



Københavns Universitet

Evaluering af naturfaglige kompetencer

Dolin, Jens; Nielsen, Jan Alexis; Tidemand, Sofie

Published in:

Acta Didactica Norge - nasjonalt tidsskrift for fagdidaktisk forsknings- og utviklingsarbeid

DOI:

[10.5617/adno.4702](https://doi.org/10.5617/adno.4702)

Publication date:

2017

Document Version

Også kaldet Forlagets PDF

Citation for published version (APA):

Dolin, J., Nielsen, J. A., & Tidemand, S. (2017). Evaluering af naturfaglige kompetencer. Acta Didactica Norge - nasjonalt tidsskrift for fagdidaktisk forsknings- og utviklingsarbeid, 11(3), 1-28. [2].
<https://doi.org/10.5617/adno.4702>

Jens Dolin

Københavns Universitet

Jan Alexis Nielsen

Københavns Universitet

DOI: <http://dx.doi.org/10.5617/adno.4702>

Evaluering af naturfaglige kompetencer

Sammendrag

Artiklen skitserer kort de sidste 40-50 års udvikling i naturfagene op til det nuværende fokus på undersøgelsesbaseret undervisning og udvikling af kompetencer. Den påpeger hvorledes mange traditionelle evalueringsformer, især brugt ved eksamen, ikke er i stand til at indfange de ønskede kompetencer, hvorfor de har svært ved at slå igennem i den daglige undervisning.

Med udgangspunkt i en model af sammenhænge mellem formativ og summativ brug af evaluering argumenteres for nødvendigheden af større alignment mellem naturfagenes formål, pædagogik og evalueringsformer. Artiklen giver eksempler fra en række danske og internationale forskningsprojekter på udvikling og implementering af evalueringsformer, der kan indfange de nye læringsmål og som brugt formativt kan fremme deres læring. Den viser hvilke muligheder og udfordringer de rummer for lærere, og hvorledes en meget struktureret brug af evalueringer kan risikere at elevmotivationen forskydes fra en mestringsorientering hen mod en præstationsorientering.

Afslutningsvis bliver der peget på forskellige måder til at håndtere modsætningerne mellem en formativ og en summativ brug af evalueringer. Dels gennem tiltag, der mindsker karakterpresset i skolen, og dels gennem udvikling af nye eksamensformer, der er i bedre overensstemmelse med en kompetenceorienteret undervisning.

Nøgleord: kompetencer, evaluering, naturfag, motivation, karakterer

Assessment of science competences

Abstract

The article briefly outlines the past 40-50 years of development in science education up to the current focus on inquiry-based teaching and competence

development. It points out how many traditional forms of assessment, especially used for examinations, are unable to capture the desired competences, which makes them difficult to realize in daily teaching.

Based on a model of the relationships between formative and summative use of assessment, it is argued for the need for greater alignment between the goals in science subjects, teaching and assessment. The article provides examples from a number of Danish and international research projects on the development and implementation of assessment methods that can capture the new learning objectives and which used formatively can promote their learning. It demonstrates what opportunities and challenges they have for teachers, and how a highly structured use of assessment may risk shifting student motivation from a mastering orientation towards a performance orientation.

In conclusion, different ways are identified to deal with the contradiction between a formative and a summative use of assessment. Partly through actions that reduce the level of pressure in school for high marks, and partly through the development of new forms of assessment that are in better accordance with a competence-oriented education.

Keywords: competences, assessment, science education, motivation, grading

Evaluering af naturfaglige kompetencer

Naturfagsdidaktikken i Danmark så vel som i mange andre lande (herunder i Norden) har i det sidste årti udviklet et stærkt fokus på at undervisningen bør være *undersøgelsesbaseret* og have som mål, at elever tilegner sig *kompetencer* og ikke blot tilegner sig viden igennem mødet med et pensum (Nielsen, Waadegaard, Dolin & Bruun, 2017). I denne artikel redegør vi kort for den historiske udvikling i det naturfagsdidaktiske felt bag dette kompetenceorienterede fokus, og vi argumenterer for at dette fokus skaber et behov for ændringer i den fremherskende evalueringspraksis. På baggrund af erfaringer fra tre nyligt afsluttede projekter, præsenterer vi en række muligheder og udfordringer forbundet med ændringer i tilgange til evaluering af elevers kompetencetilegnelse.

Naturfagsdidaktikken omhandler – ligesom alle fags didaktik – de grundlæggende spørgsmål om *hvorfor* unge skal have naturvidenskab, *hvad* de så skal lære om, og endelig *hvordan* læringen kan foregå (Duit 2015; Krogh 2017). De tre spørgsmål – *Hvorfor? Hvad? Hvordan?* – hænger uløseligt sammen. En bestemt begrundelse peger på et vist indhold som bedst læres på visse måder, og bestemte måder at arbejde med et fag på giver nogle bestemte indsigter og holdninger og opfylder dermed nogle bestemte læringsmål.

Før vi dykker ned i naturfagsdidaktikken, er det nødvendigt med en begrebsafklaring. Vi anvender 'naturfag' som samlebetegnelse for de

undervisningsfag, der arbejder med naturvidenskabelige problemstillinger, herunder teknologi og engineering. Det er forskelligt fra skolesystem til skolesystem og også afhængigt af uddannelsestrin i hvilket omfang naturvidenskaberne behandles i et fælles science-fag eller i adskilte disciplinfag, som fysik, biologi, kemi, geografi, ligesom teknologi og engineering i nogle lande indgår i naturvidenskabelige fag og i andre udfoldes i deres egne fag. I det danske skolesystem møder eleverne først et integreret natur-/teknologifag i 1.-6. klasse for i 7.-9. klasse at have tre adskilte fag: Fysik/kemi, biologi og geografi. Der er dog en fælles afgangsprøve for de tre fag, som fagene gennem tværfaglige forløb skal forberede eleverne på. I gymnasiet (videregående skole) er der et introducerende Naturvidenskabeligt Grundforløb som overgang til en række selvstændige naturvidenskabelige fag, mens det tekniske gymnasium desuden har selvstændige teknologiske fag.

Naturfag dækker således over mange forskellige konkrete udformninger, der dog deler en lang række fælles problemstillinger, og det er dem, vi vil fokusere på i det følgende. Vi vil første kort sammenfatte de sidste 40-50 års centrale udviklingstendenser i dette komplekse og sammenhængende felt med henblik på at (i) belyse udviklingen af det stærke fokus på undersøgelsesbaseret og kompetencer, og på at (ii) udfolde hvad vi anser for en af de mest centrale problemstillinger i naturfagsundervisningen – nemlig, spørgsmålet om hvorledes man validt kan evaluere avancerede naturfaglige kompetencer som fx modellering og innovation, med fokus på fremme af læring i stedet for blot måling af læring.

Det naturfagsdidaktiske forskningsfelt

Naturfagsdidaktik er et ungt forskningsfelt både internationalt og i Norden. For at specificere hvornår det naturfagsdidaktiske forskningsfelt begynder, læner vi os op ad Fenshams (2004) kriterier for hvornår man kan tale om et forskningsfelt: anerkendelsen af at feltet er et akademisk felt via eksistensen af felt-specifikke forskningstidsskrifter, felt-specifikke professionelle organisationer og felt-specifikke konferencer. En autoritet som van den Akker (1998) angiver 1950'erne som en start: "... when science education began to be a subject of serious policy making and when the first large-scale science curriculum development activities were initiated" (ibid. s. 423). De store, internationale tidsskrifter i Science Education blev etableret i 1960'erne og 70'erne, den europæiske forskersammenslutning, European Science Education Research Association, blev etableret i 1995, mens de Nordiske forskersymposier om undervisning i naturfag startede allerede i 1984. Den første nordiske professor i naturfagsdidaktik var Björn Andersson i 1997 og det nordiske tidsskrift for naturfagsdidaktik, NorDiNa, startede i 2005. (For en autoritativ gennemgang af feltets historie indtil starten af dette årtusind se Fensham (2004))

og van den Akker (1998)). Det er klart at der før 1960'erne, var tanker om undervisning og læring af naturfagligt indhold. For eksempel tematiserede Dewey (1910) i starten af det 20. århundrede elevens manglende interesse i naturfag og relaterede dette til mangler i undervisningen.

Langt op i 1970'erne lå fokus på indholdet og naturfagene var en bearbejdet udgave af videnskabsfagene. Lærebøgerne var skrevet af universitetsforskere og fagenes discipliner blev i timerne gennemgået slavisk med henblik på at lære det videnskabelige indhold og metoder. Som van den Akker (1998) opsummerer: "Preparation for the next academic level seemed to be the almost exclusive goal of most teachers. Science instruction appeared to be overly dependent on textbook use. Direct experience, inquiry approaches and other forms of intellectual stimulation were uncommon" (s. 425). Det store skred i retning af den nuværende elev- og omverdensorienterede tilgange til naturfagsundervisningen skete med konstruktivismens indtog op gennem 1970'erne og 80'erne i først den naturfagsdidaktiske forskning og efterfølgende i undervisningen. Rosalind Driver kan med rette siges at have startet denne bølge med sine undersøgelser af elevens opfattelse af naturvidenskab og naturvidenskabelige sammenhænge. Titler som *The Pupil as Scientist* (Driver, 1983), *Children's Ideas in Science* (Driver, 1985) og *Young people's images of science* (Driver et al., 1996) afspejler klart konstruktivismens grundideer. Hun viste at elever har et sammenhængende verdensbillede baseret på snusfornuftig tolkning af deres erfaringer. Frem for udtryk for Piagetske kognitive udviklingsstadier er der snarere tale om paradigmeskift, hvor én forståelse afløser en anden. Man går altså fra en opfattelse af læring som en generisk eller menneskeindlejret proces hen imod en opfattelse af læring som en forståelse, der udvikler sig, der konstrueres, gennem arbejde med erfaringer. Denne læringsopfattelse fik helt naturligt konsekvenser for undervisningsformen. Når elever ikke lærte via passiv tilegnelse, men via aktiv skabelse af viden, måtte undervisningen gøre dette muligt.

Den konstruktivistiske forståelse som begyndte at præge naturfagene kan ses i sammenhæng med en fremvoksende opfattelse af læring og uddannelse som deltagelse i en kultur. Dette sociokulturelle grundsyn blev klart formuleret af Jerome Bruner i hans bog *The Culture of Education* fra 1996. Bruner viser hvorledes børn – og dermed også uddannelsen af dem – udvikler sig bedst gennem at deltage i kulturens måde at handle og tænke og tale på. Sproghandlinger var en central del af kulturindførelsen: "Learning science is talking science", som Jay Lemke (1990) udtrykte det.

Den konstruktivistiske tilgang satte en bølge af undersøgelser i gang af elevens forståelse af naturfænomener og udbytte af naturfagsundervisningen – med ret deprimerende resultater! Fx undersøgte de danske fysikdidaktiske forskere Henry Nielsen og Poul V. Thomsen i begyndelsen af 1980'erne danske gymnasielevens forståelse og motivation og fandt, at eleverne i gymnasiet opfattede fysikken abstrakt og virkelighedsfjernt og at de var demotiverede og

uinteresserede i faget (Nielsen & Thomsen 1985, 1986 og 1988). Disse undersøgelser blev fulgt op af nogle der viste, at jo længere tid eleverne havde fysik i skolen, jo mere demotiverede blev de (Krogh et al. 2001). Inddragelsen af en elevvinkel åbnede op for et selvkritisk og fornyende blik på naturfagsundervisningen.

De konstruktivistiske og sociokulturelle indsigter initierede også en række forsøg med nye undervisningsformer og nyt indhold. Projektarbejdet blomstrede op gennem 1970'erne, 1980'erne og 1990'erne. Mange arbejdede med omverdensrelevante problemstillinger, ofte i tværfaglige sammenhænge, og hele naturvidenskabens processide blev vægtet højere. Eleverne skulle ikke kun lære et konkret indhold, de skulle lære at arbejde naturvidenskabeligt. Wolff-Michael Roths (1995) *Authentic School Science* kan ses som eksponent for en sådan 'undersøgelsesbaseret' tilgang til naturfagsundervisningen.

I slipstrømmen af konstruktivismen blev der stillet spørgsmål til 'Hvorfor?'. Den herskende pipe-line approach (dvs. van den Akkers "preparing for the next academic level") blev problematiseret i takt med at en stadig større andel af de unge gik i gymnasiet – og de skulle ikke alle have en naturvidenskabelig uddannelse. De rapporter, som i 1990'erne kom i forbindelse med det amerikanske *Projekt 2061* introducerede en forståelse af naturfagene som havende både en demokratisk og en personlig betydning, og den kobledes denne betydning af naturfagene med deres samfundsmæssige og økonomiske værdi. I publikationen *Science for All Americans* (AAAS, 1990) begrebssættes den såkaldte *scientific literacy*, som alle elever bør have opnået som resultat af deres samlede skoleoplevelser i naturvidenskab, matematik og teknologi på det tidspunkt, hvor de forlader high school, nemlig:

- at være fortrolig med den naturbundne verden og forstå både dens mangfoldighed og dens helhed
- at forstå nogle af de vigtige måder, hvorpå naturvidenskab, matematik og teknologi afhænger af hinanden
- at forstå naturvidenskabens nøglebegreber og principper
- at have evne for naturvidenskabelig tænkemåde
- at forstå at naturvidenskab, matematik og teknologi er menneskeskabte projekter; og have forståelse for, hvad det betyder for deres styrker og begrænsninger
og
- at kunne anvende naturvidenskabelig viden og tænkemåde til individuelle og samfundsrelaterede formål.

Projekt 2061 ændrede – sammen med en række andre projekter, som fx *Beyond 2000: Science education for the future* (Millar & Osborne, 1998) i Storbritannien – fokus fra en videnskabsorientering med vægt på konkret viden i undervisningen hen mod en mere procesorienteret, dannende, science-for-all

tilgang. I de nordiske lande blev diskussionerne af formålene med naturfagsundervisningen desuden inspireret af OECD-rapporter som *Key Competencies for a Successful Life and Well-Functioning Society* (Rychen & Salganik, 2003). Disse argumenterede for at globaliseringen og videnskabspløsonen og den hurtige samfundsudvikling med den stærkt øgede kompleksitet gjorde det vanskeligt at præcisere de fremtidige krav til borgere og arbejdskraft og det var derfor nødvendigt at lægge mere vægt på almene, sociale og personlige egenskaber – som en del af de faglige krav. Dette, sammen med et ønske om at eleverne i højere grad skulle være i stand til at anvende det, de lærte i skolen, i situationer uden for skolen, satte i det meste af Europa gang i at formulere målene for naturfagene i kompetencetermer (Waddington, Nentwig & Schanze, 2007). Essensen i faglige kompetencer er at de forsøger at indfange det almene i faget, det, der er hævet over det konkrete indhold: måder man arbejder med faget på – på tværs af de forskellige fagdiscipliner. I det danske 'faglighedsprojekt' blev matematik, dansk, sprogfagene og naturfagene således formuleret i kompetencetermer. I den opsamlende rapport *Fremtidens uddannelser. Den ny faglighed og dens forudsætninger* (Undervisningsministeriet, 2004) defineres faglige kompetencer som "[...] en vidensbaseret parathed til at handle hensigtsmæssigt i situationer som rummer en bestemt slags [...] faglige udfordringer" (s. 19). I rapporten *Fremtidens naturfaglige uddannelser Naturfag for alle - vision og oplæg til strategi* (Andersen et al., 2003) blev de naturfaglige kompetencer formuleret, og der arbejdes i Danmark nu med fire naturfaglige kompetencer:

- Undersøgelseskompetence (empirikompetence)
- Modelleringskompetence
- Perspektiveringskompetencen
- Kommunikationskompetence

Det vil sige at uanset hvilket naturfag eller hvilken disciplin, der arbejdes med, kan det konkrete arbejde ses som bidragene til at udvikle elevernes naturfaglige kompetencer.

I Europa satte rapporten *Science Education now: A Renewed Pedagogy for the Future of Europe* (European Commission, 2007) en markant dagsorden for naturfagsdidaktisk undervisning og forskning. Rapportens titel afspejler at forfattergruppen mente at den største årsag til elevernes manglende interesse for naturfag skulle findes i den måde, som der blev undervist på i naturfagene - ligesom Dewey 100 år tidligere! Som konsekvens heraf koncentrerede gruppens anbefalinger sig om en fornyelse af naturfagspædagogikken, fra en overvejende deduktiv til en udforskningsbaseret tilgang. En sådan pædagogik skulle også give øgede muligheder for samarbejde mellem skolen og uformelle miljøer, som virksomheder, forskere, museer, science centre, etc. Rapporten gav direkte anstød til en bølge af europæiske projekter inden for det 7. Rammeprogram

'Science & Society' indsats, der alle arbejdede med at indføre undersøgelsesbaseret naturfagsundervisning. Således er der listet 135 projekter (se www.cordi.europa.eu) i det 7. Rammeprogram med nøgleordene 'inquiry' og 'science education' – herunder projekter som PROFILES, SAILS, PRI-SCI-NET, IRRESISTIBLE, PRIMAS, ASSIST-ME, ESTABLISH, FIBONNACCI (alle disse akronymer indeholder minimum én gang bogstavet 'I' for 'Inquiry').

Opsummerende er det rimeligt at sige at der i naturfagene gennem de sidste godt 40 år er arbejdet på at etablere en balance mellem formålet, indholdet og pædagogikken på de forskellige uddannelsestrin – ofte drevet af de ændrede formål der opstilles for fagene, som igen skyldes konsekvenser af en uhensigtsmæssig pædagogik eller et utidssvarende indhold. Der arbejdes nu i et vist omfang undersøgelsesbaseret med for eleverne relevante problemstillinger, og der er gennem en lang række projekter (fx de ovennævnte EU-finansierede projekter) og efteruddannelsesinitiativer opbygget såvel en omfattende litteratur som en bred accept af det fornuftige i at undervise undersøgelsesbaseret og kompetenceorienteret.

Drivkræfterne i at indføre undersøgelsesbaserede arbejdsformer har været at fremme elevmotivation og at give eleverne et mere realistisk billede af hvorledes naturvidenskaben arbejder. Kompetencer blev indført som uddannelsesmål for at eleverne skulle tage noget med sig fra naturfagene, som de kunne bruge uden for skolen, så de kunne anvende deres viden i hverdags- og arbejdsliv.

Men i de sidste ca. 20 år har et fremvoksende testregime (EACEA; Eurydice, 2009) baseret på relativt enkle testspørgsmål (Bundsgaard, 2016) vanskeliggjort realiseringen af de nye mål og arbejdsformer.

Evalueringsproblematikken

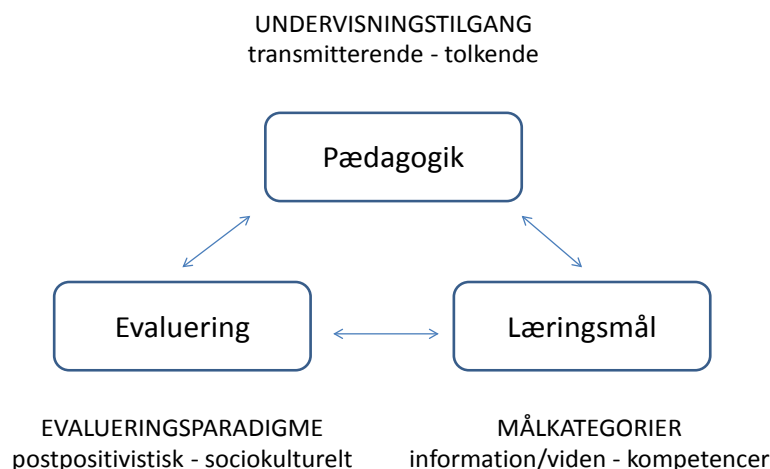
I dette afsnit vil vi argumentere for, at de eksisterende evalueringsformer må tilpasses de nye målformuleringer og arbejdsformer, for at hele det arbejde, der er lagt i nye målformuleringer og nye undervisningsformer, ikke skal være forgæves. Sat på spidsen kan man sige, at eksamen er den hale, der logrer med hunden.

Det er efterhånden bredt accepteret, at evaluering og evalueringspraksisser spiller en helt central rolle for elevers læring (for et tidssvarende overblik se Nielsen et al., 2017). Evaluering skal her forstås som indsamling, fortolkning og anvendelse af informationer i forhold til at nå til en vurdering af en elevs faglige niveau og eventuelt med henblik på at træffe beslutning om fremtidige undervisningssituationer (Harlen, 2007). Der er omfattende evidens for, at (god) formativ evaluering har en markant og positiv indflydelse på elevers læring (se fx Black & Wiliam, 1998; Black, Harrison & Lee, 2004; Hattie & Timperley, 2007). Samtidig er summativ evaluering ikke i sig selv en kilde til læring (se fx

Butler, 1987; 1988); men summativ evalueringspraksis – især de eksamensformer der afslutter fagene – sætter dagsordenen for, hvad der sker i den daglige undervisning (se fx Nordenbo et al., 2009). For eksempel har målformuleringer i naturfagene i Danmark ofte en dimension, hvor eleven skal kunne bruge faget til at forholde sig kritisk til samfundsmæssige nøgleproblemer (i den internationale naturfagsdidaktik er denne dimension ofte behandlet som *socioscientific issues*; se Sadler & Zeidler, 2003; Nielsen, 2012); men da sådan en dimension stort set ikke testes til eksamen er der tydelige tegn på at denne dimension udvandes i den daglige undervisning på trods af, at den er en eksplicit del af målformuleringen (se Tidemand & Nielsen, 2017).

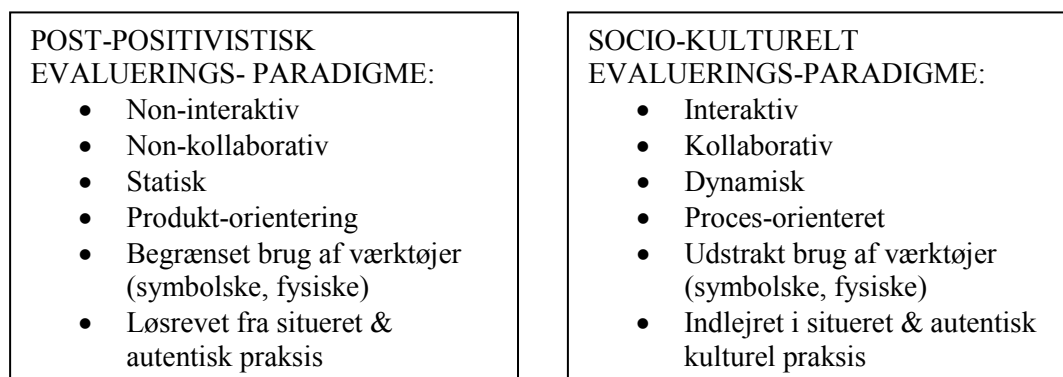
Som det er beskrevet i det foregående afsnit stilles der stadig nye og større krav til hvad eleverne skal have ud af naturfagsundervisningen. Ud over centrale naturvidenskabelige begreber skal eleverne nu også tilegne sig både naturfaglige kompetencer, såsom evne til at modellere, eksperimentere etc., og mere generiske kompetencer, såsom evne til at perspektivere, kommunikere, være innovativ osv. Men de evalueringsformer, som traditionelt har været fremherskende i det danske uddannelsessystem, kan kun dårligt indfange disse nye mål i undervisningen. Kompetencer er jo udtryk for en evne til at handle vidensbaseret i bestemte, autentiske situationer. De kan derfor kun evalueres validt ved at se elever handle i tilsvarende situationer – og at sidde alene i en gymnastiksal eller snakke 20 minutter i en eksamenssituation er ikke en rimelig afspejling af en relevant situation (for et eksempel på hvor 'få' kompetencer en skriftlig eksamen i fx faget Bioteknologi kan dække se Elmeskov, Bruun & Nielsen, 2015). Sådanne eksamenssituationer kan til en vis grad teste elevens evne til at reproducere viden og tankeprocesser. Men massiv forskning viser hvorledes evalueringsformen har en afgørende afsmitning på undervisningen (Harlen, 2012; Nordenbo et al., 2009), så et uddannelsessystem, der har evalueringsformer, der kun vurderer 'traditionelle' uddannelsesmål, vil fremme undervisning, der prioriterer læring af sådanne uddannelsesmål. Det vil sige at de eksisterende evalueringsformer i vid udstrækning blokerer for at eleverne får mulighed for at tilegne sig de nye læringsmål.

Disse sammenhænge er illustreret i evalueringstrekanten i Figur 1. Pointen er, at der skal være alignment mellem de mål, der er opstillet for undervisningen, den undervisning, som eleverne deltager i, og det evalueringsgrundlag, som man evaluerer eleverne ud fra.



Figur 1. Evalueringstrekanten. Nye mål kræver nye undervisningsformer, der gør det muligt for eleverne at tilegne sig målene, og nye evalueringsformer, der kan indfange de nye mål.

Forskellige typer faglige mål (viden, færdigheder, kompetencer) kan fremelskes gennem forskellige pædagogiske tilgange og skal ligeledes evalueres på forskellig vis. Mens mere enkle læringsmål, som fx konkret viden, kan læres via 'traditionelle' formidlende undervisningsformer og validt testes i relativt enkle tests som fx multiple choice test, kræver læring af kompetencer, der jo består af situerede handle-mønstre, en mere involverende og tolkende undervisning i et relevant miljø, og kompetencetilegnelsen kan kun validt testes i en kontekst der svarer til lærings-situationen. Der er således tale om to forskellige evaluering-paradigmer, som skematisk er opstillet i Figur 2. Et *post-positivistisk paradigme* ser læring af viden som et spørgsmål om at overføre objektificeret viden til eleven, og i dette paradigme vil standardiserede opgaver i begrænsede respons-formater (fx multiple choice, korte åbne svar) være tilstrækkeligt til at checke om den autoriserede viden er modtaget (Krogh & Dolin, 2016, s. 21f).



Figur 2. Oversigt over to evaluering-paradigmer. Inspireret af Gipps, 1999; Bell, 2007, her fra Krogh & Dolin, 2016, s. 21f.

Herover for står det *socio-kulturelle paradigme*, som ”stresses how knowledge is conditioned and constrained by the technology, information resources, representation systems, and social situations with which people interact” (Mislevy, Steinberg & Almond, 2003). Viden er her ikke kun et flytbart objekt, som den enkelte besidder i større eller mindre grad, men findes indlejret i de situationer og processer, som man indgå i sammen med andre (Krogh & Dolin, 2016, s. 21f).

Før validt at kunne vurdere kompetencer er det derfor nødvendigt at der udvikles nye evalueringsmetoder baseret på et sociokulturelt evalueringsparadigme, hvor altså eleverne kan udfolde sig i autentiske, virkelighedsnære kontekster. Det vender vi tilbage til. Men et nok så vigtigt hensyn at tage ved udvikling af evalueringsmetoder, er at de kan levere valide data til politikerne. Dels fordi politikernes ønske om at få data overskygger hensynet til læringen og dels fordi de data, som politikere overvejende bygger deres beslutninger på, ikke altid er valide i forhold til det, de skal bruges til.

Der er i de seneste fem til otte år i alle de nordiske lande fra politisk side etableret et relativt stærkt testregime. Det kan ses som et element i den New Public Management-bølge (Greve, 2002), der skyllede ind over uddannelsessystemet i løbet af 2000’erne, ligesom den gjorde over hele den offentlige sektor. Politikerne ville vide om de fik ’value for money’ og de ville have redskaber til at styre udviklingen. Dette kræver målbare data og opbygning af en evaluerings-/testkultur. PISA kan ses i dette lys ligesom beslutningen om indførelsen af nationale tests i Danmark i 2005 (Dolin, 2016). Problemet er, at indførelsen af tests og summative evalueringer ikke er neutrale ’måleredskaber’, men ændringer som har en række negative konsekvenser for såvel undervisningen som for eleverne. Nordenbo et al (2009) gennemførte et systematisk review over hvorledes indførelsen af testtyper, som indgår i nationale test i de nordiske lande, indvirker på læreres didaktiske beslutninger og elevers læringsadfærd. Hovedresultaterne kan groft opsummeres som:

Undervisningen ændres i retning af at

- curriculum indsnævres eller fordrejes, idet faglige tankegange forsimples, faktaviden og mekaniske færdigheder betones på bekostning af kreative og æstetiske perspektiver
- undervisningstid allokeres til det/de fag og fagområder, der skal testes i, på bekostning af de fag og fagområder, der ikke testes i, og
- undervisningen kan forfalde til træning til testen og udenadslæren.

Elevernes læringsadfærd ændres i retning af at

- elevernes testresultater stiger ved indførelse af test, men først efter nogle år
- når en test annonceres, kan det udløse følelsesreaktioner som nervøsitet og angst
- eleven forbereder sig ved at lære udenad og memorere sætninger
- for bedre præsterende elever stiger motivationen, mens svagere præsterende taber modet
- det testresultat, som eleven får ved testen, kan virke ind på fremtidig motivation og selvværd.

Der sker således dramatiske ændringer af såvel undervisningen som af elevernes adfærd i og oplevelse af undervisningen når evalueringers summative brug kommer i fokus. Især effekten på eleverne har været meget markant de sidste år. En undersøgelse af stress i gymnasiet i Danmark i 2016 viste at stressniveauet hos 55 procent af de elever i anden gymnasieklasse, der indgik i undersøgelsen, var så højt, at det svarer til niveauet hos de 20 procent mest stressede fra resten af befolkningen (Nielsen & Lagermann, 2017):

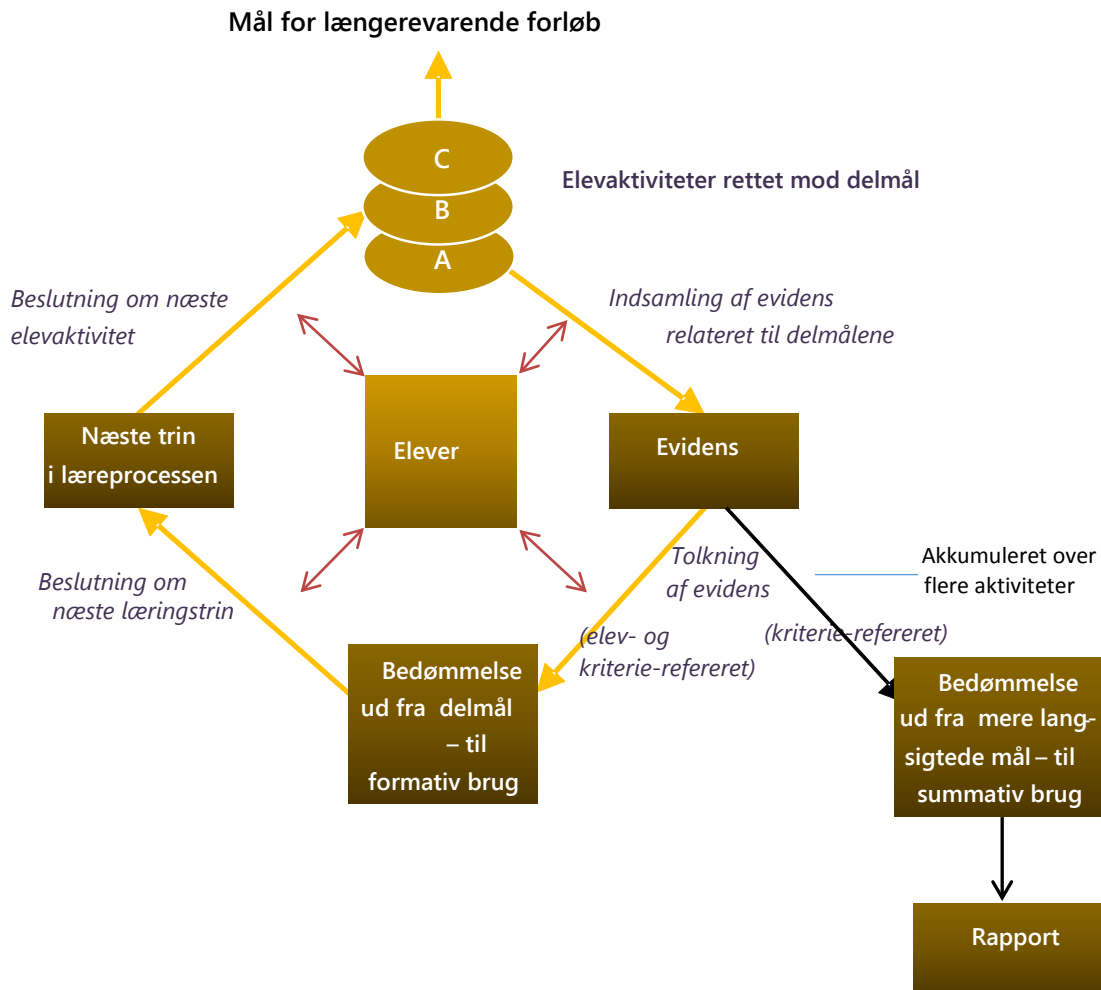
Undersøgelsen viser et hverdagsliv i gymnasiet med stort præstationspres og høj konkurrence. Ifølge eleverne er det, som stresser i gymnasiet, konstante høje krav, hvor de skal præstere både i skriftlige afleveringer (der kan ligge samtidigt i forskellige fag) og ved at forberede sig til og deltage i undervisningen (hvor der kan blive givet nye lektier for fra dag til dag), så mange af dem ofte må tage natten til hjælp for at nå det hele.

De anstrenger sig konstant for at præstere bedst muligt, fordi karakterer er afgørende for deres fremtid, videre uddannelse, karriere og succes i livet. Kampen for høje karakterer skaber konkurrence både fagligt og socialt: der kæmpes i klasserne om at blive bemærket som dygtige af lærerne, der bedømmer og tildeler karakterer. Man skal helst være sammen med de dygtige, så øges muligheden for at lærerne ser en selv som dygtig – og det er elever med høje karakterer, som andre vil lytte til og arbejde sammen med (Nielsen & Lagermann, 2017, s. 6).

Vi står altså i et evalueringsmæssigt spændingsfelt mellem på den ene side et ønske om at kunne evaluere en række ønskede kompetencer, både formativt for at fremme læring af dem, og summativt af accountability-hensyn, og på den anden side en virkelighed med dominerende summative evalueringer, der har en række meget negative effekter på såvel undervisningen som elevernes læring, og som ikke er i stand til at levere valide data.

Formativ og summativ brug af evalueringer

Spørgsmålet er, hvorledes det er muligt at opnå så stor alignment som muligt mellem mål, undervisning og evaluering. Her er vægningen og samspillet mellem formativ og summativ brug af evalueringer helt central. Forskellen på de to brug af evaluering fremgår af modellen i Figur 3.



Figur 3. Evalueringsprocessen for såvel formativ som summativ evaluering (baseret på Dolin, Black et al., 2017).

Modellen er udviklet af Wynne Harlen og videreudviklet i ASSIST-ME projektet (Dolin et al., 2017). Som det fremgår, kan al dokumentation af elevpræstationer (evidensen) bruges såvel formativt som summativt. Den afgørende forskel er hvilke kriterier, der ligger til grund for bedømmelsen, og hvad formålet med bedømmelsen er. Ved den summative brug af evalueringer bedømmes alle efter de samme kriterier, typisk opstillet i overensstemmelse med de faglige mål (kriterie-refereret), mens disse kriterier nuanceres ud fra den konkrete elevs optimale læreproces ved formativ brug af evalueringen (elev- og kriterie-refereret). Formålet med den summative evaluering er at give en

bedømmelse af eleven, typisk i form af en karakter, ud fra de mere overordnede opstillede kriterier, fx de faglige mål for et længerevarende undervisningsforløb, og at videregive denne bedømmelse til ekstern brug. Ved den formative evaluering bruges bedømmelsen til at sætte en reflektiv proces i gang hos eleven for at fremme læringen, således at eleven aktivitet for aktivitet nærmer sig de overordnede faglige mål. Ideelt set handler det om forskellige måder at forbinde den formative og summative brug af evalueringer på, idet det er ønskværdigt at udvikle en evalueringspraksis, der

- bruger tilpassede evalueringsformer, der validt kan indfange ønskede kompetencer og dermed bruges både formative og summativt
- har fokus på formativ brug af evalueringer, da forskning viser at dette har en stor læringsmæssig effekt (Black & Wiliam, 1998; Hattie & Timperley, 2007)
- og
- aktivt mindsker ukonstruktive effekter af presset fra de summative evalueringer.

I det nedenstående vil vi opridsse resultater og tematikker fra tre projekter: ASSIST-ME (*Assess Inquiry in Science, Technology and Mathematics Education*; Dolin, 2012; Dolin, 2016; Dolin & Evans, 2017), *Gymnasiet tænkt forfra* (Nielsen, 2015a, 2015b, 2015c) samt en undersøgelse af en række danske gymnasiers indsats omkring at nedtone summativ evaluering. Tilsammen indikerer forskningserfaringerne fra disse tre indsatser veje frem i forhold til at nå de tre ovenstående målsætninger for evalueringspraksis i naturfagene.

Erfaringer med progression og målorientering i formative evalueringsprocesser

ASSIST-ME var et fireårigt projekt, der involverede ti partnere fra otte lande. Det løb fra 2013-2016 og var finansieret af EU's FP7-program. Som nævnt vil vi her udelukkende fokusere på den del af projektet, der omhandlede arbejdet med at generere viden om hvordan lærere bedst arbejder med deres evalueringspraksis for at deres formative og summative evaluering kan indfange de komplekse kompetencer der er i spil når elever arbejder undersøgelsesbaseret. Mere overordnede resultater fra projektet kan findes i Dolin & Evans (2017).

ASSIST-ME-projektet udviklede en række evalueringsformer, der blev implementeret af lærere, der indgik i lokale arbejdsgrupper sammen med forskere. Den danske arbejdsgruppe bestod af 23 lærere fra såvel folkeskolen som gymnasiet og repræsenterende alle de største naturfag, inklusiv teknologi. Implementeringen af de valgte evalueringsformer forgik i tre runder, hver af et semesters varighed. Gruppen mødtes fire gange i semesteret, første gang i

august 2014 og sidste gang i december 2015. Såvel forskningsspørgsmål som metode er grundigt beskrevet i Dolin (2016), og vi vil her fremhæve de resultater, der knytter sig til formativ evaluering af komplekse kompetencer. Vi gennemførte to interviewrunder med lærerne, hvor lærerne i grupper på tværs af skoleform og skole skulle diskutere et sæt spørgsmål. Det første sæt omhandlede forholdet mellem formativ og summativ evaluering, og det andet sæt spørgsmål omhandlede arbejdet med de formative evalueringer, der endte med at blive et spørgsmål om progression og målorientering.

Angående det første, var der bred enighed om at eksaminer generelt ikke kan indfange de komplekse kompetencer, som læreplanerne kræver:

Jeg synes – når man ser på målene i læreplanerne, så er der så mange ting man ikke kan måle (ved eksamen).

Hvis vi ser på eksaminationen som den ser ud på nuværende tidspunkt, ift. at skulle måle de kompetencer vi arbejder med, der skal da ske en drastisk ændring i forhold til den måde man eksaminerer på nu, tænker jeg. Hvis vi skal afsøge det hele menneske, lidt mere bredt, så skal man have andre kompetencer i spil. Så er det vel ikke bare at trække et spørgsmål og en artikel, så er det vel også noget eksperimentelt.

Jeg synes det er et problem, at tit står der alle de kompetencemål, og ofte bliver eleverne målt fagfagligt. Der et stykke vej til at alle kompetencemålene bliver implementeret i evalueringen.

(Lærerudtalelser, citeret i Dolin, 2016).

En undtagelse herfra er teknologifaget, der har en projekteksamen, som blev anset for en relativ valid eksaminationsform.

Angående det andet spørgsmål, viste det sig hurtigt at det at undervise målrettet med formativ evaluering af elevers tilegnelse af bestemte kompetencer, forudsætter at man er i stand til at opstille nogle operationaliserbare mål, nogle kriterier for at de er opfyldt, og en måde at evaluere deres grad af opfyldelse på. Netop grad af opfyldelse er vigtig. Elevers vej til et mål vil være en proces hvor eleven via en række trin gradvist nærmer sig målet, således som det er illustreret i Figur 3. Undervisningen skal derfor være bygget op om en progression inden for det pågældende faglige område. Hvilket progressionstrin den enkelte elev er på, skal desuden kunne bestemmes, dvs. der skal være mulighed for at 'måle' elevens grad af mål-opfyldelse og sikre eleven feedback herpå, en feedback der også kan angive det næste skridt i progressionen. Det blev derfor et væsentligt fokuspunkt i projektet at understøtte lærernes arbejde med at udvikle konkrete læringsprogressionsskemaer for de kompetencer, som de underviste i.

Der viste sig tre didaktiske pointer i dette arbejde. Den første er at det for lærerne var uvant at skulle arbejde eksplicit med læringsprogression:

Jeg ville normalt aldrig skrive progressionstrin ned til mig selv. Vi kender dem jo, de sidder dybt i os, det skal da kun være til eleverne, så de bliver klar over kravene. Vi ved jo godt hvad vi gør.

Der er forskel på at arbejde med læringsprogression og at være bevidst om læringsprogression.

Jeg arbejder ikke så meget med det, men er nok bevidst om det. Jeg tænker progressionen meget klart ind i hvordan jeg opbygger en time, og i hvordan jeg spørger eleverne, men jeg har ikke før dette projekt tydeliggjort det for eleverne at der er denne her forskel i læringsprogression.
(Lærerudtalelser, citeret i Dolin, 2016).

Den anden pointe er, at de sædvanlige progressionsforståelser, som de fx er formuleret i læreplaner, ikke fungerede i praksis. En lærer forklarede hvorledes hans undervisning fulgte en faglig logik frem for den progression som udtrykkes i mange læringstaxonomier:

Det er noget andet end at bruge Bloomske termer, der vil gøre det uoverskueligt og uforståeligt. Det her er faglige mål. Man har fokus på de faglige mål frem for læringsmål.

Det viste sig, at de fleste lærere i deres planlægning af undervisningen, og dermed i den progression, de stillede op, ikke fulgte en generisk taxonomi som fx Blooms taxonomi eller SOLO-taxonomien, på trods af at læreplanerne i vid udstrækning er beskrevet i disse termer. De fleste lærere så progression mere som et puslespil, hvor man starter med nogle karakteristiske brikker, for så efterhånden at fylde hele billedet ud. Og dette puslespil skulle designes konkret for hvert fagligt forløb. Dette er i overensstemmelse med andres erfaringer, som viser at lærere må opstille en progression for hver enkelt fagområde, der undervises i (Alonzo & Gotwals, 2012). Dette er et meget tidskrævende arbejde, som kun få lærere er i stand til at udføre alene.

Den tredje didaktiske pointe er nok den mest dramatiske, set i lyset af at det øgede fokus på formative evalueringer gerne skulle fremme en bedre læreproces for eleverne. Ikke alle lærerne opfattede det positivt at arbejde så indgående med at operationalisere læringsmålene i progressionsskemaer:

Jeg tror det er en uheldig ting, hvis man laver en undervisning, hvor progressionstrinnene bliver fuldstændigt pindet ud. Så mister man noget – så går det over mod præstation i stedet for at handle om læring. Jo mere man pinder det ud, jo mere fokus bliver der på at nu har jeg fået to plus fem kompetencen i hus. Hvis vi hele tiden fokuserer på hvordan de kan evaluere sig selv, så fjerner vi nysgerrigheden og glæden og fokuserer på at nu skal I også præstere.
(Lærerudtalelser, citeret i Dolin, 2016).

På den måde peger erfaringerne fra ASSIST-ME projektet på et underliggende dilemma mellem på den ene side at skulle arbejde med at operationalisere kompetencemål (for at understøtte gode læreprocesser) og på den anden side faren for at evaluering bliver instrumentalistisk og fremmer en præstationsorientering hos eleverne, frem for den mestring, som gerne skulle være kernen i de formative processer.

Disse to motivationstyper kan defineres således:

Mestringsmål fokuserer personen på den aktuelle opgave og relaterer sig især til at udvikle kompetence og opnå forståelse og indsigt. Præstationsmål fokuserer personen på selvet og relaterer sig især til hvorledes evner bliver bedømt, og hvorledes man selv præsterer, især i forhold til andre.

(Midgley, Kaplan & Middleton, 2001, s. 77, egen oversættelse.)

Her er det vigtigt at huske, at forskellige målorienteringer hos elever giver forskellige handlinger i læringsituationer og dermed forskelligt læringsudbytte. Som Harlen (2012) udtrykker det:

Der er stærk evidens for at ydre belønning underminerer den indre motivation. [...] Der er indikationer på at ydre motivation leder til 'overflade'-læring snarere end 'dyb' læring. (Harlen, 2012, s. 174, egen oversættelse.)

Det viste sig altså, at stor fokus på formativ evaluering kunne risikere at fremme uhensigtsmæssige motivationsorienteringer hos elever, den selv samme motivationsorientering som følger af stærk fokus på summative evalueringer! Dette skal ikke forstås som, at den formative evaluering var en skjult summativ evaluering; men at en meget detaljeret formativ evalueringspraksis (fx baseret på, at læreren arbejde med detaljerede tjeklister for hvad eleverne skal kunne) kan blive for instrumentel (se Torrance, 2007) stik imod en lærers intention om at lave formativ evaluering af høj kvalitet.

Det handler således om at finde en balance mellem formativ brug af evaluering, der ikke er for stramt styret og instrumentel, og summativ evaluering, der rent faktisk bedømmer i hvilken grad elever mestrer de mere komplekse kompetencer. Pointen ved det sidste er, at hvis det elever bedømmes på i den summative evaluering er deres mestring af en kompetence, vil dette fjerne den primære bevæggrund til præstationsorientering, fordi de er nødt til at mestre kompetencerne for at kunne klare sig godt til eksamen.

Der er derfor brug for eksamensformer der involverer relevante situationer som elever kan aktivere deres kompetencer i. Det vil næste afsnit give et eksempel på.

Erfaringer om udviklingen af et fælles sprog om kompetenceevaluering

Vi ser nu på et andet større projekt: *Gymnasiet tænkt forfra* (2012-2015). Vi mener, at projektet indikerer en vej fremad når det gælder at udvikle og gennemføre kompetenceorienteret undervisning og evaluere kompetencetilegnelse (projektet er beskrevet nærmere i Nielsen, 2015a, 2015b, 2015c; Nielsen & Dolin, 2016).

Projektet involverede seks gymnasier i Region Hovedstaden, og det primære sigte var at understøtte elever i at udvikle fagligt funderede innovationskompetencer og større motivation og lyst til at lære. I kraft af at innovationskompetencer på dette tidspunkt stadig var relativt ubeskrevne og løst definerede i det danske uddannelsessystem, blev projektet fra starten et projekt, der handlede om hvordan de deltagende lærere italesatte, underviste i, og evaluerede på kompetencer som sådan. Igennem projektet opstod der på alle skoler meget fasttømrede lærerteams, der kontinuerligt forholdte sig til, hvordan en given kompetence undervises og evalueres. Disse pædagogisk-didaktiske refleksioner i lærerteam ledte til et fælles sprog om kompetenceudvikling. I processen om at etablere det fælles sprog om kompetenceudvikling kunne følgeforskningen indkredse en liste af tegn på kompetenceudvikling når det gælder innovationskompetence (se Nielsen, 2015b).

Lærernes fælles sprog om kompetenceudvikling forplantede sig til de deltagende elever. Lærerne bragte med andre ord det nye sprog med ind i deres klasser, og etablerede derigennem en fælles forståelsesramme med de deltagende elever om hvad det vil sige at udvikle en given kompetence (se Nielsen, 2015a; Nielsen & Dolin 2016). Der er tydelige indikationer på, at de deltagende elever gennem deres gymnasietid tilegnede sig dette sprog (Nielsen, 2015a) igennem en meningsforhandling som må være en essentiel del af at anerkende opstillede evalueringskriterier (for en diskussion af dette se Dysthe et al., 2008). Her er det væsentligt at bemærke, at følgeforskningen i projektet viste, at de deltagende elever i modsætning til tilsvarende kontrol elever over tid udviklede antræk til mestringsorientering. Denne udvikling kunne både ses i elevers spørgeskemabesvarelser samt i elev- og lærerinterviews (se Nielsen, 2015a; Nielsen & Dolin, 2016).

En anden væsentlig erfaring fra projektet var at selv komplekse generiske kompetencer kunne evalueres pålideligt i eksamenslignende situationer – selv af lærere uden omfattende erfaring med kompetencerne (for en detaljeret rapport se Nielsen, 2015c). Som en del af projektet blev der etableret et forsøg med at teste fem forskellige eksamensformer i fem fag (afsætning, biologi, dansk, matematik og samfundsfag). I alle eksamensformer indgik et element af at elever skulle forholde sig fagligt til en problemstilling der rækker udover faget selv. For eksempel forløb eksamensformen i biologi over to dage, hvor elever i grupper på den første dag undersøgte og foreslog løsninger til en problemstilling

omkring forringelse af økosystemet i og omkring søer som resultat af fodring af ænder. Bedømmerne var til stede under dette arbejde og havde således adgang til at følge elevernes arbejdsproces. På den anden dag var hver elev individuelt til en mundtlig eksamen der tog udgangspunkt i undersøgelsen dagen før og udviklede sig til en faglig dialog som det er kendt fra vanlige mundtlige eksaminer.

Tre af de afprøvede eksamensformer indeholdt et element af at bedømmerne observerede elevernes undersøgelser af problemstillingen og arbejdede med at opstille løsningsforslag. To eksamensformer var mundtlige eksaminer hvor det undersøgende arbejde og arbejdet med at opstille løsningsforslag foregik inden selve eksaminationen, og hvor bedømmerne kun havde adgang til processen gennem en rapport fra eleverne. Til hver afprøvning af eksamensform deltog tre til fire bedømmere, der uafhængigt af hinanden gav den enkelte elev karakterer på op til fem dimensioner af innovationskompetence ud fra et prædefineret fagspecifikt progressionsskema. Minimum to af disse bedømmere havde ingen videre forudsætninger for at bedømme innovationskompetence og havde ikke nogen tilknytning til projektet. Resultaterne viser entydigt at pålideligheden af bedømmelserne i de afprøvede eksamensformer var tilfredsstillende til meget høj. Der blev både testet for pålidelighed forstået som *enighed* (hvor tæt er bedømmernes vurderinger af den samme elev?) og som *konsistens* (i hvor høj grad følger bedømmerne samme interne logik eller skala?). Det var med andre ord muligt for bedømmerne at nå til pålidelige summative vurderinger af elevernes innovationskompetence.

Endvidere viste analyser af interviews med bedømmerne, at de bedømmere der kunne følge elevernes arbejdsproces mente at de mere validt kunne nå til en vurdering af elevernes kompetencer, men de kvantitative resultater viste dog, at der ikke var forskel på pålideligheden afhængigt af, om bedømmerne havde adgang til at observere elevernes arbejdsproces.

Vi mener, at disse erfaringer indikerer at kompetenceorienteret undervisning bør basere sig på et fælles kompetencesprog blandt lærere – fx i faglige eller tværfaglige teams – for at kunne understøtte en konstruktiv evalueringspraksis.

Erfaringer med at nedtone karakterer i et karakterdrevet uddannelsessystem

De tidligere nævnte negative konsekvenser af det øgede karakterpres betød at en række danske gymnasier vedtog indsatser som skulle forbedre den formative evaluering og feedback og lægge mindre vægt på karakterer ved uddannelsens start. I skoleåret 2016/17 var der således to gymnasier der havde fået Undervisningsministeriets tilladelse til forsøg med karakterfritagelse og en række andre, der inden for den gældende bekendtgørelse havde forskellige tiltag

til at nedtone karaktererne – typisk således at man kun gav de karakterer, der var foreskrevne ved lov.

For at danne os et overblik over disse indsatser og de deltagende aktørers erfaringer, sendte vi et åbent spørgeskema til ledere og involverede lærere fra de to forsøgsgymnasier og 12 andre gymnasier, som ministeriet var bekendt med havde en relevant indsats i 2016/17. Vi fik svar fra 12 gymnasier (incl de to forsøgsgymnasier), i alt 12 ledelser og 27 lærere. For ikke at belaste skolerne unødigt havde vi bedt ledelserne om at udvælge to til fire lærere, der kunne repræsentere forsøget. En PhD-studerende gennemgik og kategoriserede de indsamlede fritekstsvar med henblik på at identificere gennemgående tematikker (i stil med tematisk analyse beskrevet af Braun & Clark, 2006).

Lærerevalueringer

Lærerne beskriver forskellige metoder/arbejdsformer til den formative feedback, der i højere grad inddrager eleverne i evalueringsprocessen og hjælper den enkelte elev til at forstå sit eget niveau og guide eleven videre, som fx ”elevarbejde med portefølje som kan have et dagligt fokus”, ”evalueringssamtaler, hvor eleverne har evalueret deres egen faglige indsats og fået feedback/forward af os uden at tale karakterer”, ”systematisk [arbejde] med mundtlige (screencast optagne) feedback kommentarer”. Størstedelen af lærerne beskriver, hvordan de i klasserne arbejdede med fokuspunkter, og at de giver formativ feedback relateret til fælles delmål, som fx ”ved retning af elevprodukter bruges fokuspunkter som kommenteres” eller ”vi arbejder [...] med at de [eleverne] sætter sig et mål (og formulerer en masse refleksioner om, hvordan de skal nå målet, hvad der kan forhindre dem i at nå målet osv.), som de arbejder målrettet med, og som de får fokuseret feedback på”. Mange af undersøgelsens lærere beskriver, hvordan indsatsen har fokus på, at ”man bliver dygtigere i fællesskab”, og at det er okay at begå fejl, som en lærer beskriver, en ”italesættelse af at eleverne er i en læreproces i den daglige undervisning og læring sker gennem at have modet til at prøve at fejle”. Andre lærere beskriver indsatsen vedrørende det mundtlige arbejde ”mundtligt gjaldt det endvidere at [...] klassen havde ambitiøse elever (mange potentielle 12er piger [12 er den højeste karakter på den danske karakterskala]) som hellere ville dø end at sige noget ”forkert” på klassen, så her arbejdede vi meget med klasserumskultur, peer feedback [...]”, og ”det handler for mig om at få skabt et trygt undervisningsrum, hvor det er okay at fejle”.

I lærernes beskrivelser af indsatsen er det tydeligt at der er tale om en kulturændring på godt og ondt for eleverne: ”Mange af eleverne har ”manglet karakterer” forstået på den måde, at en karakter i sig selv har været motiverende for nogle af eleverne, sikkert fordi det er det, de kender, og er vant til” og eleverne reagerede ... i starten på, at de ikke fik standpunktskarakterer. Det ville de gerne have haft”. Der er dog indikationer på, at elever kan vænne sig til den nye kultur: ”Indledningsvist var der stor frustration blandt eleverne, der

manglede et tal som vurdering af deres indsats. Den indledende frustration er dog vendt til accept og ”eleverne har lært at se det som en fordel, men det har været en hård kamp” eller som en anden lærer siger: ”Jeg oplevede lidt frustration i starten omkring manglende karakterer, men dette blev hos langt de fleste noget der blev sat pris på”.

En lærer beskriver, at vedkommende stadig tror, at eleverne går meget op i karakterer på trods af deres indsats, også på grund af samfundets fokus på præstation:

Jeg tror stadig, at eleverne går meget op i karakterer. Karakterspøgelset er der, nærmest uanset hvad man gør. Man kan godt tale det ned i timerne, og det hjælper måske, men der er også nogle strukturer (konkurrencestaten, uddannelsessystemets opbygning og vægtning af eksamenskarakterer mv.), som er svære at undgå helt. Det er en skam, for det evindelige fokus på karakterer gør undervisningen til et ”spil” og en ”kamp” fremfor et fælles læringsprojekt. Jeg er glad for, at min skole sætter emnet på dagsordenen med formativ evaluering.

I lærernes beskrivelser af effekten af arbejdet med nedtoningen af karakterer på eleverne i det daglige, nævner flere netop den ovennævnte præstationsorientering som værende mindsket: ”Med hensyn til eleverne er det min opfattelse, at fokus er flyttet fra karakteren til indholdet – hvad kan forbedres” og ”det var helt tydeligt, at der kun var ganske få, der var præstationsmotiverede. På den måde har arbejdet med formativ evaluering og motivation spillet godt sammen”. En anden lærer siger: ”Evaluering er fremadrettet, og jeg oplever et mindsket stressniveau ved tilbagelevering af afleveringer. Disse ses som en del af en længere proces og ikke blot et færdigt produkt [...]. Jeg oplever, at der har været en større ro omkring arbejdet med det faglige, da de ikke er blevet vurderet. Dvs. at fejl kun giver læring og ikke en dårlig karakter. Det gør det heller ikke i praksis, men det kan være svært at overbevise eleverne om”. Det beskrives, hvordan eleverne ”tænker fremad og tager kritik mindre personligt, end man ofte ser det i andre klasser, hvilket åbner op for bedre dialog om læring”, og det har i nogle klasser særligt haft en effekt på ”det at turde fejle”. Nogle lærere gengiver, hvordan indsatsen har taget presset af det første halve eller hele år af elevernes gymnasietid: ”Eleverne får mulighed for at finde sig tilpas på skolen, uden at de skal forholde sig til et specifikt tal i de forskellige fag” eller ”jeg oplever et mindre stressniveau blandt eleverne sammenlignet med tidligere 1.g hold”, og en lærer beskriver, hvordan eleverne ”ikke i så hurtig grad fik udviklet et konkurrenceforhold i klassen. Klassen blev ikke delt i de dygtige og de mindre dygtige.” Sammenlignet med et normalt stressniveau for en afgangsklasse har det ifølge en anden lærer ligeledes haft en effekt: ”Klassen virker som helhed mindre stresset, end man ofte ser det sidst på semestret i 3.g”. En lærer beskriver, at eleverne stadig er stressede, men at indsatsen har haft en indflydelse på stressniveauet: ”Stressniveauet er

stadigvæk for højt i klassen, men nok lidt lavere end hvis ikke der var gjort denne kæmpe indsats for at dæmpe præstationslysten”.

Ifølge lærerne i denne undersøgelse, er eleverne blevet dygtigere til at analysere styrker og svagheder i egne og andres arbejde, skriftligt såvel som mundtligt og er blevet bedre til at reflektere over egen læring og indsats: ”I samtaler med eleverne er de meget reflekterede om sammenhængen mellem arbejdsindsats, motivation og læring, og de fleste har en positiv tilgang til de udfordringer de møder”, og eleverne er ”blevet bedre til at forstå feedback, at give konstruktiv feedback, at reflektere over egen læring og til at sammenligne sig mindre og koncentrere sig mere om egne mål” og at ”både elever og lærer bliver mere opmærksomme på arbejdsprocessen end selve resultatet”. Ifølge nogle lærere udviste eleverne en større glæde ved at få brugbar, gennemskuelig feedback.

I lærernes beskrivelser af indsatsens effekt for dem selv fremhæves lærernes øgede arbejdsglæde idet fokus på karakterer bliver nedtonet: ”Min egen arbejdsglæde er uden tvivl højere, fordi jeg helt generelt ikke bryder mig om karakterer” og ”jeg synes, det er rart at fokusere på elevens faglige udvikling fremfor at måle niveauet”. Andre lærere gengiver at ”det var en lettelse ikke at skulle give karakterer og dermed også skulle diskutere karakterer med eleverne” og ”det er også rart som lærer, at man kan lære eleverne at kende, både fagligt og socialt, inden man skal give de første karakterer” og ”jeg oplever også selv, at mit fokus har ændret sig til et mere fremadrettet fokus – hvordan skal denne elev hjælpes videre? Det bidrager til min arbejdsglæde, at vi (lærer og elev) kan have fokus på faglig fremgang og proces fremfor karakteren. Det giver en bedre lærer-elev relation”. Lærersamarbejdet i indsatsen fremhæves derudover som godt og motiverende: ”Dejligt at kunne samarbejde og udvikle didaktik på tværs af fag og fakulteter”.

Dog er lærernes udsagn ofte iblandet problemer med at realisere den karakterfri indsats, typisk i form af tidsforbrug og arbejdsomfang. Det påpeges, at en karakterfri indsats ”kræver systematisk tilføjelse af ressourcer i form af tid (der ikke må tages fra undervisningen eller forberedelse)”. Flere lærere nævner netop tidsaspektet forbundet med den formative evaluering: ”Som lærer kræves der dog ekstra tid, da eleverne ’piller’ mine kommentarer fra hinanden i søgningen efter en bedømmelse – hvilket betyder, at der skal reflekteres meget mere over den skriftlige feedback” og ”det er en spændende udfordring at arbejde med formativ feedback, men også enormt tidskrævende og svært at prioritere i en tid med nedskæringer og besparelser, hvor man skal have flere hold og nå mere på kortere tid. God undervisning fås ikke for færre penge” og ”det kan være en hård omgang at få talt med alle eleverne (...) det er et ekstra arbejde for læreren, som det kan være svært at vurdere frugten af endnu.”

Eleversundersøgelse

Parallelt med undersøgelsen blandt de deltagende lærere, bad vi lærerne om at få deltagende elever (samt, om muligt, kontrolelever) til at udfylde et kort spørgeskema for at undersøge indikatorer på motivation og stress. Som instrument for undersøgelsen af motivation valgte vi fem skalaer fra et internationalt valideret spørgeskema *Motivation and Engagement Scale* (Martin, 2007) om adaptive og maladaptive tankemønstre omkring skolearbejdet. Disse data er i skrivende stund kun underlagt en præliminær analyse. På skalaniveau ser der ikke ud til at være signifikante forskelle mellem effekt- og kontrolelever. Ser man på enkeltspørgsmål i skalaerne for motivation finder vi antræk til et mønster, som dog kræver en større undersøgelse for bekræftelse. Som det fremgår af Tabel 1 er 45% af effekteleverne mere eller mindre uenige i at de er usikre på hvordan en eventuel god karakter kan opnås igen. Tilsvarende er kun 21,3% af kontroleleverne mere eller mindre uenige i dette. En over dobbelt så stor andel af effekteleverne end af kontroleleverne føler sig med andre ord sikre på hvordan en god præstation gentages. Dette kan være et væsentligt fingerpeg på, at effektelevernes undervisning har integreret en mere velfungerende formativ evalueringspraksis.

Tabel 1. Besvarelsesfrekvens for enkeltspørgsmålet ”Når jeg får en god karakter, er jeg ofte usikker på, hvordan jeg får den karakter igen”. N (effekt) = 280; N (kontrol) = 47.

	Effekt		Kontrol	
	%	Σ%	%	Σ%
Meget uenig	5,0	5,0	6,4	6,4
Uenig	24,3	29,3	4,3	10,6
Lidt uenig	15,7	45,0	10,6	21,3
Hverken enig eller uenig	21,1	66,1	27,7	48,9
Lidt enig	17,1	83,2	23,4	72,3
Enig	10,7	93,9	14,9	87,2
Meget enig	6,1	100,0	12,8	100,0

Tilsvarende finder vi som angivet i Tabel 2, at 46,8% af effekteleverne er mere eller mindre uenige i at de er usikre på hvordan de kan undgå at gøre deres skolearbejde dårligt. Tilsvarende er 29,8% af kontroleleverne mere eller mindre uenige i dette. Dette kan også være en indikation på, at effektelevernes undervisning har integreret en mere velfungerende formativ evalueringspraksis.

Tabel 2. Besvarelsesfrekvens for enkeltspørgsmålet ”Jeg er ofte usikker på, hvordan jeg kan undgå at gøre det dårligt i gymnasiet”. N (effekt) = 280; N (kontrol) = 47.

	Effekt		Kontrol	
	%	Σ%	%	Σ%
Meget uenig	7,9	7,9	10,6	10,6
Uenig	26,4	34,3	6,4	17,0
Lidt uenig	12,5	46,8	12,8	29,8
Hverken enig eller uenig	19,6	66,4	23,4	53,2
Lidt enig	16,4	82,9	23,4	76,6
Enig	10,4	93,2	17,0	93,6
Meget enig	6,8	100,0	6,4	100,0

Som det fremgår af Tabel 3 og Tabel 4 er der indikationer på, at effekteleverne er mindre drevet af frygten for nederlag. Således er der fx 17,9% (100-82,1) af effekteleverne der er mere eller mindre enige i at hovedårsagen til at de laver deres skolearbejde er fordi de ikke vil have at folk tænker dårligt om dem. Den tilsvarende andel af kontrolelever er 29,8% (100-70,2). Endvidere er 45% (100-55) af effekteleverne mere eller mindre enige i at hovedårsagen til at de laver deres skolearbejde er fordi de ikke vil have at deres lærer tænker dårligt om dem. Den tilsvarende andel af kontrolelever er 53,2% (100-46,8).

Tabel 3. Besvarelsesfrekvens for enkeltspørgsmålet ”Hovedårsagen til at jeg laver mit skolearbejde er ofte, at jeg ikke vil have folk til at tænke dårligt om mig”. N (effekt) = 280; N (kontrol) = 47.

	Effekt		Kontrol	
	%	∑%	%	∑ %
Meget uenig	7,9	7,9	4,3	4,3
Uenig	16,4	24,3	10,6	14,9
Lidt uenig	8,9	33,2	10,6	25,5
Hverken enig eller uenig	21,8	55,0	21,3	46,8
Lidt enig	22,5	77,5	14,9	61,7
Enig	15,4	92,9	23,4	85,1
Meget enig	7,1	100,0	14,9	100,0

Tabel 4. Besvarelsesfrekvens for enkeltspørgsmålet ”Hovedårsagen til at jeg laver mit skolearbejde, er ofte fordi jeg ikke vil have at min lærer tænker dårligt om mig”. N (effekt) = 280; N (kontrol) = 47.

	Effekt		Kontrol	
	%	∑%	%	∑ %
Meget uenig	26,1	26,1	19,1	19,1
Uenig	26,4	52,5	21,3	40,4
Lidt uenig	9,3	61,8	17,0	57,4
Hverken enig eller uenig	20,4	82,1	12,8	70,2
Lidt enig	7,5	89,6	14,9	85,1
Enig	6,8	96,4	4,3	89,4
Meget enig	3,6	100,0	10,6	100,0

Konklusioner og perspektiver

Naturfagene har ændret sig afgørende i løbet af de sidste 40-50 år, dels som reaktion på samfundsmæssige krav og forandringer og dels påvirket af naturfagsdidaktisk forskning. Der er således formuleret grundlæggende nye formål i form af såvel faglige som generiske kompetencer og udviklet undersøgelsesbaserede undervisningsformer, der gør det muligt for eleverne at tilegne sig disse kompetencer.

Men evalueringsformerne og brugen af dem til summative formål har ikke ændret sig tilsvarende. Der er stadig et misforhold mellem en kompetenceorienteret undervisning og eksamensformer der ikke er i stand til at

indfange kompetencerne, ligesom den formative brug af evaluering, evaluering for læring, overskygges af den summative brug, evaluering af læring. Dominansen af summative evalueringer har desuden en lang række negative effekter på såvel eleverne som på undervisningen.

Artiklen har dokumenteret disse misforhold og ved at trække på resultater fra tre forskellige forsknings- og udviklingsprojekter omkring evaluering har vi vist at:

- En nødvendig forudsætning for at opstillede kompetencemål indfries og undersøgelsesbaserede arbejdsformer får bedre vilkår, er at der etableres en passende balance mellem evalueringernes summative og formative brug med udgangspunkt i en forståelse af deres forskelligheder.
- Det er især nødvendigt at mindske læringshæmmende effekter af summative evalueringer og arbejde på at øge den formative brug af evalueringer.
- Udviklingen af et (lokalt) fælles sprog om de kompetencer, der skal evalueres, understøtter læreres praksis i at undervise og evaluere kompetencerne, og bidrager til, at eleverne bliver bedre til at forholde sig til deres egen læring.
- Dækkende beskrivelser af de kompetencer, der skal evalueres kan bidrage til at kompetencerne vurderes pålideligt.
- Hvis det fælles sprog og beskrivelserne af de kompetencer, der skal evalueres, bliver for detaljeret er der en risiko for, at evalueringens positive effekt på elevernes læring udvandes.
- Arbejdet med at nedtone karakterfokus i et ellers meget karakterdrevet system anerkendes generelt som positivt af lærere; og de elever der udsættes for færre karakterer ser ifølge lærerne ud til at blive bedre til at forholde sig til deres egen læring.
- Arbejdet med at nedtone karakterfokus i et ellers meget karakterdrevet system er forbundet med en række udfordringer, herunder at det som lærer kan være tidskrævende at fokusere på formativ evaluering, hvor man normalt blot ville give en karakter.
- Det er sandsynligt, at elever der udsættes for en nedtoning af karakterfokus med tiden udvikler en mere mestringsbetonet motivation for deres skolearbejde, samt at det nedtonede karakterfokus kan være med til at understøtte en undervisning med en bedre integreret formativ evalueringspraksis.

Realiseringen af disse forudsætninger og muligheder kræver koordinerede indsatser på et systemisk niveau. Det fremgik bl.a. af det arbejde, der i 2016/17 blev udført for at udarbejde en strategi for de naturvidenskabelige fag i det

danske uddannelsessystem (se <http://astra.dk/nyhed/nyt-strategiforslag-skal-styrke-undervisning-i-naturvidenskab>). Her samlede Undervisningsministeriet en såkaldt strategigruppe på 44 personer repræsenterende et bredt udvalgt af organisationer og fagmiljøer. Arbejdet blev ledet af et formandskab på 5 personer og gennem hjemmeside-input, bilaterale møder med udvalgte interessenter og møder i strategigruppen blev det tydeligt hvor mange interesser, der er i spil, når der skal opnås konsensus om uddannelsesforhold. Ud over traditionelle modsætninger mellem lærere og skoleledere og skoleejere blev der afdækket mere værdimæssige forskelle, som gik på kryds og tværs af interessentgrupperne. Det blev på den ene side klart, at uddannelsesforskningen havde en stor rolle at spille ved at kunne levere neutrale input til processen, men det var også klart at værdier og følelser og politiske holdninger ofte spillede en lige så stor rolle som forskningen, når der skulle tages beslutninger. Forskning i sig selv er derfor ikke nok til at ændre uddannelsessystemet. Lige så vigtigt er det at forskerne forstår at arbejde sammen med og formidle sine resultater til alle de, der skal være med til at tage beslutningerne (Dolin et al., 2017). Dette gælder ikke mindst vedrørende så ømtålelige områder som evaluering, hvor et centralt element vil blive forskning i at udvikle en balanceret evalueringskultur, der har fokus på formative processer og som samtidig kan levere valide summative data til det politiske system.

Om forfatterne

Jens Dolin er professor emeritus i naturfagsdidaktik ved Københavns Universitet. Han forsker i undervisning og læring af naturvidenskab, med fokus på evaluering og kompetenceudvikling, og i organisatorisk udvikling, med fokus på reform processer, curriculum udvikling, læreruddannelse og lærerefteruddannelse.

Institution: Institut for Naturfagernes Didaktik, Københavns Universitet, Øster Voldgade 3, DK-1350 København K

Epost: dolin@ind.ku.dk

Jan Alexis Nielsen (ph.d.) er lektor i naturfagernes didaktik. Hans forskning centrerer sig om evaluering af komplekse kompetencer i naturfagene, balancering mellem tværfaglige/generiske kompetencemål og fagspecifikke læringsmål, og inddragelsen af samfundsforhold og autentiske problemer i naturfagsundervisningen.

Institution: Institut for Naturfagernes Didaktik, Københavns Universitet, Øster Voldgade 3, DK-1350 København K

Epost: janielsen@ind.ku.dk

Referencer

- AAAS (1990). *Science for All Americans*, Cary, NC 27513: Oxford University Press.
<http://www.project2061.org/publications/sfaa/online/sfaatoc.htm?txtRef=&txtURIId=%2Ftools%2Fsfaaol%2Fsfatoc.htm> (lokalisert 15. april 2017).
- Alonzo, A.C. & Gotwals, A.W. (2012). *Learning progressions in science. Current challenges and future directions*. Rotterdam: Sense Publishers.
- Andersen, N. O., Busch, H., Horst, S., & Troelsen, R. (2003). *Fremtidens naturfaglige uddannelser Naturfag for alle - vision og oplæg til strategi*. København: Undervisningsministeriet.
- Bell, B. (2007). Classroom Assessment of Science Learning. In S. K. Abell & Lederman, N. (Eds.), *Handbook of Research on Science Education* (pp. 965-1006). Mahwah, N.J.: Lawrence Erlbaum Associates.
- Black, P. & Wiliam, D. (1998). *Inside the black box*. London: King's College London.
- Black, P., Harrison, C., & Lee, C. (2004). *Working inside the black box: Assessment for learning in the classroom*. London: Granada Learning.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101.
- Bruner, J. (1996). *The Culture of Education*. Cambridge, Massachusetts: Harvard University Press.
- Bundsgaard, J. (2016). *Test måler det forkerte*. Åben høring i Folketingets Børne- og undervisningsudvalg om de nationale test 14. september 2016.
<http://www.ft.dk/webtv/video/20151/buu/tv.3353.aspx?as=1#pv> (lokalisert 15. april 2017).
- Butler, R. (1987). Task-involving and ego-involving properties of evaluation: Effects of different feedback conditions on motivational perceptions, interest, and performance. *Journal of Educational Psychology*, 79(4), 474-482.
- Butler, R. (1988). Enhancing and undermining intrinsic motivation: The effects of task-involving and ego-involving evaluation on interest and performance. *British Journal of Educational Psychology*, 58(1), 1-14.
- Dewey, J. (1910). Science as Subject-Matter and as Method. *Science*, 31(787), 121-127.
- Dolin, J. (ed.) (2012). ASSIST-ME proposal. www.assistme.ku.dk. (lokalisert 15. april 2017).
- Dolin, J. (2016). Idealer og realiteter i målorientert undervisning. I: Krogh, E. & Holgersen, S.E. (2016) (red.). Sammenlignende fagdidaktik 34. *Cursiv 19*. s. 67-87.
www.edu.au.dk/cursiv (lokalisert 15. april 2017).
- Dolin, J. & Evans, R. (eds.) (2017). *Transforming assessment – through interplay between practice, research and policy*. Cham, Switzerland: Springer.
- Dolin, J., Black, P., Harlen, W. & Tiberghien, A. (2017). Exploring relations between formative and summative assessment. I: Dolin, J. & Evans, R. (eds.) (2017). *Transforming assessment – through an interplay between practice, research and policy*. Springer.
- Dolin, J., Bruun, J., Constantinou, C., Dillon, J., Jorde, D. & Labudde, P. (2017). Policy aspects – how to change practice and in what direction. I: Dolin, J. & Evans, R. (eds.) (2017). *Transforming assessment – through an interplay between practice, research and policy*. Cham, Switzerland: Springer.
- Driver, R. (1983). *Pupil as scientist*. Milton Keynes: Open University Press.
- Driver, R. (1985). *Children's ideas in science*. Milton Keynes: Open University Press.
- Driver, R., Leach, J., & Millar, R. (1996). *Young people's images of science*. Buckingham: Open University Press.

- Duit, R. (2015). Didaktik. I: Gunstone, R. (ed). *Encyclopedia of Science Education*. Cham, Switzerland: Springer.
- Dysthe, O., Engelsen, K. S., Madsen, T., & Wittek, L. (2008). A theory-based discussion of assessment criteria – The balance between explicitness and negotiation. I: A. Havnes & L. McDowell (Eds.). *Balancing dilemmas in assessment and learning in contemporary education* (pp. 121-131). New York: Routledge.
- EACEA; Eurydice (2009). *National Testing of Pupils in Europe: Objectives, Organisation and Use of Results*. Brussels: Eurydice.
- Elmeskov, D. C., Bruun, J., & Nielsen, J. A. (2015). Evaluering af bioteknologi A som forsøgsfag i stx og htx. *MONA Forskningsrapportserie, 1*(1), 1-99.
- European Commission (2007). *Science Education Now: A Renewed Pedagogy for the Future of Europe*. Brussels, European Commission. http://ec.europa.eu/research/science-society/document_library/pdf_06/report-rocard-on-science-education_en.pdf (lokaliseret 15. april 2017)
- Fensham, P. J. (2004). *Defining an Identity: The Evolution of Science Education as a Field of Research*. Dordrecht: Springer.
- Gipps, C. (1999). Socio-Cultural Aspects of Assessment. *Review of Research in Education, 24*, 355-392.
- Greve, C. (2002). *New public management*. København: Nordisk KulturInstitut.
- Harlen, W. (2007). *Assessment of Learning*. London: Sage Publications Ltd.
- Harlen, W. (2012). The role of assessment in developing motivation for Learning. I: J. Gardner (red). *Assessment and Learning* (s. 61-80). London: SAGE Publications Ltd.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of educational research, 77*(1), 81-112.
- Krogh, E. (2017). Fagdidaktik. I: Dolin, J., Ingerslev, G.H. & Jørgensen, H.S. (red.). *Gymnasiepædagogik, en grundbog* (s. 561-576). København: Hans Reitzels Forlag.
- Krogh, L. B., Arnborg, P. & Thomsen, P. V. (2001). *Hvordan gik det så med fysikundervisningen og elevernes udbytte?* Aarhus Universitet: CND's skriftserie, no. 3.
- Krogh, L. B. & Dolin, J. (2016). PISA 2006 Science testen og danske elevers naturfaglige formåen. *MONAs Forskningsrapportserie, 2*. <http://www.ind.ku.dk/mona/serie/2016-2/>.
- Lemke, J. L. (1990). *Talking Science: Language, Learning, and Values*. Norwood, NJ: Ablex Publishing Corp.
- Martin, A. J. (2007). Examining a multidimensional model of student motivation and engagement using a construct validation approach. *British Journal of Educational Psychology, 77*(2), 413-440. doi:10.1348/000709906X118036
- Midgley, C., Kaplan, A. & Middleton, M. (2001). Performance-approach goals: Good for what, for whom, under what circumstances, and at what cost? *Journal of Educational Psychology, 93*(1), 77-86.
- Millar, R. & Osborne, J. (1998). *Beyond 2000: Science education for the future*. London, King's College London.
- Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). On the Structure of Educational Assessments. *Measurement: Interdisciplinary Research and Perspectives, 1*, 3-62.
- Nielsen, A. M. & Lagermann, L. C. (2017). *Stress i gymnasiet. Sammenfatning af resultater og anbefalinger*. Aarhus Universitet: DPU. <http://edu.au.dk/aktuelt/nyhed/artikel/stress-truer-elevers-laering-og-trivsel/> (lokaliseret 15. april 2017)
- Nielsen, H. & Thomsen, P. V. (1985; 86 og 88). Physics in upper secondary schools in Denmark. *European International journal of Science Education, 7*, 95-106; 8, 315-24; 10, 189-202.

- Nielsen, J. A. (2012). Science in Discussions: An analysis of the use of science content in socio-scientific discussions. *Science Education*, 96(3), 428–456.
- Nielsen, J. A. (2015a). *Evaluering af projektet Gymnasiet tænkt forfra 2012-2015*. Institut for Naturfagenes Didaktik, Københavns Universitet.
- Nielsen, J. A. (2015b). Assessment of Innovation Competency: A Thematic Analysis of Upper Secondary School Teachers' Talk. *Journal of Educational Research*, 108(4), 318-330.
- Nielsen, J. A. (2015c). *Rapport fra arbejdsgruppe for prøveformer der tester innovationskompetencer i gymnasiet*. Institut for Naturfagenes Didaktik, Københavns Universitet.
- Nielsen, J. A., & Dolin, J. (2016). Evaluering mellem mestring og præstation. *MONA*, 2016(1), 51-62.
- Nielsen, J. A., Waaddegaard, N., Dolin, J., & Bruun, J. (2017). Undervisning og læring i STEM. I: J. A. Nielsen (red.). *Litteraturstudium til arbejdet med en national naturvidenskabsstrategi* (s. 19-49). København: Institut for Naturfagenes Didaktik, Københavns Universitet.
- Nordenbo, S., Allerup, P., Andersen, H. L., Korp, H., & Dolin, J. (2009). *Pædagogisk brug af test – en systematisk review*. København: Danmarks pædagogiske universitetsforlag og Dansk Clearinghouse for Uddannelsesforskning.
- Roth, W.-M. (1995). *Authentic School Science. Knowing and Learning in Open-Inquiry Science Laboratories*. Cham, Switzerland: Springer.
- Rychen, D. S. & Salganik, L. H. (Eds.) (2003). *Key Competencies for a Successful Life and Well-Functioning Society*. Göttingen: Hogrefe and Huber Publishers.
- Sadler, T. D., & Zeidler, D. L. (2003). The morality of socioscientific issues: Construal and resolution of genetic engineering dilemmas. *Science Education*, 88(1), 4-27.
- Tidemand, S., & Nielsen, J. A. (2017). The role of socioscientific issues in biology teaching: from the perspective of teachers. *International Journal of Science Education*, 39(1), 44-61.
- Torrance, H. (2007). Assessment as learning? How the use of explicit learning objectives, assessment criteria and feedback in post-secondary education and training can come to dominate learning. *Assessment in Education: Principles, Policy & Practice*, 14(3), 281-294.
- Undervisningsministeriet (2004). *Fremtidens uddannelser. Den nye faglighed og dens forudsætninger*. København: Undervisningsministeriet.
<http://static.uvm.dk/Publikationer/2004/fremtidens/> (lokaliseret 21. april 2017).
- van den Akker (1998). The Science Curriculum: Between Ideals and Outcome. I: Fraser, B J. & Tobin, K. G. (eds.). *International Handbook of Science Education*. s. 421–448. Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Waddington, D., Nentwig, P. & Schanze, S. (Eds.) (2007). *Standards in Science Education*, Münster: Waxmann Verlag.