *SDQ* eller *Rosenbergs Self-Esteem Scale*, mens størstedelen af studier anvendte mindre kendte indikatorer for trivsel (Rambøll, 2014a).

UNDERSØGELSER OM BØRN OG UNGES TRIVSEL

På baggrund af de inkluderede studier fandt man i kortlægningen fra Rambøll blandt andet, at elevers mestring af deres egen adfærd og læreres tydeliggørelse af acceptabel adfærd i klassen korrelerer positivt med både elevers trivsel og læring (Keilow m.fl., 2014; Rambøll, 2014a). Endvidere fandt man i kortlægningen, at skolens fysiske og strukturelle rammer også har en betydning i relation til undervisningsmiljø og trivsel blandt elever og, i mindre grad, organiseringen af undervisningen (ibid.).

En forskningskortlægning foretaget af Gutman og Vorhaus (2012) konkluderede ligeledes, at der er en klar statistisk sammenhæng mellem elevers trivsel og deres efterfølgende faglige kompetencer målt

via faglige tests. Forfatterne undersøger blandt andet spørgsmålet om elevers trivsel ved at afdække elevernes glæde ved at gå i skole samt deres engagement i skolearbejdet. Studiet fandt, at disse to parametre var signifikant forbundet med elevers faglige udvikling – særligt i de senere skoleår. Studiet fandt endvidere, at følelsesmæssig trivsel er vigtig i de tidlige skoleår, mens problematisk adfærd hos eleven samt elevens engagement i skolen er vigtige parametre for læring og faglig udvikling i de senere skoleår (Gutman m.fl., 2012; Keilow m.fl., 2014).

Hvad angår undersøgelser af børns og unges trivsel i en skandinavisk kontekst, foretages der både i Norge og Sverige nationale målinger af trivsel. Norges nationale måling, *Elevundersøgelsen*, er obligatorisk for 7-11. klassetrin og har først og fremmest fokus på klasseledelse og relationer i skolen, mens andre spørgsmål er centreret omkring aspekter af elevens læring. Undersøgelsen er derudover frivillig for 5. og 6. klassetrin, og data indsamles to gange årligt via et elektronisk spørgeskema, som omfatter 65 obligatoriske spørgsmål samt supplerende frivillige spørgsmål (Keilow m.fl., 2014).

I Sverige er den nationale måling af trivsel, *Attityder till skolan*, derimod frivillig, foretages hvert tredje år og genereres på baggrund af en stikprøve af lærere og elever i 4. til 9. klasse. I 2015 bestod stikprøven af 3.869 4.-6.-klasses-elever, 1.472 7.-9.-klasses-elever og 1.566 lærere (Keilow m.fl., 2014; Skolverket, 2016). Temaerne i den svenske undersøgelse drejer sig om trivsel, tryghed, relationen mellem lærer og elever, lærernes kompetencer, stressniveau, krav til eleverne samt krænkelse og mobning. Undersøgelsen er her delt op i to versioner, hvor undersøgelsen rettet mod 4.-6. klasse indeholder 80 spørgsmål, mens versionen til 7.-9. klassetrin består af 115 spørgsmål. Alle svar angives på en fem-trins-skala (Keilow m.fl., 2014).

I en dansk kontekst kan blandt andet nævnes *Skolebørnsundersøgelsen*, som er det danske bidrag til det internationale forskningsprojekt, som er varetaget af World Health Organization, og i Danmark udføres af Statens Institut for Folkesundhed. Den seneste spørgeskemaundersøgelse blev gennemført i 2013/2014 med 44 deltagende lande på tværs af Europa og Nordamerika (Inchley m.fl., 2016). Undersøgelsen dækker et bredt spekter af temaer inden for børns sundhedsadfærd og selvrappede trivsel. Hvad angår elevtrivsel, dækker undersøgelsen her temaer omkring selvvurderet helbred, livstilfredshed, skoletilfredshed og mobning. Undersøgelsen tager afsæt i en række internationalt anerkendte skaler, fx *Schwarzers skala for self-efficacy*, som er rettet mod barnets tiltro til egne kompetencer og formåen (Inchley m.fl., 2016; Keilow m.fl., 2014).

I mange kommuner foretages ligeledes forskellige undersøgelser med fokus på sundhed og trivsel. Af eksempler kan blandt andet nævnes *Københavnbarometeret*, udviklet af Københavns Kommune, og Århus Kommunes *Store Trivselsdag*, mens mange mindre kommuner også har foretaget selvstændige trivselsundersøgelser (Keilow m.fl., 2014).

DEN DANSKE TRIVSELSMÅLING

Et centralt fokus i den seneste folkeskolereform, som blev gennemført i Danmark i august 2014, var en styrkelse af elevers trivsel. I den forbindelse blev det fra skoleåret 2014/2015 derfor obligatorisk for alle folkeskoler at gennemføre en årlig trivselsmåling med fokus på at afdække elevers trivsel, undervisningsmiljø samt ro og orden i klasse- og skolemiljø (Keilow m.fl., 2014; Undervisningsministeriet, 2014).

I definitionen og målingen af *trivsel* fremhæves tre vigtige dimensioner, nemlig: elevens psykiske og fysiske velbefindende, elevens oplevelse af faglige og personlige kompetencer samt elevens oplevelse af støtte og inspiration fra omgivelserne. I forhold til begrebet *undervisningsmiljø* blev der lagt vægt på oplevelsen af at være inkluderet fagligt og socialt, at have tillid til klassekammerater og voksne på skolen, samt fokus på klasseledelse, undervisningsdifferentiering og skolens fysiske rammer. Hvad angår begrebet *ro og orden* var fokus på, at eleven ikke må føle sig forstyrret af larm og uro, men også, at eleverne stimuleres og er fri for kedsomhed og passivitet (Undervisningsministeriet, 2014).

Trivselsmålingen består i sin endelige version af 40 spørgsmål til elever i 4.-9. klasse og 20 mere enkle spørgsmål til elever i 0.-3. klasse og blev gennemført første gang i foråret 2015. Målingens resultater på klasseniveau er tilgængelige for alle undervisere, skoleledelser, skolebestyrelser, skoleforvaltninger og kommunalbestyrelser. Eleverne selv præsenteres for deres egen classes resultater af deres lærer, og forældre får præsenteret klassens resultater af enten lærer eller skoleleder. Målingen er som udgangspunkt frivillig for den enkelte elev, men obligatorisk for klassen, og alle elever er anonyme i deres besvarelse (dog med kommunal variation). Dette skyldes først og fremmest et etisk hensyn til målingens personfølsomme oplysninger og samtidig det forhold, at respondenter generelt svarer mere ærligt på spørgsmål, hvis de er sikret anonymitet (Ministeriet for børn, undervisning og ligestilling, 2016b). I forhold til visualisering af data er det her værd at bemærke, at informationer om elevernes trivsel dermed ikke er mulig at følge på individniveau for fagprofessionelle.

HVAD MÅLES PROGRESSION OP IMOD?

I dette afsnit sætter vi fokus på forskningsspørgsmålet omkring referencegrupper og mere specifikt, *hvad progression måles op imod*. De fleste skoler har et mål om at sikre bedst mulig undervisning og høj indlæring hos elever på alle klassetrin. For at opnå disse mål er det nødvendigt for lærere, skoleledere og andre beslutningstagere at have redskaber til at identificere, blandt andet hvilken undervisning der fungerer bedst, og hvilke elever der fagligt har det svært. Lærere behøver også egnede redskaber

for at sikre, at alle elever, både de fagligt stærkeste og de fagligt svageste, gør fremskridt og udnytter deres faglige potentiale.

Lande, skoler og lærere angriber progressionsmåling forskelligt. I følgende afsnit vil vi forsøge at belyse både, hvordan elevers progression måles, hvordan elevers progression analyseres, og hvad progression måles op imod i en international kontekst. I afsnittet berører vi, hvilke målinger eller sammenligninger der giver den mest valide beskrivelse af en enkelt elevs, eller en samlet klasses, progression.

Tests i folkeskolen udføres ofte for at evaluere og målrette undervisningen, så niveauet for både den enkelte elev og skolen samlet set bliver så højt som muligt. Faglige tests bør også anvendes til at forbedre den enkelte lærers undervisning og støtte deres arbejde med undervisningsdifferentiering. Forventningerne til, hvordan disse tests skal anvendes, kan afhænge af, hvilken aktør, man beskæftiger sig med. Der eksisterer forskellige opfattelser af, hvordan data skal bruges, hvis man sammenligner skoleledere, lærere og forældre. Der eksisterer også forskellige opfattelser af, hvordan elevers progression måles bedst muligt, eller i hvert fald, hvordan elevers faglige progression skal fremlægges (OECD, 2013). Forældre til elever i folkeskolen har ofte et ønske om at vide, hvordan deres eget barn klarer sig og gør fremskridt. Forældre anser derfor ofte sammenligninger på tværs af jævnaldrende eller resultater af, hvordan barnet har klaret sig på standardiserede tests, for at være anvendelige. Lærere har derimod primært et ønske om at anvende testresultaterne til at forstå elevers styrker og svagheder på samtlige faglige niveauer og ud fra dette kunne målrette og forbedre undervisningen. Skolelederes brug af testresultater er ofte primært rettet mod evaluering af skolens samlede progression eller faglige niveau (OECD, 2013). Det er vist ved flere lejligheder, at læreres brug af monitorering af elevers kompetencer eller fremskridt øger elevernes præstation (fx Black & Wiliam, 2009; Fuchs & Fuchs, 2002 og gennemgangen i afsnittet Hyppigheden af målinger).

TO TYPER AF PROGRESSIONSMODELLER

Veldefinerede og valide prøver bidrager med viden om, hvad enten den enkelte elev eller grupper af elever kan og har forstået på et givent tidspunkt. Men *enkelstående tests* kan ikke fortælle noget om, hvilke fremskridt en elev eller en klasse har gjort. Enkelstående tests kan anvendes af skoleledere eller andre beslutningstagere til at give en indikation af, om den enkelte skole gør fremskridt i forhold til andelen af elever, der klarer sig tilfredsstillende. Hull (2007) beskriver to overordnede typer af progressionsmodeller, der anvendes ved enkeltstående tests:

1. Progressionsmodellen (*The Improvement Model*)
2. Præstationsindeksmodellen (*Performance Index Model*).

I den første model, 'progressionsmodellen', sammenlignes en klasses eller en kohortes præstationer med tilsvarende elevers præstationer fra en anden kohorte/klasse. Sammenligningen sker mellem to grupper af elever, som i teorien burde præstere ens. Fx, 55 pct. af eleverne i sidste års 4. klasse opnåede tilfredsstillende resultater (bestemt ved en grænseværdi for karakterer, score eller lignende), hvorimod kun 52 pct. af eleverne i dette års 4. klasse fik tilfredsstillende resultater. De nuværende 4.-klasses-elever har klaret sig ringere end tilsvarende elever sidste år, og skolen har derfor ikke opnået fremskridt i forhold til andelen af elever, der klarer sig tilfredsstillende.

Denne sammenligningsmodel/progressionsmodel evaluerer ikke den enkelte elevs eller klasses fremskridt, da den ikke måler den samme kohorte af elever kontinuerligt, men i stedet sammenligner på tværs af kohorter. Modellen er derimod relativ simpel, nem at implementere og forstå. På trods af, at den ikke måler individuelle elevers fremskridt, kan modellen godt give en indikation af, hvorvidt flere eller færre elever på forskellige klassetrin klarer sig tilfredsstillende fra år til år. Ændringen, der observeres fra den ene kohorte til den anden kohorte, kan være et resultat af mange andre ting end undervisningen alene (Hull, 2007). Forskellene mellem kohorternes præstationer kan afspejle, at niveauet mellem kohorterne af elever i udgangspunktet er forskelligt.

Denne type progressionsmåling anvendes, når succesfulde skoler skal adskilles fra knap så succesfulde skoler, fx i USA's 'No Child Left Behind Act' (NCLB).⁸ Skolers præstation og progression afgøres på basis af proportionen af elever inden for et bestemt klassetrin, der opnår tilfredsstillende resultater. Ved indførelse af NCLB blev det pålagt, at forskellige stater skulle fastsætte egne passende mål for årlige fremskridt. Fordelene ved fremskridtsmodellen er, at resultaterne er lette for offentligheden at forstå, samtidig med at modellen identificerer *nogle* af de succesfulde skoler og nogle af de mindre succesfulde skoler. Metoden er derimod blevet kritiseret for ikke at være i stand til at beskrive de skoler (eller elever), som ligger lige omkring grænseværdien. Et andet kritikpunkt ved denne type sammenligning er, at skolers/elevers udgangspunkt ikke er medregnet, hvilket resulterer i, at nogle skoler/elever vil have let ved at opnå det fremsatte mål, mens andre skoler skal gøre betydelige fremskridt for at nærme sig målet (Heck m.fl., 2006).

I en artikel af Allen og Hancock (2016) undersøges effekten af læreres brug af individuelle kognitive profiler til elever i undervisnings-sammenhænge. Selve artiklens problemstilling er mindre interessant for dette afsnit, men måden, hvorpå de måler elevernes fremskridt, er relevant. Allen og Hancock benytter sig af en metode, som lægger sig tæt op

8. No Child Left Behind (NCLB) var en reform, der blev vedtaget i 2002 i USA. Formålet med reformen var at sikre, at alle børn, med særligt fokus på de fattigste, havde mulighed for at vælge en skole, som levede op til nationale standarder.

ad progressionsmodellen. Elevernes præstation i undersøgelsen, og dermed effekten af intervention, blev opgjort ved ”*Benchmark grade level expectations*”. Ved brug af denne metode eksisterer der et overordnet mål for, hvordan man forventer, at elever klarer sig fagligt, når de bliver stillet overfor en tekst (Allen & Hancock, 2016). Andelen af elever, der opnåede benchmark grade level expectations i interventionsgruppen, blev sammenlignet med andelen, der opnåede benchmark grade level expectations i kontrolgruppen.

I den anden progressionsmodel, som Hull (2007) beskriver, er målet at sammenligne flere grupper end blot elever, der opnår en fremsat grænseværdi, og elever, der ikke gør. Ved brug af denne model inddeles eleverne i mindre sammenlignelige grupper, sammenligningen sker stadig på tværs af klassetrin fremfor at sammenligne de samme elever over tid. Ved at inddele elever i fx fire grupper: 1. *Fremragende præstation*, 2. *God præstation*, 3. *Jævn præstation*, 4. *Utilstrækkelig præstation* sammenlignes andelen af elever, der flytter sig fra den ene gruppe til den anden. Fordelen ved at lave flere inddelinger er, at de fremskridt, som de svageste elever gør, også godskrives. En mindre grov inddeling af elever skulle også kunne afhjælpe, at lærerne udelukkende fokuserer deres undervisning og kræfter på de såkaldte ’bubble kids’, eller at ledere fokuserer på ’bubble-schools’. Bubble kids/schools er defineret som elever (eller skoler), der scorer lige *under* eller lige *over* tilfredsstillende.

Et svensk projekt ved navn DigiLys, som havde til formål at udvikle læringsmiljøet i skolen, anvendte en lignende inddeling af eleverne. Elevernes præstationer blev anvendt som screeningværktøj for læringsmiljøet. Elevernes resultater inddeles i tre kategorier: en grøn kategori, som indikerer et tilfredsstillende resultat, gul kategori, som indikerer, at der er plads til forbedring, og rød som indikation af, at det er nødvendigt at lave en indsats. Elevernes præstationer bliver sat i relation til en normalfordelingskurve, så det er synligt, hvordan en lærers elever klarer sig. Intentionen er at skabe et gennemsigtigt sammenligningsgrundlag, som gør det muligt at se mønstre i forskellige grupper og dermed vurdere, hvordan læringsmiljøet fungerer (Törner, 2015). Præstationsindeksmodellen er velegnet, når skoler skal vurdere, om de formår at flytte de fagligt svageste elever, men ligeledes udfordre de fagligt dygtigste elever.

Målinger henover tid (før- og eftermålinger eller målinger foretaget kontinuerligt henover en periode) kan beskrive en enkelt elevs progression. Målingerne kan give en indikation af, hvorvidt en elev udvikler sig i den rigtige retning, og om eleven opnår de faglige mål, der er sat. En simpel sammenligning af før-/eftermåling anvendes ofte, når effekten af interventioner, som har til formål at kvalificere læreres undervisning eller støtte elevens indlæring, skal evalueres. Før- og eftermålinger er fx anvendt i mange randomiserede kontrollerede forsøg, hvor effekten af indsatser og interventioner bliver evalueret. I denne type af forsøg sammen-

lignes også fx læsefærdighederne hos eleverne i kontrolgruppen med læsefærdighederne hos eleverne fra interventionsgruppen. Begge gruppers læsefærdigheder måles ved hjælp af tests henover tid (se fx Connor m.fl., 2013). Denne form for progressionsmåling er meget simpel og implementeres derfor ofte, hvis elevers faglige fremskridt skal evalueres. En af ulemperne ved denne type af model er, at det kan være svært at udvikle sammenlignelige tests, der er i stand til at måle, hvorvidt et barn har gjort faglige fremskridt. En anden ulempe er, at målinger i sig selv ikke nødvendigvis beskriver, hvorvidt eleven gør de fremskridt, der forventes, især hvis testene resulterer i en score (Hull 2007). Hvis denne type progressionsmodel skal være anvendelig, skal lærerne have et funderet grundlag, som de kan sammenligne elevernes resultater med.

MÅLINGER MED INDIVIDUELLE PROGRESSIONSMÅL

Simple tilgange til progressionsmåling anvendes hyppigt, selvom der eksisterer andre, mere avancerede, måder at måle progression på, hvor også elevernes forskellige udgangspunkt inkluderes. I Seltzer, Choi og Thum (2003) var formålet at undersøge, om der eksisterede en sammenhæng mellem elevers faglige udgangspunkt og hastigheden af deres faglige fremskridt. Hvis elevers udgangspunkt er afgørende for, hvor hurtigt de gør fremskridt, bør forskellige elevers succeskriterier for progression variere. Et standardiseret mål og ideal for elevers progression kan derfor opfattes som utilstrækkelig. Et standardiseret mål for progression forudsætter, at alle elevers progression er/bør være det samme. Der eksisterer progressionsmodeller, som medtager elevers udgangspunkt, når deres fremskridt skal evalueres. Modellerne er ofte mere statistisk avancerede, hvilket kan være årsagen til, at de ikke anvendes så hyppigt som de mere simple modeller. For at vurdere elevernes progression medtages elevernes tidligere præstationer for at forudsige, hvor meget en elev skal præstere det følgende år. Almindeligvis inddrager man i beregningen, hvordan tidligere lignende elever, med samme historik i forhold til faglighed, har gjort fremskridt (Hull, 2007).

Dette resulterer i, at elevernes individuelle fremskridt kan evalueres på en retfærdig måde, selvom de fagligt har forskelligt udgangspunkt og læringskurve. Hvis eleven flytter sig mere end forventet, har eleven gjort positive fremskridt, hvis eleven gør de forventede fremskridt, har eleven opnået det forventede, men hvis eleven ikke har opnået det forventede, har eleven gjort negative fremskridt. Så selvom eleven gør fremskridt, er det ikke sikkert, at de gør de fremskridt, de ”bør”. Denne måde at måle progression på kan både anvendes af beslutningstagere på et mere overordnet niveau, men lærere kan også inddrage metoden i planlægningen af deres undervisningsforløb.

Lærerne har ofte en ide om, hvordan elever bør bevæge sig gennem de forskellige undervisningstemaer, men med individuelt fremsatte

mål for hver enkelt elev får lærerne et tydeligt og gennemsigtigt indblik i, hvordan undervisningen påvirker elevernes læring (OECD, 2013). Hvis samtlige elever i en klasse har opnået deres individuelt fremsatte mål, har læreren formået at differentiere undervisningen tilstrækkeligt, og alle eleverne får udnyttet deres potentiale. Hvis der derimod er en gruppe elever, som ikke har nået deres mål, kan det indikere, at undervisningen enten ikke formår at trække de fagligt svage frem eller udfordre de fagligt stærke forventeligt. Selvom denne type af progressionsmåling kan give en indikation af lærerens undervisningskvalitet, bør det understreges, at elevernes fremskridt også bliver påvirket af andre faktorer, der ikke relaterer sig direkte til lærerens undervisning. Hvordan elever klarer sig, påvirkes også af fx elevens skolemiljø, elevens sociale miljø og forhold i elevens hjem. De progressionsmål, der sættes for en elev, er et resultat af, hvordan lignende elever tidligere har klaret sig, og hvilke mål, den elev har opnået, men dermed ikke sagt, at eleverne oplever samme undervisning eller samme miljø.

For at denne type af mere avancerede progressionsmodeller skal fungere godt, kræver det derfor, at en skole har adgang til tilstrækkelig med data for at lave en god forudsigtelse om elevens progression. At bruge en mere avanceret statistisk model kræver også flere kompetencer. De mere retvisende sammenligninger er dermed en mere ressourcekrævende metode.

VISUALISERING

I dette afsnit vil vi forsøge at belyse, *hvordan data visualiseres bedst, så de giver det rigtige grundlag for beslutningstagen*. Det er forventet, at data anvendes i skole- og undervisningsplanlægning, hvor skoleledere, lærere og forældre skal benytte sig af data til at fremme elevens præstation og trivsel. Skoler indsamler meget data om eleverne, men har på trods af dette været beskyldt for at være informationsfattige (Wayman, 2005). En af årsagerne til dette kan være, at store mængder af data ikke er tilgængelige i et format, som muliggør, at enten skoleledere eller lærere kan anvende eller tolke data, så det giver værdi i det videre arbejde med fx undervisnings- eller skoleplanlægning.

I følgende afsnit præsenterer vi, hvilke erfaringer der er beskrevet fra udlandet i forhold til visualisering af data. Søgningen frembragte ingen artikler med fokus på visualisering af *ikke-faglige* kompetencer. Det har derfor ikke været muligt at beskrive, hvilke erfaringer der er, når fx trivselsmålinger skal visualiseres. Følgende er udelukkende koncentreret om præsentation og visualisering af elevens *faglige* præstation, progression og resultater.

Der er udgivet overraskende lidt litteratur siden år 2000, som beskriver, hvordan data bedst visualiseres, så de giver et godt grundlag for planlægning og beslutningstagen. Den sparsomme litteratur på området og de manglende undersøgelser er gældende, uanset hvilken gruppe af aktører, man fokuserer på. Størstedelen af litteraturen omhandler, hvordan data skal præsenteres og illustreres, så *lærere* er i stand til at tolke resultater og data generelt. Goodman og Hambleton (2004) beskrev, at der var forbløffende få guidelines, hvis man ønskede at afrapportere testresultater i skole regi på den mest hensigtsmæssige måde. Den begrænsede litteratur på området kan indikere, at der fortsat er et behov for at undersøge emnet yderligere.

SIMPEL OG LETLÆSELIG VISUALISERING AF DATA

Alverson m.fl. (2014) udførte et studie, hvor formålet var at evaluere, hvordan forskellige aktører (heriblandt lærere og administratorer) foretrak, at elevdata blev præsenteret. Informanterne i studiet blev præsenteret for fire forskellige simple søjlediagrammer og blev bedt om at udpege, hvilket diagram de mente var det bedste til at beskrive data. For lærerne, som indgik i studiet, var det vigtigt, at et diagram var nemt, pålideligt og bekvemmeligt. Begreberne indikerer, at lærerne har et behov for hurtigt at kunne gennemskue, hvilke informationer diagrammer indeholder. En anden fordel ved det fortrukne diagram, som lærerne fremhævede, var, at denne type diagramtype var en type, som de havde set før, og den var derfor vant til at aflæse. I samme studie blev en gruppe administratorer interviewet og en gruppe forældre. Begge grupper fremhævede, ligesom gruppen af lærere, at det var vigtigt, at et diagram var let forståeligt og indeholdt få budskaber, da mange informationer samlet i ét diagram ofte forvirrer mere, end det gavner. Simple og letlæselige grafer og diagrammer er et gennemgående tema i de inkluderede studier.

Goodman og Hambleton (2004) er en af de få undersøgelser, som kommer med konkrete forslag til, hvordan testdata bør rapporteres. Selvom deres anbefalinger er udarbejdet med henblik på amerikanske scorerapporter (*score reports*), hvor tekst også indgår, kan anbefalingerne være brugbare ved rapportering af elevtests i en dansk kontekst. Deres forskningskortlægning resulterede i følgende ni anbefalinger, når testresultater skal præsenteres:

1. Præsentationer af elevers resultater bør være klare, kortfattede og visuelt interessante
2. Præsentationer bør inkludere en letlæselig tekst, som støtter og forbedrer tolkningen af diagrammer og tabeller
3. Datavisninger (*data displays*) skal designes, så de udfylder mindre håndgribelige formål, og man må ikke forsøge at udtrykke for meget ved hver enkelt præsentation

4. Værktøjer såsom bokse og grafiske fremstillinger skal kun anvendes til at fremhæve hovedfund
5. Ved præsentation af data bør den måde, hvorpå data grupperes, være meningsfuld og klart defineret
6. Tekst i lille størrelse, fodnoter og statistiske begreber bør undlades
7. Nøglebegreber bør defineres i en ordliste, som nemt kan findes af brugere
8. Man bør foretage pilottest af præsentationerne, så det intenderede publikum kan komme med kommentarer og input
9. Præsentationer bør specialdesignes til forskellige brugere, så de opfylder de forskellige behov, som de forskellige brugere har. Fx kan en lærer have gavn af detaljerede præsentationer på individniveau, hvorimod skoleledere har mere gavn af at se præsentationer på klasseniveau. Yderligere kan forældre have gavn af en præsentation, der er mere simpel end den, en lærer ville have gavn af.

Ovenstående anbefalinger af Goodman og Hambleton (2004) underbygger, at fremstillinger af data er mest anvendeligt, når det holdes simpelt med få informationer.

EKSEMPLER PÅ BRUGEN AF DATAVISUALISERING

På trods af, at flere artikler beskriver problematikken om, at afrapporteringer ofte er for avancerede og komplicerede, beskriver meget få studier helt konkret, hvordan data ellers bør fremstilles og visualiseres (Goodman m.fl., 2004; Rankin m.fl., 2013). Ved gennemgangen af de relevante studier, som blev identificeret i søgningen, finder vi dog eksempler på, hvordan elevdata anvendes til at informere og guide aktører. I følgende præsenteres et mindre udsnit af eksempler på, hvordan data kan bruges.

Zapata-Rivera m. fl. (2012) har udgivet en artikel, hvor formålet var at udvikle og evaluere brugen af scorerapporter (*score reports*). Scorerapporter er et evalueringsværktøj, der ofte anvendes i USA, når elever tager tests, svarende til de danske nationale tests. Scorerapporterne bliver anvendt af både elever, forældre, skoleledere og politiske beslutningstagere. Scorerapporterne kan være udformet forskelligt, men præsenterer enten en enkelt elevs eller en samlet klasses præstation. Scorerapporterne indeholder både tekst og diagrammer og er ofte tilgængeligt online. I bilag 1 (<http://www.sfi.dk/elev-bilag1/>) vises eksempler af de scorerapporter, som blev evalueret i undersøgelsen. Tidligere undersøgelser har vist, at brugere af scorerapporter har svært ved at forstå og gøre brug af dem. I nævnte undersøgelse konkluderes det, at lærerne overordnet havde positive tilbagemeldinger vedrørende det visuelle udtryk og brugen af scorerapporter. Zapata-Rivera m.fl. (2012) beskriver, som mange andre, at scorerapporterne og de informationer, de skal levere, ikke er simple nok. Lærere vælger ofte at ignorere lange forklarende sætninger i

præsentationer af elevdata og foretrækker korte præcise informationer. Yderligere beskriver de i studiet, at lærerne ofte havde svært ved forstå hovedprincipperne bag de statistiske beregninger i scorerapporterne. I OECD (2013) fremlægges et eksempel på, hvordan nogle skoler i Holland siden midten af 80'erne har anvendt et værktøj kaldet *Leerling Volg Systeem* (forkortes LVS, oversat til dansk elevvurderingssystem). Værktøjet baseres på standardiserede longitudinelle tests af eleverne, som skal måle elevernes præstation og fremskridt i forskellige fag. Resultaterne fra testene bliver præsenteret i en graf, som visualiserer elevernes fremskridt gennem årene, og de kalder denne præsentation af data for en elevrapport. Nationale undersøgelser bliver anvendt som referencegruppe, og elevernes elevrapport viser, hvor en elev placerer sig i forhold til landsgennemsnittet. Hos elever, som modtager specialundervisning, fremlægges der ligeledes en graf, som sammenligner eleven med elever uden særlige behov. Et eksempel på en rapport fremgår af bilag 2. (<http://www.sfi.dk/elev-bilag2/>).

Der eksisterer altså et behov for helt tydeligt at fastlægge, hvordan data bedst visualiseres, så det er tilgængeligt for de målgrupper, som kan drage stor nytte af den (Sabbah, 2011). Der er mangel på og et behov for studier, som evaluerer og undersøger brugen af forskellige grafiske præsentationer og andre rapporteringsformer i formidlingen af data til målgrupper såsom lærere, skoleledere og andre beslutningstagere (Zapata-Rivera m.fl., 2012). Der eksisterer nemlig en enighed om, at der ligger et uudnyttet potentiale i allerede eksisterende data, og at grafer og andre visualiseringsværktøjer kan være effektive, når elevens progression skal diskuteres. Visualisering af data kan være en hjælp i forhold til at identificere, hvornår der er behov for at ændre eller justere undervisning (Capizzi & Barton-Arwood, 2009).

I en nyligt udgivet rapport, hvor formålet var at evaluere lærere og leders brug af de nationale tests, var konklusionen, at lærere og leders manglende evne til at fortolke testresultater resulterer i, at de nationale testdata ikke bliver inkluderet i skole- eller undervisningsplanlægning i så høj grad, som man havde håbet (Bundsgaard & Puck, 2016). Korrekt visualisering og præsentation af eksisterende data, som fx de nationale tests, kan her bidrage til, at lærere og ledere oplever de nationale tests som anvendelige i deres daglige planlægning, idet testen derved peger på, hvor læreren skal have særlig opmærksomhed ift. at indsamle mere viden/data om eleven.

DET FAGPROFESSIONELLE ARBEJDE MED ELEVENES PROGRESSION

Dette kapitel omhandler forskellige aspekter af det fagprofessionelle arbejde med elevers progression. Kapitlet er inddelt i 4 afsnit, som fokuserer på henholdsvis processen fra progressionsmåling til tilpasning og udvikling af undervisning, it-understøttelse, dataforståelse og kompetencer samt den institutionelle kapacitet til at arbejde med progressionsmåling.

PROGRESSIONSMÅLING SOM EN DEL AF DET FAGPROFESSIONELLE ARBEJDE

I dette afsnit vil vi sætte fokus på processen fra progressionsmåling til tilpasning og udvikling af undervisningen. Afsnittet er centreret omkring de to forskningsspørgsmål om, *hvordan progressionsmåling kan indgå som en del af det fagpædagogiske arbejde med at fremme elevens faglige og alsidige kompetencer samt trivsel, og hvilke praktiske erfaringer der er*. Vi vil her opridsende tendenser omkring fagprofessionelles arbejde med data og progressionsmåling, baseret på flere forskningskortlægninger samt rapporter fra en både dansk og international kontekst.

Mange faktorer spiller ind, når man ser på skolers brug af progressionsmåling, herunder både nationale, kommunale og lokale forhold, såsom politiske uddannelsesmæssige interesser, kommunale ressourcer og prioriteringer samt skoleledere, lærere og pædagogers motivation for brug af data. I afsnittet Kendetegn ved skoler og kommuner, som benytter progressionsmåling, vil vi mere overordnet sætte fokus på kommuner og skolers institutionelle kapacitet i forhold til arbejdet med progressionsmåling. Dette afsnit vil derimod først og fremmest være centreret omkring lærere og pædagogers arbejde med elevdata.

FRA DATAINDSAMLING TIL TILPASNING OG UDVIKLING AF UNDERVISNING

Inden for litteraturen om data og progressionsmåling er der flere forskningsoversigter og rapporter, som har fokus på at formulere praksisguides til skolers brug af data, og der er her mange lighedspunkter på tværs af studier (fx Hamilton m.fl., 2009; Means m.fl., 2010 og Ministry of Education in Ontario, 2011). Ulempen ved disse praksisguides er dog, at anbefalinger ofte er så generelle, at de kan være svære at inddrage i praksis.

Nogle af de gennemgående pointer i disse praksisguides er vægtingen af, at der afsættes tid til, at lærere i fællesskab kan analysere elevdata, udveksle undervisningserfaringer og reflektere over egen undervisning og på baggrund heraf træffe beslutninger om den fremadrettede tilrettelæggelse af undervisning. En anden vigtig pointe i studierne er også anbefalingen om at inddrage data løbende i det fagpædagogiske arbejde og dermed behandle progressionsmåling som en konstant proces (Hamilton m.fl., 2009; Means m.fl., 2010; Ministry of Education in Ontario, 2011 og OECD, 2013).

I forlængelse heraf bliver databrug og progressionsmåling i alle de ovennævnte studier illustreret som en cyklus i flere trin, der alle ligner hinanden i dets indhold. I forskningsoversigten udført af Uddannelsesministeriet i Ontario (2011) beskrives denne cyklus i fire trin: planlægning, handling, observation og refleksion.

Planlægningsfasen er rettet mod indsamling og analyse af elevdata for på den baggrund at afdække, hvor eleverne har behov for yderligere støtte og tilrettelægge undervisningen herefter. Hamilton m.fl. (2009) understreger desuden vigtigheden af, at denne analyse baseres på flere forskellige datakilder frem for at vægte resultaterne af en enkelt test. I begge forskningsoversigter anbefales det endvidere, at lærere her planlægger undervisningen med øje for slutresultatet. Dette indebærer, at lærere forholder sig til, hvilke data der kan belyse, om elevens behov er blevet mødt, og udvikler en evalueringsstrategi på baggrund heraf. Derudover anbefales det, at lærere anvender forskningsbaserede undervisningsstrategier (Hamilton m.fl., 2009; Ministry of Education in Ontario, 2011).

I *handlingsfasen* implementeres de udvalgte undervisningsaktiviteter, som efterfølgende justeres efter behov på baggrund af løbende dataindsamling og elevfeedback. Det anbefales her, at lærere gør brug af professionel sparring for på den måde at opbygge en fælles forståelse omkring undervisningstilgange.

Observationsfasen har fokus på at monitorere både elevs og lærers læring. Det anbefales, at lærere deler elevdata, herunder elevfeedback, for i fællesskab at analysere, hvorvidt der er sket en progression i den enkelte elevs læring, og på baggrund heraf vurdere næste skridt. Begge forskningsoversigter fremhæver desuden vigtigheden af, at lærere deler undervisningserfaringer og -udfordringer med hinanden for herved at udvikle og

skabe en fælles praksis for undervisning og vurdering af elevers præstation (Hamilton m.fl., 2009; Ministry of Education in Ontario, 2011).

I *refleksionsfasen* analyseres og evalueres elevers resultater og samlede arbejde, og baseret på evidens bliver det her vurderet, hvorvidt elevens behov er blevet mødt, og lærere kan på denne baggrund også reflektere over egen læring. På baggrund af den samlede vurdering besluttes næste skridt i undervisningen (Ministry of Education in Ontario, 2011).

FÆLLES KULTUR OG KLARE MÅL OMKRING DATABRUG

For at data kan anvendes mest effektivt, argumenterer flere studier desuden for nødvendigheden af, at der på skoler og i distrikter opbygges en kultur omkring databrug (fx Hamilton m.fl., 2009; Ministry of Education in Ontario, 2011; OECD, 2013 og Wayman, Jimerson & Cho, 2012). En datakultur defineres her som et læringsmiljø, der inkluderer holdninger, værdier, mål og normer omkring undervisningspraksis fulgt op af en eksplicit vision for databrug samt en anerkendelse af dataindsamling som en nødvendig del af de fagprofessionelles arbejde.

Hamilton m.fl. (2009) anbefaler i den forbindelse, at målinger følges op af en klar vision for databrug på den enkelte skole. Dette kan blandt andet understøttes ved at etablere et datateam på tværs af skolens ansatte, som kan sikre input fra alle led, og som løbende kan udbyde vejledning til alle ansatte i brug af data i tråd med skolens vision. Ifølge Hamilton m.fl. bør datateamet endvidere skabe og udbrede en fælles forståelse omkring læringskoncepter såsom progression, læring og evidens. Herudover bør teamet formulere en skriftlig plan med relevante og realistiske mål for, hvordan skolen vil anvende data til at understøtte den enkelte skoles mål om at forbedre elevers præstationer (Hamilton m.fl., 2009).

I Danmark indførte man i 2003-2006 Fælles Mål i folkeskolen, som i 2015 ændrede format til de nuværende Forenklede Fælles Mål. Formålet med de Forenklede Fælles Mål er ligeledes at understøtte fagprofessionelles arbejde med elevers læring og progression ved at opstille klare kompetencemål for, hvad eleverne skal kunne på de enkelte klassetrin i de forskellige fag. Der lægges i den forbindelse op til, at lærere og pædagoger kan arbejde læringsmålstyret med tilrettelæggelsen af undervisningen. Med de fælles nationale læringsmål ønskede man altså at give både fagprofessionelle, forældre og eleverne selv et klart billede af, hvad det forventes, at eleverne skal lære. I forlængelse af anbefalingerne fra Hamilton m.fl. kan disse mål være en mulig hjælp til skolers fremadrettede arbejde med elev-data (Ministeriet for børn, undervisning og ligestilling, 2016c).

FORMATIVE OG SUMMATIVE VURDERINGER AF ELEVERS LÆRING

Som tidligere belyst er der i dag et større fokus på, at vurderinger af elevers læring i højere grad bør anvendes formativt i tilrettelæggelsen af un-

dervisning og generelt en øget tendens mod brug af formative tests (fx Means m.fl., 2010 og OECD, 2013).

Som eksempel på formative vurderinger i en dansk kontekst fremhæver man i OECD (2013) den danske brug af elevplaner (*individual student plans*), som blev indført i folkeskolen i 2006. Elevplanen har fokus på at dokumentere elevens progression med vægt på elevens fremtidige læring frem for summative resultater af læring. En sådan elevplan kan netop understøtte læreres arbejde med elevens progression og har også fungeret som et godt redskab i læreres kommunikation med forældre. En af udfordringerne ved elevplanen er dog den ekstra arbejdsbyrde, som den tilfører lærerne (OECD, 2013).

En evaluering af de nationale tests i den danske folkeskole, udført af Rambøll i 2013, pointerede endvidere, at testene i lovgivningen også var tiltænkt til formativ (fremadrettet) brug i tilrettelæggelse af undervisningen (Rambøll, 2013). I evalueringen konkluderes det dog, at både lærere og ledere i overvejende grad benytter testresultaterne summativt – til at skabe overblik og monitorere det faglige niveau blandt elever, klasser og skoler som helhed. Evalueringen konkluderer endvidere, at kommunerne anser de nationale tests som en væsentlig indikator til at vurdere skolernes faglige resultater.

ELEVERS INDDRAGELSE I EGEN LÆRING

I flere forskningsoversigter og rapporter fremhæves det endvidere, at elever i højere grad selv bør inddrages i egen læring (Hamilton m.fl., 2009; Means m.fl. 2011; OECD, 2013). Hamilton m.fl. (2009) samt OECD (2013) anbefaler således, at lærere sætter tid af til at træne eleverne i, hvordan de kan bruge data og resultater til løbende at reflektere og kritisk analysere deres egen indsats i klasselokalet og opstille egne mål for fremtidig læring. Det pointeres her, hvordan en sådan analyseproces kan være med til at motivere eleverne ved at opsætte mål, som er realistiske at nå, belyse, hvor der er sket fremskridt samt give eleven en følelse af kontrol over egen læring (Hamilton m.fl., 2009; OECD, 2013). Hamilton m.fl. (2009) formulerer i den forbindelse anbefalinger til, hvordan dette kan implementeres i praksis. Det indebærer blandt andet, at lærere grundigt forklarer, hvilke forventninger der stilles til eleverne, og hvilke kriterier de bedømmes ud fra. Herudover, at lærere følger resultaterne op med rettidig, konkret og konstruktiv feedback til den enkelte elev samt værktøjer til, hvordan eleven kan lære af denne feedback. Dette kunne fx indebære at afsætte undervisningstid til, at elever kan tolke og lære fra deres egen data (Hamilton m.fl., 2009). I OECD (2013) anbefales det ligeledes, at elever, ud over at vurdere deres eget arbejde, også roterer og vurderer hinandens arbejde.

UDFORDRINGER I ARBEJDET MED PROGRESSIONSMÅLING

I gennemgangen af forskningsoversigter, deskriptive studier og praksiserfaringer, der omhandler arbejdet med progressionsmåling, er der dog også flere udfordringer og implikationer for praksis, som går igen.

MANGLENDE VIDEN OM, HVAD DER VIRKER

En grundlæggende problematik i arbejdet med progressionsmåling er det forhold, at værdien af data er bredt anerkendt i forskningen, mens vores viden om, hvordan brugen af data til tilrettelæggelse af undervisning kan få den største effekt på elevers læring, stadig er begrænset (fx Hamilton m.fl., 2009; Means m.fl., 2010; Ministry of Education in Ontario, 2011; Slavin m.fl., 2013). Der findes således mange specifikke casestudier og deskriptive studier, som har fokus på, hvordan skoler kan bruge data mest effektivt, og mange af anbefalingerne refereret i denne kortlægning er dermed også baseret på denne type af studier. Omvendt er der færre studier, som reelt har fokus på koblingen mellem effektiv databrug og elevudfald i form af øget læring. Dette skyldes blandt andet, at det kan være svært at finde en passende kontrolgruppe eller sammenligningsgrundlag, der kan belyse, hvad der ville være sket uden en given indsats. Samtidig kan det være svært at isolere effekten af en given indsats, idet mange studier indebærer flere former for indsatser, alt imens mange omkringliggende faktorer ligeledes spiller ind (Hamilton m.fl., 2009). I en forskningskortlægning om resultatbaseret styring udgivet af KORA i indværende år konkluderes det desuden, at der blandt de studier, som faktisk måler på elevlæring, identificeres en korttids-, men ingen langtidseffekt af skolers brug af resultatbaseret styring (Møller, Iversen & Andersen, 2016). Der mangler således mere viden omkring effekten af datainformeret skoleudvikling.

RISIKO FOR AT "LÆRE TIL TESTEN" OG SKABE MÅLFIKSERING

En anden problematik omhandler faren for, at skolernes undervisning bliver rettet mod at "lære til testen" (*teaching to the test*), mens aspekter af læring, som er mindre målbare, mister fokus (fx Møller, Iversen & Andersen, 2016 og OECD, 2013). Dette kan resultere i forskydning af fokus i undervisningen, fx ved at der allokeres ekstra ressourcer til de undervisningsområder, som testes, at elever trænes i specifikke typer af spørgsmål eller udsættes for gentagne træningstests (OECD, 2013). En yderligere risiko er desuden den, at resultater fra enkeltstående tests tillægges for stor betydning, eller at der direkte manipuleres med præsentationen af resultater (Deming & Figlio, 2016; Jacob & Levitt, 2003). Der kan her ligge en udfordring i, at aktører i en datainformeret styringssammenhæng kan være mere motiveret af at opfylde mål på papiret frem for reel målopfyldelse (Møller, Iversen & Andersen, 2016; OECD, 2013). Det tidligere refererede danske studie af Bundsgaard og Puck (2016)

omhandlende lærere og skolelederes anvendelse af de nationale tests er i denne henseende interessant. Her fandt man, at der blandt lærerne var en gennemgående oplevelse af, at testens resultater tillægges for stor betydning af særligt den kommunale forvaltning og til dels også af skoleledelsen. Studiet fandt desuden, at en stor del af lærerne ikke har oplevelsen af, at de nationale tests er et værdifuldt redskab, der giver et kvalificeret bidrag til deres viden og fremadrettede professionelle handlen⁹ (Bundgaard & Puck, 2016). I Rambølls evaluering af de nationale tests (2013) fandt man et meget differentieret billede af, hvorvidt der var tendenser på skolerne til at lære til testen. Man fandt her, at dette var mest udbredt i udskolingen og på skoler, hvor ledelsen har fokus på faglige resultater (Rambøll, 2013).

OECD (2013) anbefaler i deres samlede rapport, at man opretter sikkerhedsinstanser mod overdreven vægtning af specifikke standardiserede tests, samt at man generelt inddrager en række forskellige informationer og data om elever, når der skal foretages vurderinger om elevers progression (OECD, 2013). Flere studier understreger i den forbindelse vigtigheden af, at læreres individuelle skøn og vurdering af elever fortsat vægtes som en helt central del af arbejdet med elevers progression. Herved skabes mulighed for, at kompetencer, som ikke er målelige i standardiserede vurderinger, fortsat gives plads (Ministry of Education in Ontario, 2011; Møller, Iversen & Andersen, 2016; OECD, 2013).

Møller, Iversen og Andersen (2016) problematiserer endvidere, at brugen af resultatbaseret styring generelt kan risikere at føre til målfiksering og tjeklisteadfærd, frem for et redskab til refleksion og forbedring af den faglige omsætning af mål til praksis. Arbejdet med progressionsmåling kan dermed blandt fagprofessionelle opfattes mere som kontrol end som et redskab til læring (ibid.). Uddannelsesministeriet i Ontario anbefaler i den forbindelse i deres forskningsoversigt blandt andet, at lærere inden for den enkelte skole eller det enkelte lærerteam får størst mulig autonomi til selv at udvælge de data, som de finder mest relevant i arbejdet med elevers progression. Som tidligere nævnt vægter de her ligeledes skolers opbygning af en fælles datakultur med klare visioner for databrug og forståelser for dets begrænsninger (Ministry of Education in Ontario, 2011).

DISKUSSION AF MÅLENES KARAKTER

Som tidligere nævnt i dette afsnit, pointerer mange studier vigtigheden af, at der i arbejdet med progressionsmåling opstilles klare mål for databrug, og at ansatte fra alle faggrupper er med til at formulere disse mål for at

9. Fx er under 30 pct. af lærerne enige i de positive udsagn om, at testen er et godt udtryk for elevernes faglige kunnen, at den giver lærerne værdifuld viden, de ellers ikke ville have haft, eller at den hjælper dem til at målrette undervisningen til den enkelte elev eller til at tilrettelægge den videre undervisning (Bundgaard & Puch, 2016).

sikre deres implementering. Dette var ligeledes en af konklusionerne i Møller, Iversen og Andersen (2016). I litteratursøgningen fandt man dog kun et enkelt studie, der teoretisk fokuserer på betydningen af målenes karakter – altså hvorvidt karaktergennemsnit eller testresultater er et passende mål for resultatopnåelse. Dette til trods for, at mange af de inkluderede studier benytter eksamenskarakterer som måleenhed og opstiller eksamenskarakterer som indikator på effekten af resultatbaseret styring. Det formodes i studiet, at dette kan skyldes, at der på grundskoleområdet er etableret en form for konsensus om, at eksamenskarakterer kan fungere som et validt udfaldsmål på målopfyldelse (ibid.). I studiet nævnes kvalitetsrapporter, evalueringer, budgettal og forældretilfredshedsundersøgelser som andre brugte mål, men det problematiseres dog, at man på skoleområdet, i forhold til fx socialområdet, har en mere begrænset opfattelse af, hvordan elevlæring og skolepræstationer kan måles. I den forbindelse stiller forfatterne her spørgsmålstegn ved, hvorvidt tilstedeværelsen af bestemte data fører til deres brug, frem for fx udvikling af mere velegnede indikatorer eller indsamling af andre data (ibid.). I kortlægningen fandt man ligeledes overraskende få studier, som fokuserer på målingen af aspekter som tværgående kompetencer og trivsel, hvilket ligeledes kunne tyde på en ensretning i forhold til, hvordan vi definerer og måler elevers læring og progression.

TID TIL IMPLEMENTERING OG SAMARBEJDE PÅ TVÆRS

På tværs af studier fremhæves vigtigheden af, at kommune og skoleledelse afsætter struktureret tid til, at lærere og pædagoger kan opbygge datakompetencer og analysere data i fællesskab inden for lærerteams eller fælles klassetrin. I forlængelse heraf, at der skabes muligheder for at sparre på tværs, dele undervisningserfaringer og bruge disse input til at sammensætte og tilpasse den fremadrettede undervisning. I en nylig rapport fra KORA om pædagogiske medarbejderes oplevelser og erfaringer med folkeskolereformen i 2014/2015, fremgår det, at en af de største udfordringer er, at lærere og pædagoger oplever at have svært ved at finde tilstrækkelig tid til forberedelse af undervisningen med de nye arbejdstidsregler (Bjørnholt m.fl., 2015). Samme rapport viste omvendt, at flere skoleledere mener, at lærernes forberedelsestid er tilstrækkelig (ibid.).

De gældende arbejdstidsbestemmelser skaber mulighed for medarbejdernes fælles tilstedeværelse på skolen og rummer dermed et potentiale for fagprofessionelt samarbejde om arbejdet med elevdata. Undersøgelsen her belyser dog samtidig en ledelsesmæssig udfordring i forhold til at skabe denne strukturerede tid til, at lærere og pædagoger kan analysere elevdata i og på tværs af faggrupper og opbygge datakompetencer. Selvom behovet for, at fagprofessionelle investerer mere tid i starten, kan være en udfordring nu og her, indeholder den datainformerende skole dog også muligheder på sigt. Det skyldes, at en bedre adgang, anvendelse og

visualisering af eksisterende data, og øget brug af digitale læremidler mere generelt, har potentiale til at forenkle fagprofessionelles arbejde. Det gælder fx i forhold til funktioner, som gør det muligt at skabe en differentieret undervisning, tilbyde samlede digitale undervisningsforløb og skabe nem adgang til at give feedback og rette elevers opgaver digitalt (Rambøll, 2014b).

IT-UNDERSTØTTELSE AF DET FAGLIGE ARBEJDE MED DATA

I dette afsnit sætter vi fokus på forskningsspørgsmålet omhandlende, *hvordan man it-understøtter det faglige arbejde med progressionsmåling*. It-understøttelse refererer her til de datasystemer, som ligger til grund for indsamling og analyse af data om den enkelte elev såsom trivselsrapporter, fremmøde og karakterer. For at belyse dette forskningsspørgsmål og fremhæve eksempler på forskellige it-programmer trækker vi i afsnittet her på flere forskningsoversigter, rapporter fra en både dansk og international kontekst samt to effektstudier.

På tværs af studier er der et gennemgående fokus på, hvordan it-systemer kan skabe bedst mulig tilgængelighed til data samt understøtte fagprofessionelles analyse og tolkning af data (Connor m.fl., 2013; Hamilton m.fl., 2009; Rankin, 2013). De to effektstudier (Connor m.fl., 2013; Rankin, 2013), som er inddraget i dette afsnit, er netop eksempler på, hvordan it-systemer med fordel kan understøtte fagprofessionelle i deres anvendelse af data. Andre studier sætter derimod fokus på selve udviklingen af it-systemer, hvilke krav de bør leve op til, og hvordan de bedst muligt implementeres på skolerne (Hamilton m.fl., 2009; OECD, 2013). I afsnittet her vil vi først sætte fokus på, hvilke behov et it-system bør leve op til, og hvordan det kan implementeres for dernæst at belyse, hvordan it-programmer kan understøtte fagprofessionelles anvendelse af data på forskellig vis.

KRAV TIL IT-SYSTEMER OG DERES IMPLEMENTERING

I et omfattende studie fra det amerikanske uddannelsesministerium udført i 2007-2008 blev der sat fokus på skoler og distrikters kapacitet og brug af data (Means m.fl., 2010). Studiet var baseret på et spørgeskema blandt 427 skoledistrikter, casestudier af distrikter, som var særligt aktive i deres brug af elevdata, samt et lærerspørgeskema med 2.509 respondenter. I forhold til skolers it-systemer var et af studiets konklusioner her, at manglende kobling mellem forskellige datasystemer er en barriere for at kunne udvide skolens brug af data. Denne udfordring gjorde sig gældende for 60 pct. af de deltagende distrikter, mens distrikterne, som var del af casestudiet, i gennemsnit opererede med mellem 2-7 forskellige elektroniske datasystemer. Endvidere viste studiet, at de fleste distrikter

har data omkring elevers demografi og testresultater, mens langt færre distrikter har mulighed for at kombinere data fra forskellige datasystemer eller koble undervisningsaktiviteter til elevpræstationer. Studiet viste her, at under halvdelen af distrikterne havde datasystemer, som gør det muligt at koble udfaldsmål til praksis, som er en nødvendighed for at kunne skabe løbende forbedringer (ibid.).

Hamilton m.fl. (2009) bringer i den forbindelse anbefalinger til, hvordan man med fordel kan skabe og opretholde et fælles it-system på tværs af skoler inden for et givent distrikt. For at et sådant system kan oprettes og implementeres med succes, anbefaler Hamilton m.fl., at der i hvert distrikt oprettes et *it-panel*, som i fællesskab udvikler et datasystem, der kan imødekomme skolernes og brugernes forskellige behov. Et sådant panel må derfor også bestå af repræsentanter på tværs af interessegrupper såsom administrativt personale, teknisk personale, lærere, pædagoger, forældre og elever. Hamilton m.fl. anbefaler endvidere, at der på hver skole oprettes et *lokalt datateam*, som understøtter lærernes brug af data og skaber en fælles plan for databrug. For at udvikle et fælles it-system er det her nødvendigt, at it-panelet er i tæt kommunikation med en repræsentant fra hver skoles datateam. Målet her er derved at skabe et it-system, baseret på de behov, som brugerne reelt har i forhold til brug af data – ikke omvendt.

Hamilton m.fl. opstiller i den forbindelse nogle brugbare krav til it-systemer, herunder:

- Datasikkerhed og adgang til systemet – fx kan man kun tillade it-personale at taste og redigere data samt overveje at indføre forskellige grader af adgang til data.
- Et konsistent elev- og lærer-ID – således at den enkelte elev og hans eller hendes udvikling kan følges over tid og på tværs af skoler.
- Opbevaring af data – det er her vigtigt, at systemet kan inkorporere mange typer af data og skabe mulighed for, at alle data kan kobles og analyseres samlet.
- Sikring af datakvalitet og timing af data – data skal være rensset for fejl og indtastes umiddelbart efter indsamling, så undervisere har mulighed for at analysere og give feedback til elever kort tid efter.
- Løbende professionel udvikling for både brugere og it-personale – baseret på den enkeltes behov i forhold til fx tekniske kompetencer, undervisningsfelt og ansvar.

Flere af anbefalingerne i Hamilton m.fl. (2009) til et it-system går igen i flere andre studier. I OECD's samlede rapport (2013) fremhæves således også vigtigheden af *timing* i forhold til hurtig feedback af testresultater til både lærere og elever (OECD, 2013). Det anbefales i den forbindelse, at skoler benytter computerbaserede tests frem for papirbaserede. Dette

betyder, at lærere og elever kan få resultatet den følgende dag, frem for uger eller måneder senere. Dette muliggør i sidste ende en større brug af testresultaterne i forhold til at tilpasse undervisning og læring for at understøtte den enkelte elevs progression.

IT-UNDERSTØTTELSE AF DATAANALYSE OG -ANVENDELSE

Som nævnt i afsnittets indledning har flere studier et direkte fokus på, hvordan man med it-programmer kan understøtte læreres konkrete fortolkning og brug af data. Flere studier, i både dansk og international kontekst, påpeger netop, hvordan mange fagprofessionelle har problemer med at tolke data og efterspørger yderligere støtte i it-systemerne til at analysere data (fx Bundsgaard & Puck, 2016; Means m.fl., 2010; Rankin, 2013).

Et eksempel på, hvordan it-systemer konkret kan yde en sådan støtte kan blandt andet ses i et amerikansk effektstudie (Rankin, 2013). Studiet belyser, hvordan det at tilføje analyseredskaber i it-systemerne kan have en positiv effekt på nøjagtigheden af underviseres analyse af data. Rankin gjorde i studiet brug af tre forskellige værktøjer til at understøtte underviseres fortolkning af data. Det første redskab var her tilføjes af uddybende fodnoter ved præsentationen af data med information om, hvordan data bør læses. Herudover blev der ved afrapportering af data vedlagt et resumé af formål og fokus i den givne test samt mulige faldgruber ved tolkningen af dens resultater. Det sidste analyseværktøj i studiet var en guide til tolkning af data med fokus på konkrete eksempler på, hvordan dataresultater skulle tolkes.

Studiet involverede 211 lærere med forskellig baggrund og funktioner på 9 forskellige grundskoler i Californien. Data blev indsamlet via et elektronisk spørgeskema. Resultaterne viste, at lærernes analyse af data blev 307 pct. mere nøjagtig, når der blev tilføjet en fodnote til data, 205 pct. mere præcis, når et resumé var vedhæftet, og 273 pct. mere præcis, når en guide til fortolkning af data var vedlagt. Selv hvor deltagerne specifikt oplyste at have brugt analyseredskaberne, resulterede ingen af de tre redskaber dog i en 100 pct. nøjagtig analyse af data. Den samlede nøjagtighed af lærernes dataanalyse steg i gennemsnit op til 48 pct. (med et udgangspunkt på 11 pct. nøjagtighed før indsats).

Et andet eksempel på, hvordan it-systemer kan understøtte underviseres faglige arbejde med progressionsmåling, udvalgt på grund af dets omfattende størrelse, er en række amerikanske lodtrækningsforsøg udført af Carol M. Connor og medforfattere (se fx Connor m.fl., 2007, 2013). Studierne undersøgte effekten af et algoritme-baseret it-program med fokus på sammenhængen mellem vurdering og instruktion (A2i) i forhold til individuel læseundervisning. På baggrund af den enkeltes elevs præstationer skaber programmet blandt andet konkrete individuelle anbefalinger til læreren i forhold til, hvilken mængde og type af videre un-

dervisning, eleven fremadrettet bør modtage. Disse anbefalinger opdateredes månedligt.

Resultaterne viste generelt en signifikant positiv effekt på elevers progression i læsning ved brug af individuel læsevejledning. I det studie, der følger eleverne gennem længst tid (Connor m.fl., 2013), deltog 47 lærere og 616 1.-klasser-elever fra 10 forskellige skoler i et etnisk og økonomisk diverst skoledistrikt i Florida. Studiet bestod her af en kontrolgruppe, som ikke gjorde brug af it-programmet, A2i, i deres læseundervisning, samt en indsatsgruppe, som modtog oplæring i programmet og anvendte dette i deres undervisning. Både indsats- og kontrolgruppe blev bedt om dagligt at afsætte 90 minutter af undervisningen til læseundervisning i små grupper.

Studiet havde dog også visse komplikationer, idet it-programmet og brugen af individuel læsevejledning var en udfordring for flere lærere på trods af støtte fra professionelle. De lærere, der gjorde mest brug af programmet, havde omvendt også en tendens til at implementere den individuelle vejledning af elever bedre i klasseværelset.

Hvor effektstudiet af Rankin (2013) altså har fokus på, hvordan man via it-systemer kan støtte og forbedre læreres *tolkning* af data, sætter effektstudiet af Connor m.fl. (2013) først og fremmest fokus på, hvordan et it-program kan understøtte læreres *anvendelse* og *handlen* på baggrund af data. Begge studier er således relevante eksempler på, hvordan it-systemer på forskellig vis kan understøtte fagprofessionelle i deres arbejde med progressionsmåling. Der er i øvrigt ikke, i vores udvalg af studier, mange eksempler fra grundskolen på den type af mere sofistikerede hjælpemidler, der foretager fortolkning og forudsigelser i stedet for fagprofessionelle, som beskrives i et studie udført af Nicolae-Bogdan m.fl. (2015). Forfatterne udvikler en maskinlæringsalgoritme, der ud fra data om elevernes fravær, transporttid, karakterer med mere, med over 90 procents sandsynlighed kan udpege elever, der er i fare for at droppe ud af gymnasiet i de nærmeste 3 måneder.

FAGPROFESSIONELLES KOMPETENCER TIL AT FORTOLKE DATA

I dette afsnit fokuserer vi på forskningsspørgsmålet omhandlende, *hvilke kompetencer fagprofessionelle har i forhold til fortolkning af data, hvilke kompetencer der kræves, og hvordan det udvikles.*

Fortolkning af data er et bredt begreb. I besvarelsen af dette spørgsmål vil vi her tage udgangspunkt i, hvordan lærere og pædagoger kan *tolke* de vurderinger og datainformationer, som indsamles om eleverne. I forlængelse heraf vil vi ligeledes sætte fokus på, hvordan lærere kan *handle* på baggrund af data ved at målrette og tilpasse undervisningsakti-

viteter med henblik på at fremme elevers læring og trivsel. I afsnittet vil vi trække på forskningsoversigter, rapporter fra en både dansk og international kontekst af nyere dato samt tre forskellige effektstudier.

En gennemgående pointe inden for forskningen omkring progressionsmåling og datainformeret skoleudvikling generelt omhandler netop spørgsmålet omkring fagprofessionelles kompetencer, og udviklingen af kompetencer, til at *tolke* data (Hamilton m.fl., 2009; Rankin, 2013). Flere studier påpeger i den forbindelse på, hvordan mange fagprofessionelle har problemer med at aflæse og forstå data, hvilket blandt andet kan føre til forkerte vurderinger af elevers præstation, mindre brug af data og en generel modstand ved at bruge data (Bowen m.fl., 2011; Means m.fl., 2010, 2011; Rankin, 2013). I visse studier understreges det dog, at det ikke er nok at fokusere på visualisering og fortolkning af data. En anden gren af studier fokuserer her i højere grad på fagprofessionelles muligheder for at *handle* på baggrund af data som et afgørende forhold, der må understøttes og udvikles (Connor m.fl., 2013; Slavin m.fl., 2013). Som det ligeledes blev belyst under afsnittet om it-understøttelse, ligger der altså for fagprofessionelle både en udfordring i at lære at tolke korrekt på data samt viden om, hvordan disse resultater kan implementeres og omsættes til ændringer i undervisningspraksis. Mange studier påpeger, at begge komponenter må understøttes.

I få studier diskuteres dog også mere overordnet forholdet mellem output og udfaldsmål i den eksisterende forskning inden for skolers brug af progressionsmåling og resultatbaseret styring (Møller, Iversen & Andersen, 2016; Slavin m.fl., 2013). Et amerikansk studie påpeger blandt andet, hvordan mange deskriptive studier fremhæver nødvendigheden af at implementere databrug og understøtte fagprofessionelle i at tolke og træffe beslutninger om undervisning på baggrund af data (Slavin m.fl., 2013). Mens der i disse studier ofte ikke etableres en klar forbindelse mellem effektiv brug af data (output) og forbedring af elevers trivsel og læring (udfald). Slavin m.fl. pointerer endvidere, at effekterne af at implementere benchmark-målinger, understøttet af professionel udvikling til lærere i forhold til tolkning og anvendelse af data, indtil videre har været meget moderate. Noget tyder altså på, at målinger og korrekt brug af data ikke i sig selv har en stor effekt på elevers læring, men at fokus også må rettes mod de indsatser, der implementeres på baggrund af data.

Afsnittet her vil derfor være delt i to og fremhæve studier, der belyser henholdsvis fagprofessionelles fortolkning og anvendelse af data i en undervisningskontekst.

HVILKE KOMPETENCER KRÆVES FOR AT KUNNE FORTOLKE DATA, OG HVORDAN KAN DE UDVIKLES?

I et tidligere refereret dansk studie blev der sat fokus på lærere og skolelederes brug, holdninger og viden om de nationale tests, som blev indført

på de danske folkeskoler tilbage i 2010 (Bundsgaard & Puck, 2016). En af rapportens konklusioner var her, at både skoleledere og læreres viden om, hvordan resultaterne af de nationale tests skal fortolkes, er begrænset. Konklusionen var baseret på respondenternes svar til 6 forskellige udsagn omkring, hvordan resultaterne af testen skal fortolkes. Her svarede lærerne i gennemsnit rigtigt på 2 ud af de 6 spørgsmål, og skolelederne på gennemsnitligt 2½ ud af de 6 spørgsmål (ibid).

Som indledningsvist nævnt er fagprofessionelles manglende kompetencer til at tolke og forstå data, et fund, der ligeledes går igen i mange internationale studier. Det påpeges endvidere, at lærere i de fleste lande modtager meget sparsom eller slet ingen undervisning om dataforståelse under deres uddannelses- og oplæringsforløb (OECD, 2013; Rankin, 2013). Mere viden om, hvilke kompetencer der kræves for at tolke data, og hvordan disse kompetencer kan udvikles, er derfor yderst relevant.

En rapport fra det amerikanske undervisningsministerium fremhæver i denne sammenhæng 5 forskellige kompetenceområder, som lærere med fordel må mestre for at kunne benytte elevdata til at tilrettelægge og forbedre undervisning (Means m.fl., 2011). Anbefalingerne her er udformet i samarbejde med statistiske eksperter og er baseret på et eksplorativt studie blandt 52 individuelle lærere og 70 mindre grupper af lærere, suppleret med ekstern data. Det anbefales på denne baggrund, at lærere er i stand til at:

- Finde de relevante data i det tilgængelige it-system (lokalisering af data)
- Forstå, hvad de givne data viser (aflæsning af data)
- Finde ud af, hvad de givne data betyder (tolkning af data)
- Udvalge en undervisningstilgang eller indsats, der retter sig mod den situation, som er identificeret via data (beslutningstagen angående undervisning)
- Fremsætte relevante spørgsmål relateret til undervisningen, som kan blive besvaret ved hjælp af de data, som er tilgængelige i it-systemet (stille spørgsmål).

Forskningen omhandlende, hvordan sådanne kompetencer kan udvikles blandt fagprofessionelle, har gennemgående fokus på, hvordan man bedst muligt kan implementere og understøtte den daglige brug af data. Her er der særligt *tre fokusområder*, som går igen i mange studier. Den ene pointe er den, at fagprofessionelle tilbydes professionel kompetenceudvikling i databrug. Derudover anbefales det, at der på skolerne udbydes løbende assistance til fagprofessionelle i deres tolkning og anvendelse af data i form af lokalt forankrede datateams eller dataeksperter. Den sidste anbefaling fokuserer på, at der afsættes struktureret tid til, at lærere i samarbejde kan analysere og tolke data (fx Hamilton m.fl., 2013; Means

m.fl., 2010; Ministry of Education in Ontario, 2011; Rankin, 2013; Slavin m.fl., 2013).

Hamilton m.fl. (2009) giver i forlængelse heraf flere anbefalinger til, hvordan fagprofessionelles tolkning og brug af elevdata kan understøttes og udvikles. Det anbefales her blandt andet, at skoleledere bør tilbyde datafacilitatorer, som løbende mødes med skolens lærerteams for at diskutere data og træne lærere i at tolke data. Hamilton m.fl. påpeger, at disse facilitatorer, afhængig af skoler og distrikters budget, kan være fx fuldtidsundervisere, som tilbyder sparring til alle skolens ansatte, administrativt personale, der er tilknyttet og understøtter flere skoler, eller en ansat, som støtter medlemmerne i sit individuelle team eller afdeling. Det anbefales her, at sådanne facilitatorer ikke skal bære det fulde ansvar for dataanalyse og fortolkning, men i stedet træne fagprofessionelle til selvstændig brug af data. Dette gælder både i forhold til tolkning af data samt anvendelse af data til at forbedre undervisning og elevers præstation.

Endvidere anbefales det i kortlægningen, at skoleledere afsætter og prioriterer struktureret tid til, at lærere, som underviser på samme klassetrin eller i samme team, kan tolke elevers data i fællesskab. Afhængig af skolens individuelle behov bør der afsættes tid flere gange ugentligt til dette. Ifølge Hamilton m.fl. kan en sådan procedure understøtte, at lærere kan dele undervisningserfaringer med hinanden og dermed skabe en *best practice* baseret på kollegaers erfaringer. Samtidig kan det at tolke data i fællesskab være med til at skabe en fælles referenceramme blandt lærere for, hvordan der tolkes på data fx i forhold til, hvordan elever vurderes, og hvilke forventninger der stilles til dem (Hamilton m.fl., 2009).

Færre studier har dog fokus på, hvordan fagprofessionelles tolkning af data kan understøttes i selve it-systemet eller afrapporteringen af resultater. Som refereret i afsnittet omhandlende it-understøttelse er Rankin (2013) et konkret eksempel på, hvordan tilføjelsen af hjælpemærktøjer ved præsentation af data kan være med til at forbedre fagprofessionelles tolkning heraf. I studiet fandt man, at den samlede korrekthed af undervisernes dataanalyse i gennemsnit lå på 11 pct. før indsats og i gennemsnit steg op til 48 pct. efter indsats (Rankin, 2013).

HANDLING PÅ BAGGRUND AF DATA

Selvom forskningsspørgsmålet i denne kortlægning er centreret omkring fagprofessionelles kompetencer til at fortolke data, peger flere studier dog på, at dette fokus i sig selv ikke er tilstrækkeligt. For at brugen af målinger og tests skal have en virkning på elevers læring, er fagprofessionelles respons på disse informationer afgørende. Det er derfor også nødvendigt at understøtte lærere i forhold til, hvordan de kan handle på baggrund af det givne elevdata. Dette indebærer blandt andet en større viden om, hvordan man på baggrund af data kan justere vejledning og under-

visning af elever eller implementere nye indsatser. Flere studier peger her på, at mange fagprofessionelle også på dette område efterspørger mere støtte og større viden (Connor m.fl., 2013; Means m.fl., 2010; Rankin 2013; Slavin m.fl., 2013). En af hovedkonklusionerne i Means m.fl., (2010) var ligeledes, at det område, hvor distrikterne havde mest brug for hjælp, omhandlede modeller for, hvordan elevdata kan kobles til undervisningspraksis.

Anbefalingerne fra blandt andet What Works Clearinghouse angående struktureret tid til, at lærere i fællesskab kan tolke data, kan ligeledes understøtte anvendelsen af data. Samarbejdet kan her skabe mulighed for, at lærere kan bruge hinandens erfaringer fx til at afprøve nye metoder ved undervisning i komplekse emner (Hamilton m.fl., 2009).

Et mere konkret eksempel på, hvordan man kan understøtte fagprofessionelles anvendelse og handling på baggrund af data, er her det amerikanske effektstudie udført af Carol M. Connor m.fl., som blev beskrevet under afsnittet omhandlende it-understøttelse (Connor m.fl., 2013). Formålet med studiet var at undersøge effekten af et algoritmebaseret it-program (A2i), der på baggrund af den enkelte elevs præstationer tilbyder konkrete individuelle anbefalinger til underviser, i forhold til hvilken mængde og type af undervisning, eleven fremadrettet bør modtage.

Et andet studie med fokus på, hvordan handling på baggrund af data med fordel kan understøttes, er et amerikansk 4-årigt effektstudie udført af Slavin m.fl. (2013). Indsatsen bestod af en reformmodel implementeret på distrikt niveau. Første del bestod her af professionel assistance til distrikts- og skoleledere til at implementere kvartalvise benchmark-målinger samt støtte af læringskonsulenter til at tolke og identificere faglige problemer på baggrund heraf. Anden del bestod i professionel support til at udvælge og implementere et undervisningsprogram med god evidens for effekt, rettet mod det givne problemområde. Studiet var et randomiseret kontrolleret forsøg blandt 608 skoler fordelt over 59 distrikter i 7 forskellige stater i USA.

I løbet af studiets første 2 år, hvor benchmark-målinger blev implementeret samt støtte til tolkning af data, viste studiet ikke nogen signifikant effekt på elevers færdigheder. I løbet af studiets fjerde år viste resultaterne dog en signifikant positiv effekt på elevers færdigheder i særligt læsning. En eksplorativ analyse viste her, at effekten på elevers læsefærdigheder var større blandt de skoler, som havde implementeret evidensbaserede læseprogrammer. Resultaterne tyder her på, at det at understøtte fagprofessionelles fortolkning af data ikke er tilstrækkeligt, men også må følges op af forandringer i undervisning og læring.

Omgivende faktorer, som ligger uden for selve indsatsen, kan dog også have haft en indflydelse på studiets fund. Som i flere andre effektstudier på området kan forhold såsom reformvillighed, motivation for databrug, skoleressourcer og forældrestøtte også bære en del af for-

klaringen (Møller, Iversen & Andersen, 2016; Slavin m.fl., 2013). Derudover var andelen af skoler, som valgte at implementere et evidensbaseret undervisningsprogram, langt mindre end forventet. Omkring en tredjedel af skolerne valgte at implementere et læseprogram, mens kun 8 pct. valgte at implementere et matematisk undervisningsprogram. Dette kan til dels skyldes, at sådanne evidensbaserede programmer er omkostningsfulde og kan kræve store forandringer i praksis (Slavin m.fl., 2013).

En forskningsoversigt omhandlende skolers brug af evidensbaseret praksisser understøtter disse pointer og påpeger ligeledes flere udfordringer ved brugen af evidensbaserede undervisningsprogrammer (Bowen m.fl., 2011). Dette gælder blandt andet store omkostninger ved opstart, krævende træning af fagprofessionelle, stor afhængighed af lærerne i forhold til implementering samt omfattende programmer, der kræver intensive reformer, og som ikke nødvendigvis er tilpasset den enkelte skoles behov. I studiet af Slavin m.fl. problematiseres det, at der i store dele af USA implementeres undervisningsprogrammer, som ofte har meget lidt eller slet ikke nogen evidens for effekt, mens brug af evidensbaserede programmer kan have en større effekt på elevers færdigheder. Studierne her belyser dog altså også flere udfordringer ved implementering af sådanne programmer på lokalt niveau.

KENDETEGN VED SKOLER OG KOMMUNER, SOM BENYTTER PROGRESSIONSMÅLING

I dette afsnit vil vi sætte fokus på skoler og kommuners institutionelle kapacitet i forhold til implementeringen af databrug og arbejdet med elevers progression. Mere specifikt er afsnittet bygget op omkring besvarelsen af forskningsspørgsmålet omhandlende, *hvad der kendetegner skoler og kommuner, hvor progressionsmåling er blevet en naturlig del af det faglige arbejde på skolen.*

De studier som, på baggrund af screening og kodning, faldt inden for dette forskningsspørgsmål, er typisk teoretiske studier med afsæt i organisations- og ledelsesteori eller casestudier og deskriptive studier, som fokuserer på eksempler fra skoler, som gør brug af data og generelle anbefalinger til skoler og distrikter. Besvarelsen af dette spørgsmål er dermed hovedsageligt baseret på forskningsoversigter, deskriptive studier og casestudier.

En pointe, der understreges på tværs af studier, er den, at udviklingen af skolers brug af data og arbejdet med progressionsmåling mere specifikt sker i et komplekst samspil af interaktioner mellem både skoleledere og lærere, kommunale ledere/distriktsledere og skoleledere samt politiske beslutninger foretaget på nationalt niveau (fx Anderson, Leithwood & Strauss, 2010; Louis m.fl., 2010; Ministry of Education in Ontario, 2011). Endvidere, at der ikke er én bestemt måde at arbejde med

progressionsmåling, og der således er store variationer i praksis blandt skoler og distrikter (Datnow, Park & Wohlstetter, 2007; Hamilton m.fl., 2009). På tværs af studier er der dog også flere gennemgående praksisser og anbefalinger, som vi vil tage op i dette afsnit, og derudover give eksempler på studier af skoler, som gør brug af data i udviklingen af skole og elevers læring.

I forskningskortlægningen fra Hamilton m.fl. (2009) fremhæver man overordnet fem organisatoriske forhold, som anses som afgørende for en effektiv brug af data på både klasse-, skole- og kommuneniveau. Flere af anbefalingerne er blevet taget op enkeltvist i tidligere afsnit, hvor det har været relevant, men de fem forhold udgør tilsammen forfatterens vigtigste anbefalinger til effektiv brug af data. Anbefalinger er, (1) at data inddrages som del af en løbende cyklus af forbedring af vejledning og undervisning, (2) at elever lærer at undersøge deres egen data og sætte læringsmål for sig selv, (3) at der etableres en klar vision på skolen omkring databrug, (4) at der tilbydes støtte til fagprofessionelle, som er med til at underbygge skabelsen af en fælles datakultur på skolen og, (5) at der oprettes et fælles datasystem inden for et givent skoledistrikt (Hamilton m.fl., 2009).

Et omfattende 6-årigt amerikansk studie udført af Louis m.fl. (2010) sætter blandt andet fokus på skolefagligt lederskab på skole-, distrikts- og statsniveau, herunder hvordan distrikter påvirker skolelederes brug af data, og hvordan elevers læring er forbundet til skolers brug af data. Studiet bestod af spørgeskemadata fra i alt 8.391 lærere og 471 skoleledere; interviewdata fra 581 lærere og ledere, 304 ansatte på distriktsniveau og 124 ansatte på statsniveau samt observationsdata fra 312 klasseværelser. Derudover blev der indsamlet data om elevers præstation i matematik og læsefærdigheder i grundskole og på mellemtrin.

Studiet fandt blandt andet, at skoleledere spiller en central rolle i at etablere formålet med og forventninger omkring skolers brug af data, samt allokeringen af planlagt tid til at arbejde med data, fx via lærersamarbejde, professionel kompetenceudvikling og adgang til ekspertviden. Dette underbyggedes af det fund, at lærere typisk ikke arbejdede med data på egen hånd på de skoler, hvor ledere ikke gjorde databrug til en prioritet (Louis m.fl., 2010). Derudover fandt man desuden, at skoler med et højere præstationsniveau generelt var kendetegnede ved en større mulighed for indflydelse fra lærerteams, forældre og elever. Forældres involvering i elevers læring samt læreres oplevelse af delt lederskab mellem skoleledere og lærere havde her en korrelation med elevers faglige præstation. I forlængelse heraf, at skoleledere var mest effektive, når de oplevede at samarbejde mod klare, fælles mål på tværs af distriktsansatte, andre skoleledere og skolens lærere.

Forventninger og *accountability* mål var ligeledes to aspekter, som fyldte meget i studiets data. Resultaterne viste blandt andet, at distriktsle-

dere i distrikter, hvor niveauet for elevers læring var højt, var mere tilbøjelige til at fremhæve mål og initiativer, der rækker ud over de fælles nationale forventninger til elevers præstation. Derudover var det kendetegnende for skoler, som klarer sig godt, at lærere og skoleledere inddrogede flere forskellige typer af mål for elevers succes (ibid).

I Means, Padilla og Gallagher (2010) blev der ligeledes sat fokus på skoler og distrikters kapacitet til og brug af data. Studiet belyser også, hvilke barrierer der omvendt kan være i forhold til at skabe en øget brug af data. Blandt fagprofessionelle nævntes her udfordringer i forhold til manglende tid til dataanalyse, problemer med at anvende it-systemer, oplevelsen af, at de data, der er tilgængelige, ikke er relevante samt oplevelsen af, at distrikters politik omkring opfyldelse af pensumkrav forhindrer fagprofessionelle i at sætte tid af til at justere undervisningen til elevernes behov.

En gennemgående pointe på tværs af studier er altså blandt andet det, at skoleledere har en afgørende indflydelse på den enkelte skoles brug af data. Flere studier viser her, at data oftere bliver anvendt, hvor der fra lederes side er en motivation for databrug, og i forlængelse heraf en klar forventning om og mål for fagprofessionelles brug af data samt fokus på delt lederskab mellem ledelse og lærere. En anden vigtig pointe fra studier af skolers arbejde med progressionsmåling er spørgsmålet om at opbygge fagprofessionelles kapacitet til databrug. Som belyst i de foregående afsnit er en central anbefaling her, at der både på skole- og kommunalt niveau afsættes struktureret tid til, at lærere og pædagoger kan analysere data i fællesskab og planlægge undervisningsindsatser på baggrund heraf. Endvidere, at fagprofessionelle tilbydes professionel kompetenceudvikling i deres brug af data og har mulighed for løbende assistance via en datafacilitator eller et datateam.

DISKUSSION OG VIDERE PERSPEKTIVER

I denne kortlægning har vi haft fokus på at finde studier inden for tre hovedtemaer: Forståelser og definitioner af læring og progression; Måling af elevens læring og progression og Det fagprofessionelle arbejde med elevens progression. Vi har her besvaret de enkelte forskningsspørgsmål på baggrund af en syntese af resultaterne fra de inkluderede studier.

Overordnet er det værd at gentage, hvad vi *ikke* har fundet. Der er generelt få effektstudier inden for mange områder, hvorfor det også er begrænset, hvilken type af svar, vi kan give. Vi kan således beskrive erfaringer, men ofte ikke sige, hvad der virker bedst, eller hvad der har størst effekt. I litteraturen er der her generelt et behov for mere viden omkring sammenhængen mellem effektiv databrug og elevudfald.

Størstedelen af de studier, som er inkluderet i kortlægningen, stammer fra USA. I forhold til mulighederne og udfordringerne ved at overføre disse fund til en dansk kontekst er der generelt få emner og erfaringer, som er helt fremmede for en dansk kontekst, og man bør derfor også kunne lade sig inspirere af disse erfaringer i danske skoler. Derimod bør vi være forsigtige med direkte at overføre effektresultater og effektstørrelser, når det gælder fx hyppige målinger og modeller for data-drevet udvikling, da konteksten ofte spiller en rolle for effektstørrelser. Når det gælder effektstudier, kan man derfor ikke være sikker på at opnå samme resultater i en dansk kontekst. Vi diskuterer, hvor det er relevant, løbende muligheder og udfordringer ved at overføre fund til en dansk kontekst samt forholder os til, hvorvidt der er typer af læring, som det kan være vanskeligt at fremskaffe data om.

FORSTÅELSER OG DEFINITIONER AF LÆRING OG PROGRESSION

Overordnet er det vigtigt at have fokus på, hvad det er for et miljø, som skoler arbejder i, for eksempel i forhold til herskende tendenser omkring måling af læring og progression. På den ene side kan man her se en tendens mod et bredere fokus end kun elevers fag-faglige kompetencer. Det øgede fokus på socio-emotionelle kompetencer og det faktum, at man i mange nationale pensum nu har inkluderet den type af færdigheder og kompetencer, som samles under begrebet det 21. århundredes kompetencer, er evidens for denne strømning. En anden strømning på skoleområdet er et øget fokus på data-drevet udvikling og herunder et større fokus på mere formative målinger. På den anden side er der dog også en tendens til, at der laves flere målinger, der er – eller opfattes – som kun summative og i forlængelse heraf et større fokus på *accountability* for skoler. En evaluering fra Rambøll om brugen af de danske nationale tests fandt i den forbindelse, at testens resultater oftest bruges summativt, selvom de også var tiltænkt til formativ brug.

Hvorvidt man bør foretage formative eller summative målinger afhænger dog i høj grad også af målingens formål, idet der er brug for forskellige typer af data på forskellige organisatoriske niveauer. Lærere og pædagoger har således brug for meget detaljerede data om elevernes læring og progression, mens en skoleledelse ofte vil have brug for mere overordnede data om elevernes samlede resultater. Formative og summative målinger tjener altså hvert sit formål og er på forskellig vis nødvendige for udviklingen af den datainformerende skole og arbejdet med elevers læring og progression.

Der ligger dog her en mulig udfordring for kommuner og skoler, idet der på ene side er kommet et øget politisk fokus på skolers testscore og rankings, hvormed skolerne har større fokus på *accountability*. På den anden side må skolerne samtidig balancere implementeringen af formative målinger. Der er her en risiko for, at formativ brug af tests til intern læring kan have fået mindre fokus på bekostning af skoler og kommuners fokus på *accountability* udadtil. Hvorvidt målinger betragtes som formative eller summative, afhænger dog delvist af øjet, der betragter. Denne pointe er interessant fra flere perspektiver. Ikke mindst peger det på, at det er nødvendigt at gennemtænke, hvilke forskellige incitamenter der kan være til stede under samme type af måling – afhængigt af, på hvilket niveau målingen bruges.

I forhold til formålet med dette projekt er det værd at overveje, om det her kan skabe problemer at have visualiseringer af data, som retter sig mod lærere og pædagoger, men samtidig også skal bruges af forældre og skoleledelse. Hvis lærerne ved, at det bruges af disse interessenter, så skabes der måske også incitamenter for lærerne om at stå 'godt' ud fra de dataresultater, der bliver visualiseret. Dermed kan der

kan der være en risiko for, at undervisningen også påvirkes i en retning mod 'teaching-to-the-test' lignende adfærd, i stedet for et redskab, som skal være en hjælp til lærerne. Det vil sige, at der kan være en fare for, at lærerne kommer til at opfatte noget, som har til formål at være en hjælp til tilpasning og tilrettelæggelse af undervisning – et formativt formål – som et summativt evalueringsværktøj i stedet.

MÅLING AF ELEVENES LÆRING OG PROGRESSION

Hvad angår spørgsmålet omkring hyppigheden af målinger, er der dog foretaget en del effektstudier. Samlet set virker der til at være en sammenhæng mellem hyppigere formative målinger og bedre resultater, oftere end det omvendte. Herudover peger flere studier på, at det at lave tests i sig selv kan være en effektiv måde at lære på. Det er dog af flere grunde svært at give et mere præcist svar på spørgsmålet om, hvor hyppigt målinger skal foretages, da dette fx også afhænger af, hvilken type af test, man foretager og med hvilket formål. Det er desuden svært at svare på, hvordan "mere er bedre"-resultater, som først og fremmest stammer fra amerikanske studier, kan generaliseres til en dansk kontekst, idet vi ikke har fundet nogen danske studier, der reelt afprøver effekterne af at måle hyppigere, end man plejer at gøre i danske skoler.

I studierne af tværgående kompetencer fandt vi mange forskellige instrumenter til måling af denne type af kompetencer. Hvad angår det 21. århundredes kompetencer har vi dog ikke fundet noget eksempel, hvor man har testet, om måleinstrumentet har en selvstændig og signifikant sammenhæng med skole- og arbejdsmarkedsudfald. Derudover er der ikke et klart budskab i litteraturen i forhold til, hvilke af de 21. århundredes kompetencer, som bør måles. Derimod er der bedre afprøvede værktøjer tilgængelige til målinger af social og emotionel adfærd. Blandt denne type af instrumenter opstår der dog i stedet et andet problem, hvilket også gør sig gældende for trivselsmålinger: Eftersom der kan være tale om følsomme data og anonyme besvarelser, er de ikke altid velegnede at bruge til at følge den enkelte elevs progression.

Vi fandt, på baggrund af kortlægningen, generelt ikke nogle typer af data, som der helt mangler viden om. Men i forhold til at udvikle koncepter for en mere systematisk og kvalificeret brug af data og bedre muligheder for visninger af data er det værd at bemærke, at det ikke er alle typer af data, som er tilgængelige på individniveau. I relation til tværgående kompetencer er der ligeledes en tendens til, at det stadig først og fremmest er data om fag-faglige kompetencer, som læsning og matematik, der måles, mens der i mindre grad indsamles data om mere tværgående kompetencer og fag. Der er desuden også data, som ikke måles via tests eller mere formelle målinger, men stammer fra lærernes egne vurderinger og observationer, og her mangler der viden om, hvordan man bedst systematiserer og digitaliserer denne type af data.

I forhold til at foretage målinger og evalueringer af elevers progression hersker der ofte en antagelse om, at elever gør faglige fremskridt i samme hastighed. Det resulterer i, at progressionsmål bliver standardiserede, så de passer til den gennemsnitlige elev, frem for individuelle progressionsmål for hver enkelt elev. Elevers progression og forventningen til, hvordan de bør gøre fremskridt, bliver derfor ofte evalueret i forhold til *idealet* frem for elevens reelle mulighed for at gøre fremskridt. Man kan ikke forvente, at fagligt svage og fagligt stærke elever gør fremskridt i samme tempo. Det kan derfor være en god ide at arbejde med mere differentierede progressionsmål for elever, baseret fx på statistiske forudsigelser om progression ud fra udgangspunkt og elevtype. En ulempe ved denne type af progressionsmål er dog, at de kræver mere avancerede metoder og omfattende data og dermed er omkostningsfulde, hvorfor det også er værd at overveje, hvilken type af progressionsmål der er mest hensigtsmæssigt – afhængigt af deres formål.

I spørgsmålet omkring visualisering af data er der generelt i litteraturen en mangel på guidelines til, hvordan data bør visualiseres i forhold til både lærere, skoleledere og andre beslutningstagere på skoleområdet. Mange af de studier, vi har fundet, stammer desuden fra USA, hvor man har en tradition for at måle og præsentere data på en delvis anden måde, end vi har i Danmark. På baggrund af de inkluderede studier kan vi dog overordnet konkludere, at det – ved brug af grafer og andre visualiseringsværktøjer – er vigtigt, at fremstillingen holdes simpel og enkel; for mange informationer hindrer formidlingen og forvirrer modtageren.

DET FAGPROFESSIONELLE ARBEJDE MED ELEVENS PROGRESSION

Hvad angår skolers overordnede arbejde med elevdata fremhæves skabelsen af en datakultur flere steder i litteraturen som et vigtigt element i skolers implementering af databrug. For at skabe en fælles kultur omkring det at bruge data til udvikling af undervisning anbefales det blandt andet, at kommuner og skoleledelse udtrykker en forventning om databrug og formulerer klare visioner og mål for, hvordan data skal bruges, og hvilke læringsmål, man ønsker at opnå. I den forbindelse har man i en dansk kontekst opsat fælles nationale læringsmål, de Forenklede Fælles Mål, som indeholder klare mål for, hvilke kompetencer eleverne skal opnå på de enkelte klassetrin i de forskellige fag.

I litteraturen anbefales det endvidere, at man i høj grad medinddrager lærere og pædagoger i udviklingen og udvælgelsen af data og it-systemer. Dette er med til at skabe ejerskab og bedre implementering og sikrer samtidig, at de data, som er tilgængelige i systemet, også opleves som relevante for lærere og pædagoger i deres arbejde med elevers læring og progression.

Flere studier fremhæver desuden nødvendigheden af, at der afsættes tid og ressourcer på skolerne for at kunne implementere brugen af data til udvikling af undervisning. Det indebærer blandt andet professionel kompetenceudvikling til fagprofessionelle, løbende it-støtte samt struktureret tid til, at lærere og pædagoger kan analysere data i fællesskab. I en dansk kontekst indeholder folkeskolereformens krav om fuld tilstedeværelse umiddelbart et stort potentiale for, at fagprofessionelle nemmere kan finde tid til samarbejde i og på tværs af faggrupper, idet alle er til stede på samme tid. Et nyligt studie fra KORA, med fokus på folkeskolen efter reformen, belyser dog en mulig ledelsesmæssig udfordring i forhold til at skabe denne strukturerede tid til kompetenceudvikling og tværfagligt samarbejde mellem lærere og pædagoger. Selvom behovet for, at fagprofessionelle investerer mere tid i starten, kan være en udfordring nu og her, indeholder den datainformerende skole dog også muligheder på sigt. Det skyldes, at en bedre adgang, anvendelse og visualisering af eksisterende data, og øget brug af digitale læremidler mere generelt, har potentiale til at forenkle fagprofessionelles arbejde. Det gælder fx i forhold til funktioner, som gør det muligt at skabe en differentieret undervisning, tilbyde samlede digitale undervisningsforløb og skabe nem adgang til at give feedback og rette elevers opgaver digitalt (Rambøll, 2014b).

Ser man på, hvordan it-systemer bedst kan understøtte arbejdet med data, fremhæves det i den forbindelse, at det i høj grad er nødvendigt, at it-systemet er i stand til at koble mange typer af elevdata samt skabe mulighed for at forbinde undervisningsaktiviteter med elevpræstationer. Der findes dog ikke mange eksempler på visualiseringsværktøjer på dette område.

I relation til arbejdet med progressionsmålinger og data påpeges det i litteraturen generelt, at der er et behov for i højere grad at understøtte fagprofessionelle. Dette gælder i fortolkning af data, men særligt også i forhold til, hvordan man kan *handle* på baggrund af data og anvende elevdata til at tilpasse og justere undervisning. Et budskab fra de få effektstudier af data-drevet udvikling, vi har fundet, er her, at kendskab til elevers læring og progression, eller visualisering af data, ikke nødvendigvis er nok til at løfte eleverne. Ofte har lærere og skoler allerede en fornemmelse af, hvilke elever der har problemer, men har ikke ressourcer eller ved ikke, hvad der er mest effektivt at gøre for at støtte eleverne til at klare sig bedre. Der er derfor et behov for, at lærerne får mere hjælp til at udvælge, hvilke indsatser eller forandringer der skal laves for at forbedre undervisning og resultater.

I forhold til udfordringer ved den datainformerende skole nævnes målstyring og det at træffe beslutninger på baggrund af resultaterne fra enkeltstående tests flere steder i litteraturen som mulige faldgruber. Endvidere nævnes i litteraturen en risiko for, at undervisningen rettes mod testenes indhold og skaber en skævvridning i undervisningsfokus. I for-

længelse heraf fremhæves det i arbejdet med progressionsmåling, at det er nødvendigt, at lærere trænes i at opstille målbare mål for elevernes læring og progression – det kan i den forbindelse diskuteres, om vi herved tillægger bestemte typer af kompetencer for stor vægt og overser andre. Selvom der er et øget fokus på at forsøge at måle andre kompetencer end de traditionelt fag-faglige, er der få eksempler på skoler, kommuner eller lande, der systematisk måler fx det 21. århundredes kompetencer.

LITTERATUR

Referencer, som er markeret med en stjerne (*)forrest, er referencer til studier, som ikke opfylder kortlægningens inklusionskriterier, men som er blevet anvendt som baggrundsstudier. De resterende studier er inkluderet fra søgningen.

- Ainley, J., J. Fraillon, W. Schulz & E. Gebhardt (2016): "Conceptualizing and Measuring Computer and Information Literacy in Cross National Contexts". *Applied Measurement in Education*, 29(4), s. 291-309.
- Allen, K.D. & T.E. Hancock (2008): "Reading Comprehension Improvement with Individualized Cognitive Profiles and Metacognition". *Literacy Research and Instruction*, 47(2), s. 124-139.
- Almlund, M., A.L. Duckworth, J.J. Heckman & T. Kautz (2011): *Personality Psychology and Economics*. NBER Working Paper No. 16822.
- Anderson, S., K. Leithwood & T. Strauss (2010): "Leading Data Use in Schools: Organizational Conditions and Practices at the School and District Levels". *Leadership and Policy in Schools*, 9(3), s. 292-327.
- *Bangert-Drowns, R.L., J.A. Kulik & C-L.C. Kulik (1991): "Effects of Frequent Classroom Testing". *Journal of Educational Research*, 85(2), s. 89-99.
- *Bjørnholt, B., S. Boye, L.H. Flarup & K. Lemvigh (2015): *Pædagogiske medarbejders oplevelser og erfaringer i den nye folkeskole*. København: KORA – Det Nationale Institut for Kommuners og Regioners Analyse og Forskning.

- Black, P. & D. Wiliam (2009): "Developing the Theory of Formative Assessment". *Educational Assessment, Evaluation and Accountability*, 21(1), s. 5-31.
- Borghans, L., A.L. Duckworth, J.J. Heckman & B. ter Weel (2008): *The Economics and Psychology of Personality Traits*. IZA Discussion Paper No. 333. Bonn: Forschungsinstitut zur Zukunft der Arbeit.
- Bowen, N.K. & J.D. Powers (2011): "The Elementary School Success Profile Model of Assessment and Prevention: Balancing Effective Practice Standards and Feasibility". *School Social Work Journal*, 35(2).
- Bundsgaard, J. & M.R. Puck (2016): *Nationale test: Danske lærere og skolelederes brug, holdninger og viden*. Århus: DPU, Aarhus Universitet og Center for Anvendt Skoleforskning ved University College Lillebælt.
- Capizzi, A.M. & S.M. Barton-Arwood (2009): "Using a Curriculum-Based Measurement Graphic Organizer to Facilitate Collaboration in Reading". *Intervention in School and Clinic*, 45(1), s. 14-23. <https://doi.org/10.1177/1053451209338394>.
- Care, E., C. Scoular & P. Griffin (2016): "Assessment of Collaborative Problem Solving in Education Environments". *Applied Measurement in Education*, 29(4), s. 250-264.
- Chabrier, J., S. Cohodes & P. Oreopoulos (2016): "What Can We Learn from Charter School Lotteries?". *Journal of Economic Perspectives* 30(3), s. 57-84.
- *CITO (2014): *IRENESCHOOL WAGENINGEN - Analyse Cito Eindtoets schooljaar 2013-2014*. Hentet fra <http://docplayer.nl/9490140-Ireneschool-wageningen.html>.
- Connor, C.M., F. J. Morrison, B. Fishman, E. C. Crowe, S. A. Otaiba & C. Schatschneider (2013): "A Longitudinal Cluster-Randomized Controlled Study on the Accumulating Effects of Individualized Literacy Instruction on Students' Reading From First Through Third Grade". *Psychological Science*, 24(8) s. 1408–1419.
- Connor, C.M., F.J. Morrison, B. Fishman, C. Schatschneider & P. Underwood (2007): "Algorithm-Guided Individualized Reading Instruction". *Science*, 315, s. 464-465.
- Crooks, T.J. (2004): "Tensions between assessment for learning and assessment for qualifications". Paper presented at *The Third Conference of the Association of Commonwealth Examinations and Accreditation Bodies (ACEAB)*, Fiji, March 2004.
- Cunha, F., J.J. Heckman & S.M. Schennach (2010): "Estimating the Technology of Cognitive and Noncognitive Skill Formation". *Econometrica*, 78(3), s. 883-931
- Datnow, A., V. Park & P. Wohlstetter (2007): *Achieving with Data: How High-performing School Systems Use Data to Improve Instruction for Ele-*

- mentary Students*. Californien: Center on Educational Governance, Rossier School of Education & University of Southern California.
- *Deming, D. & Figlio, D. (2016). "Accountability in US Education: Applying Lessons from K-12 Experience to Higher Education". *Journal of Economic Perspectives*, 30(3), 33-56.
- *Datta Gupta, N. & M. Simonsen (2010): "Non-Cognitive Child Outcomes and Universal High Quality Child Care". *Journal of Public Economics*, 94, s. 30-43.
- Dobbie, W. & R.G. Fryer, Jr. (2013): "Getting beneath the Veil of Effective Schools: Evidence from New York City". *American Economic Journal: Applied Economics*, 5(4), s. 28-60.
- *Education Endowment Foundation: *Feedback*. Tilgængelig på: <https://educationendowmentfoundation.org.uk/resources/teaching-learning-toolkit/feedback/> . Besøgt d. 03.10.16.
- Ercikan, K. & M.E. Oliveri (2016): "In Search of Validity Evidence in Support of the Interpretation and Use of Assessment of Complex Constructs: Discussion of Research on Assessing 21st Century Skills". *Applied Measurement in Education*, 29(4), s. 310-318.
- Fryer Jr., R.G. (2014): "Injecting Charter School Best Practices into Traditional Public Schools: Evidence from Field Experiments". *Quarterly Journal of Economics*, 129(3), s. 1355-1407.
- Fuchs, L.S. & D. Fuchs (2002): *What is Scientifically-Based Research on Progress Monitoring?* Technical report, Nashville, TN: Vanderbilt University.
- Geisinger, H. (2016): "21st Century Skills: What Are They and How Do We Assess Them?". *Applied Measurement in Education*, 29(4), s. 245-249.
- Goodman, A & R. Goodman (2009): "Strengths and Difficulties Questionnaire as a Dimensional Measure of Child Mental Health". *Journal of the American Academy of Child and Adolescent Psychiatry*, 48(4), s. 400-403.
- Goodman, D. P., & Hambleton, R. K. (2004): "Student Test Score Reports and Interpretive Guides: Review of Current Practices and Suggestions for Future Research". *Applied Measurement in Education*, 172(2), s. 145-220.
- Gutman, L.M. & J. Vorhaus (2012): *The Impact of Pupil Behaviour and Well-being on Educational Outcomes*. London: Institute of Education, University of London. Childhood Wellbeing Research Centre.
- Greenstein, L.M. (2012): *Assessing 21st Century Skills: A Guide to Evaluating Mastery and Authentic Learning*. Thousand Oaks: Corwin Press.
- Harlen, W. (2007): "Criteria for Evaluating Systems of Student Assessment". *Studies in Educational Evaluation*, 33, s. 15-28.

- Harlen, W. (2014): *Assesment, Standards and Quality of Learning in Primary Education. A Report for the Cambridge Primary Review Trust*. York, UK: Cambridge Primary Review Trust, Derwent College M, University of York.
- Hamilton, L., Halverson, R., Jackson, S., Mandinach, E., Supovitz, J., & Wayman, J. (2009): *Using student achievement data to support instructional decision making*. Washington, DC: National Center for Education, Evaluation and Regional Assistance, Institute of Education Sciences & U.S. Department of Education.
- Hattie, J. & H. Timperley (2007): "The Power of Feedback". *Review of Educational Research*, 77(1), s. 81-112.
- Heck, R.H. (2006): "Assessing School Achievement Progress: Comparing Alternative Approaches". *Educational Administration Quarterly*, 42(5), 667-699. <https://doi.org/10.1177/0013161X06293718>
- Hull, J. (2007): *Measuring Student Growth: A Guide to Informed Decision Making*. Alexandria, USA: Center for Public Education.
- Humphrey, N., A. Kalambouka, M. Wigelsworth, A. Lendrum, J. Deighton & M. Wolpert (2012): "Measures of Social and Emotional Skills for Children and Young People: A Systematic Review". *Educational and Psychological Measurement*, 71(4), s. 617-637.
- Inchley, J., D. Currie, T. Young, O. Samdal, T. Torsheim, L. Augustson, F. Mathison, A. Aleman-Diaz, M. Molcho, M. Weber & V. Barnekow (2016): *Growing Up Unequal: Gender and Socioeconomic Differences in Young People's Health and Well-Being*. København: World Health Organization, Regional Office for Europe.
- *Jacob, B.A. & S.D. Levitt (2003): "Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating". *Quarterly Journal of Economics*, 118(3), 843-877.
- Jacob, B. & J. Rothstein (2016): "The Measurement of Student Ability in Modern Assessment Systems". *Journal of Economic Perspectives*, 30(3), s. 85-108.
- Karpicke, J.D. & P.J. Grimaldi (2012): "Retrieval-Based Learning: A Perspective for Enhancing Meaningful Learning". *Educational Psychology Review*, 24, s. 401-418.
- Keilow, M., A. Holm, S. Bagger, S. Henze-Pedersen (2014): *Udvikling af trivselsmålinger i folkeskolen – en pilotundersøgelse*. København: SFI – Det Nationale Forskningscenter for Velfærd, 14:24..
- Kiil, A., J. Nielsen Arendt, M. Rosholm (2016): *Measuring Child Well-Being*. KORA rapport, København: KORA – Det Nationale Institut for Kommuner og Regioners Analyse og Forskning.
- Kiil, A., N. Madsen Sjøe & M. Pontoppidan (2016): *Måling af børns trivsel i interventionsstudier*. Upubliceret manuskript.
- Konstantopoulos S., W. Li, S.R. Miller & A. van der Ploeg (2015): "Effects of Interim Assessments Across the Achievement Distribu-

- tion: Evidence From an Experiment". *Educational and Psychological Measurement*, p.0013164415606498.
- Kreiner, S. (2009): "Om udvikling og afprøvning af pædagogiske tests". I Bendixen, C. & S. Kreiner (red.) *Test i folkeskolen*. København: Hans Reitzels Forlag.
- Looney, J. (2011): "Integrating Formative and Summative Assessment: Progress toward a Seamless System?". *OECD Education Working Papers*, Nr. 58, Paris: OECD Publishing.
- Louis, K.S., K. Leithwood, K.L. Wahlstrom & S.E. Anderson (2010): *Learning from Leadership: Investigating the Links to Improved Student Learning*. Minnesota: University of Minnesota, University of Toronto & The Wallace Foundation.
- Lucas, B. (2016): "A Five-Dimensional Model of Creativity and Its Assessment in Schools". *Applied Measurement in Education*, 29(4), s. 278-290.
- McDaniel, M.A., H.L. Roediger & K.B. McDermott (2007): "Generalizing Test-Enhanced Learning from the Laboratory to the Classroom". *Psychological Bulletin & Review*, 14(2), s. 200-206.
- Means, B., C. Padilla, & L. Gallagher (2010): *Use of Education Data at the Local Level: From Accountability to Instructional Improvement*. Washington, D.C.: U.S. Department of Education, Office of Planning, Evaluation and Policy Development.
- Means, B., B. Chen, A. DeBarger & C. Padilla (2011): *Teachers' Ability to Use Data to Inform Instruction: Challenges and Supports*. Washington, D.C.: U.S. Department of Education, Office of Planning, Evaluation and Policy Development.
- *Ministeriet for børn, undervisning og ligestilling (2016a): *Undersøgelse af de nationale tests reliabilitet*. Tilgængelig på: <https://www.google.dk/webhp?sourceid=chrome-instant&ion=1&espv=2&ie=UTF-8#q=Unders%C3%B8gelse+af+de+nationale+tests+reliabilitet>. Besøgt 07.11.2016.
- Ministeriet for børn, undervisning og ligestilling (2016b): *Trivselsmåling*. Tilgængelig på: <https://www.uvm.dk/Uddannelser/Folkeskolen/Elevplaner-nationale-test-og-trivselsmaaling/Trivselsmaaling>. Besøgt 30.09.2016.
- *Ministeriet for børn, undervisning og ligestilling (2016c): *Fælles Mål*. Tilgængelig på: <https://www.uvm.dk/Uddannelser/Folkeskolen/Fag-timetale-og-overgange/Faelles-Maal>. Besøgt 17.10.2016.
- Ministry of Education, Ontario (2013): *Exploring Five Core Leadership Capacities. Using Data: Transforming Potential into Practice*. Canada: Ontario.

- Møller, M. Ø., K. Iversen & V.N. Andersen (2016): *Review af resultatbaseret styring – Resultatbaseret styring på grundskole-, beskæftigelses- og socialområdet*. København: KORA – Det Nationale Institut for Kommuner og Regioners Analyse og Forskning.
- Newton, P.E. (2007): “Clarifying the Purposes of Educational Assessment”. *Assessment in Education*, 14:2, s. 149-170.
- *Nicolae-Bogdan, Ş., R. Halland, C. Igel, & S. Alstrup (2015): “High-School Dropout Prediction Using Machine Learning: A Danish Large-scale Study”. I M. Verleysen (red.), *Proceedings/23rd European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, ESANN 2015, Bruges, Belgium, April 22-23-24, 2015*. Louvain-la-Neuve: Ciaco.
- Nusche, D., D. Laveault, J. MacBeath & P. Santiago (2012): *OECD Reviews of Evaluation and Assessment in Education: New Zealand 2011*. Paris: OECD Publishing.
<http://dx.doi.org/10.1787/9789264116917-en>.
- OECD (2013): *Synergies for Better Learning – An International Perspective on Evaluation and Assessment*. OECD Reviews of Evaluation and Assessment in Education. Paris: OECD Publishing,
<http://dx.doi.org/10.1787/9789264190658-en>.
- OECD (2016): *PISA 2015 Assessment and Analytical Framework: Science, Reading, Mathematics and Financial Literacy*. PISA, Paris: OECD Publishing. <http://dx.doi.org/10.1787/9789264255425-en>.
- Rambøll (2013): *Evaluering af de nationale test i folkeskolen*. Rapport for Undervisningsministeriet. København: Rambøll.
- Rambøll (2014a): *Forskningskortlægning af undervisningsmiljø og trivsel*. Rapport for Undervisningsministeriet. København: Rambøll.
- Rambøll (2014b): *Anvendelse af digitale læremidler. Effektmåling*. Rapport for Styregruppen for IT i Folkeskolen. København: Rambøll.
- Rankin, J.G. (2013): *Over-the-Counter Data's Impact on Educators' Data Analysis Accuracy*. Dissertation, Northcentral University, Prescott Valley, Arizona.
- Rawson, K.A. & J. Dunlosky (2012): “When Is Practice Testing Most Effective for Improving the Durability and Efficiency of Student Learning?”. *Educational Psychology Review*, 24, s. 419-435.
- *Rowland, C.A. (2014): “The Effect of Testing Versus Restudy on Retention: A Meta-Analytic Review of the Testing Effect”. *Psychological Bulletin*, 140(6), s. 1432-163.
- Sabbah, F.M. (2011): *Designing More Effective Accountability Report Cards*. University of California, Berkeley, California, USA.
- Seltzer, M., Choi, K., & Thum, Y. M. (2003): “Examining Relationships Between where Students Start and How Rapidly They Progress: Implications for Conducting Analyses that Help Illuminate the

- Distribution of Achievement Within Schools”. *Educational Evaluation and Policy Analysis*, 25(3), s. 263-286.
- Shewbridge, C., E. Jang, P. Matthews & P. Santiago (2011): *OECD Reviews of Evaluation and Assessment in Education: Denmark*. Paris: OECD Publishing, <http://www.oecd.org/edu/evaluationpolicy>.
- Skolverket (2016): *Attityder till skolan 2015*. Rapport 438.
- Slavin, R.E., A. Cheung, G. Holmes, N.A. Madden & A. Chamberlain (2013): “Effects of a Data-Driven District Reform Model on State Assessment Outcomes”. *American Educational Research Journal*, 50, s. 371-396.
- Törner, E. (2015): *Utvärdering av DigiLys. En modell för ständig utveckling av lärmiljön*. Rapport, Spira utvärdering.
- Undervisningsministeriet (2014): *Endelige anbefalinger fra ekspertgruppen om elevers trivsel i folkeskolen i relation til nationale trivselsmålinger*. Tilgængelig på: <https://www.uvm.dk/Uddannelser/Folkeskolen/Laering-og-laeringsmiljoe/Trivsel-og-undervisningsmiljoe/Ekspertgruppen-om-elevernes-trivsel>. Besøgt 30.09.2016.
- Van der Kleij, F.M., R.C.W. Feskens & T.J.H.M. Eggen (2015):”Effects of Feedback in a Computer-Based Learning Environment on Students’ Learning Outcomes: A Meta-Analysis”. *Review of Educational Research*, 85(4), s. 475-511.
- Wayman, J.C., J.B. Jimerson & V. Cho (2012): “Organizational Considerations in Establishing the Data-Informed District”. *School Effectiveness and School Improvement*, 23(2), s. 159-178.
- Wayman, J.C. (2005):”Involving Teachers in Data-Driven Decision Making: Using Computer Data Systems to Support Teacher Inquiry and Reflection”. *Journal of Education for Students Placed at Risk*, 10(3), s. 295-308.
- West, M.R. (2016): “Should Non-Cognitive Skills Be Included in School Accountability Systems? Preliminary Evidence from California’s CORE Districts”. *Evidence Speaks Reports*, Vol 1, #13, Washington: Brookings Institution.
- Zapata-Rivera, D., W. VanWinkle & R. Zwick (2012): *Applying Score Design Principles in the Design of Score Reports for CBAL Teachers*. Princeton, New Jersey: ETS.