



Den naturfaglige evalueringskultur i folkeskolen

- Anden delrapport fra VAP-projektet



Jens Dolin
Lars Brian Krogh

Maj 2008

Den naturfaglige evalueringskultur i folkeskolen

- Anden delrapport fra VAP-projektet

Jens Dolin
Lars Brian Krogh

Maj 2008

IND-KU skriftserie nr. 2008-17

Trykte eksemplarer af rapporten kan, så længe lager haves, rekvireres hos:

Institut for Naturfagenes Didaktik

Københavns Universitet

Tlf.: 35 32 03 94 (sekretariat)

E-mail: ind@ind.ku.dk

www.ind.dk

Alle publikationer fra IND er elektronisk tilgængelige via instituttets hjemmeside.

Indholdsfortegnelse

1. Indledning.....	4
2. Kontekst og relevans	5
Evalueringskulturbegrebet.....	8
3. Viden om evaluering i naturfag i folkeskolens ældste klasser	10
Danske undersøgelser af evalueringspraksis i naturfagene	10
Evaluering i relation til naturfagene i et internationalt perspektiv	13
4. Empirisk undersøgelse af evalueringspraksis i folkeskolens naturfag i 8.-9. klasse.	18
Metode	18
Baggrundsdata om lærersamplet.....	20
Undersøgelsens spørgsmål og indholdsmæssige afgrænsning	24
Relation til dansk evalueringspraksis i naturfag i PISA's testkoncept.....	25
Indkredsning af danske læreres evalueringspraksis i naturfag	34
5. Konklusioner	47
6. Litteratur	50
7. English summary	52
Bilag 1 Validering af PISA Science	56
Bilag 2 Lærerspørgeskema om evaluering i naturfag	59

1. Indledning

Indeværende rapport er den anden fra forskningsprojektet *Validering Af PISA Science* (VAP). Det overordnede formål med VAP er at validere PISA Science 2006 i en dansk kontekst. VAP-projektet finansieres af Undervisningsministeriet, mens indeværende delundersøgelse desuden er støttet af Danmarks Lærerforening. Forskningsprojektet er i sin helhed gennemgået i bilag 1. Prodekan for uddannelse ved Det naturvidenskabelige Fakultet ved Københavns Universitet Henrik Busch var med til at opstille konceptet for denne undersøgelse, men har på grund af arbejdsbyrden som prodekan ikke kunnet medvirke ved dens gennemførelse.

Formålet med denne rapport er at etablere et grundlag for at kunne vurdere, hvorvidt PISAs testformat og tilgang til testning svarer til den evalueringskultur, der er fremherskende i naturfagene i de ældste klasser i den danske folkeskole.

Rapporten indledes med en kortfattet oversigt over nogle centrale udviklingstræk i evalueringsfeltet – op til og under indflydelse af PISA. Selve begrebet evalueringskultur gøres her til genstand for overvejelser.

Herefter følger en gennemgang af en række undersøgelser af evaluering i naturfagene i den danske folkeskole. Disse perspektiveres i forhold til udenlandske forskningsresultater.

Rapportens centrale afsnit udgøres af empiriske undersøgelser af, hvorledes danske naturfaglærere rent faktisk anvender evaluering i deres naturfagsundervisning i 8. og 9. klasse.

Der anvendes såvel interview som spørgeskema. I et metodeafsnit redegøres for de anvendte metoder.

Hovedresultaterne er baseret på en spørgeskemaundersøgelse gennemført blandt 1159 naturfaglæreres af deres brug af evaluering. Gennem såvel åbne som lukkede spørgsmål afdækkes lærernes begrundelser for at evaluere, deres foretrukne evalueringsformer, og deres vurdering af, hvorvidt danske skoleelever er forberedt til at tage PISA science-testen.

Rapporten slutter med en konklusion om de vigtigste resultater, hvor PISA science-testen sammenholdes med den danske evalueringskultur.

2. Kontekst og relevans

Inden for uddannelsessektoren har der gennem de seneste 10-20 år været en stadig stigende fokus på evaluering. Denne tendens skyldes en række faktorer, hvor den mest overordnede vel er den markedsorienterede globaliseringsdagsorden. Uddannelserne ses i stigende grad som en international konkurrenceparameter, hvorfor de skal kunne matche en international sammenligning, samtidig med at de på nationalt niveau underlægges en markedsgørelse.

Som en del af den offentlige sektor er denne proces knyttet til bestemte tidstypiske udviklingstendenser for den offentlige sektor som helhed. 1970'ernes forsøg på at begrænse udgifterne i den offentlige sektor gennem besparelser erstattes i 1980'erne af effektiviseringstiltag. I første omgang via rammestyring, som skal give incitament til institutionsintern effektivisering. I 1990'erne sker der et skift til ekstern effektivisering gennem tilstræbt øget konkurrence mellem institutioner – både mellem de offentlige indbyrdes og mellem offentlige og private. Gennem New Public Management-orienterede tiltag opbygges en markedslignende situation, hvor offentlige institutioner skal konkurrere med hinanden og med private udbydere. Dette kan kun ske, ved at den tidligere inputstyring erstattes af outputstyring. Tidligere tiders kvalitetsnormer for skoler (elever pr. klasse, bogstandarder etc.) afløses af mål for opnåede resultater (gennemførelsesprocent, karakterniveau etc.) – og disse resultater skal evalueres i forhold til nogle opstillede standarder. Øget vægt på outputstyring kræver nødvendigvis øget vægt på at kunne måle de opnåede resultater efter nogle fælles standarder.

Disse meget overordnede udviklingstræk kan ses overalt i den udviklede verden. Inden for naturfaglige uddannelser arbejdes der med at opstille standarder (Waddington, Nentwig et al.2007) og især med forskellige metoder til at evaluere disse standarder. Det er imidlertid et grundlæggende problem at opstille standarder for så komplekse størrelser som udbytte af et uddannelsesforløb, især når de ønskede udbytter udtrykkes i kompetencetermer. Det er ikke enkelt at operationalisere et sammensat kompetencebegreb på en form, så det kan evalueres enkelt, fx i en relativ billig test. Der anvendes derfor ofte såkaldte surrogatmål (fx en skoles karaktergennemsnit som mål for dens effektivitet), og selve målebestræbelserne medfører samtidig en tendens til at opstille simple mål for undervisningen. Der kan således ses en back-to basics-tendens, hvor man evaluerer enkle, fundamentale videnselementer, samtidig med at man i retorikken opstiller krav om komplekse kompetencer.

Dette misforhold mellem de gode intentioner, som de fx udtrykkes i overordnede målformuleringer og god undervisningspraksis, og de anvendte succeskriteri-

er for gode resultater såsom kvantitative målinger og komparative opgørelser kan forklares ved en kombination af samfundsmæssige forhold som nævnt ovenfor og mere evalueringsteoretiske forhold. Jens Johansen og Søren Langager samler forklaringerne i tre grupper (Johansen and Langager 2002):

1. Senmodernitetens komplekse og kontingente vilkår – risikosamfundet – opleves som en trussel mod nationale og kulturelle værdier. Der opstår derfor en *'retraditionaliseringstendens, som slår ned i den skolepolitiske offentlighed i form af ønsker om mere eksakthed i måling (og dermed sammenlignelighed) af elevernes præstationer inden for traditionelle færdighedsområder og præciserede kundskabs- og færdighedskrav på forskellige klassetrin i skolen'* (ibid s. 14).

2. Det, der evalueres - og dermed måles - er det, der ses, og dermed det, der opleves som vigtigt.

3. Der *er* foretaget meget udviklings- og forsøgsarbejde omkring nye evalueringsformer, især med henblik på at kunne evaluere mere bløde kompetencer, men også med henblik på at kunne følge udviklingen af faglige kompetencer. Men mange af de udviklede redskaber er omstændelige og ressourcekrævende, så – *alt andet lige – er det en forståelig sag, at metoder, som er enkle, overskuelige og kun lidt ressourcekrævende, i praksis får tendens til at dominere i en skolehverdag, som i forvejen er fyldt godt op med mange forskellige gøremål'* (ibid, s. 15).

PISA kan ses som eksempel på denne udvikling, idet PISA betjener sig af en kompetenceretorik, samtidig med at testformatet er relativt traditionelt. PISA science vil måle "scientific literacy", dvs. et ganske komplekst udtryk for at kunne bruge og forholde sig til naturvidenskabelig viden i hverdagsmæssige sammenhænge (se fx Dolin, Busch & Krogh 2006). Dette gøres ved at udsætte elever for en to timers skriftlig test – en alt andet end hverdagsagtig sammenhæng. De deltagende landes resultater ses som et fuldgyldigt udtryk for, hvad landets elever får ud af science-undervisningen, og ringe resultater anvendes som begrundelse for at stramme op om kravene til elevernes grundlæggende viden – uden at forholde sig til, hvorvidt en sådan opstramning vil fremme en udvikling af elevernes naturvidenskabelige kompetence og dannelse. Det gøres for mange uddannelsespolitikere til et mål i sig selv, at landet klarer sig godt i PISA-testen.

I den danske folkeskole kan denne proces ses gennem politikernes øgede ønske om at styre og ville se, hvad der rent faktisk sker i skolen, og hvilket udbytte eleverne får ud af undervisningen. I 1993 blev der formuleret de såkaldte centra-

le færdigheds- og kundskabsområder (CFK), hvor det faglige indhold var meget bredt beskrevet. I 2001 introduceredes de 'Klare Mål', som i 2003 erstattedes af landsdækkende og bindende Fælles Mål.

De uventede svage danske resultater i PISA 2000, især i science, bliver stærkt medvirkende til indførelse af test af de opstillede mål i 2., 4. og 6. klasse. Samtidig inviteres et internationalt ekspertpanel i OECD-regi til at vurdere stærke og svage sider ved den danske folkeskole. Deres rapport (Hargreaves et al 2004) peger bl.a. på en svag dansk 'evalueringskultur' og anbefaler:

"Etableringen af en ny evalueringskultur vil blive vanskelig. Imidlertid er det formentlig den vigtigste enkeltstående forandring, som må gennemføres for at andre tiltag kan implementeres effektivt og standarderne kan hæves." (s. 67).

Der gøres ikke rede for, hvad der forstås ved en evalueringskultur, men anbefalingerne følges op af regeringen, som i sit regeringsgrundlag af 17. februar 2005 skriver:

"Undersøgelser viser, at der trods lovkrav om evaluering i folkeskolen mangler en stærk evalueringskultur. Eleverne får derfor et utilstrækkeligt tilbagespil i forhold til deres præstationer, og forældrene har svært ved at vurdere børnenes fremskridt.

Regeringen vil styrke evaluering i folkeskolen ved at give skolerne pligt til at evaluere elevernes udbytte med udgangspunkt i de bindende mål for undervisningen, som blev iværksat i 2003."

Fra skoleåret 2005/06 blev der som følge heraf indført nationale, obligatoriske tests i læsning i 2., 4., 6., og 8. klasse, matematik i 3. og 6. klasse, engelsk i 7. klasse og naturfagene i 8. klasse. På grund af ringe kvalitet af de første testeksemplere blev indførelsen udskudt, så de første elever blev testet i maj 2007.

Det er i lyset heraf vigtigt at have kendskab til den evalueringskontekst, de nye nationale tests indføres i. Er de nye tests fremmede for eleverne? Og hvad med PISA-testene – er de fremmede for eleverne? Hvorledes evalueres i det hele taget i naturfag i folkeskolen, og hvordan er danske elever gearret til at klare PISA-science testen?

Med hensyn til den ønskede evalueringskultur, er det ligeledes relevant at spørge, om der i forvejen findes en evalueringskultur, som kan integrere de nye tests

i en meningsfuld kontekst. Men før det er det nødvendigt at afgrænse, hvad der kan forstås ved en evalueringskultur.

Evalueringskulturbegrebet

Det kan være på sin plads at undersøge, hvad der kan forstås ved et så centralt, og alligevel nyt, begreb som evalueringskultur. Peter Dahler-Larsen (2006) analyserer begrebet grundigt gennem en afklaring af hhv. kultur- og evalueringsbegrebet.

Kultur anvendes i en lang række forskellige betydninger:

- som skaber af noget indiskutabelt (social orden, normer etc.)
- som styrer af adfærd (bagvedliggende styringsmekanismer)
- som forklaring på det uforklarlige (fx overskride egen selvforståelse)
- og som skaber af fællesskab.

Det er således et meget komplekst begreb, men det er fælles for alle forståelser, at forsøg på at definere entydigt og beskrive udtømmende altid vil være utilstrækkelige og forsimplende. Anvendelsen af kulturbegreber i forandrings- og ledelsesøjemed er derfor vanskelig og ofte uhensigtsmæssig, idet den formaliserer noget, som langt hen ad vejen fungerer uformelt. Som Dahler-Larsen skriver:

”Enhver praktisk mobilisering af kulturen implicerer en fiksering af enkelte af kulturens aspekter og funktioner, hvorved tiltaget står i risiko for at blive ramt af tilbageslag fra de aspekter af kulturbegrebet, der overses.” (Dahler-Larsen 2006, s. 27).

Evaluerer Dahler-Larsen ud fra Vedungs klassiske definition som en *”systematisk retrospektiv (dvs. tilbageskuende) bedømmelse af gennemførelse, præstationer og udfald i offentlig virksomhed, og det er en bedømmelse, som tåntænkes at spille en rolle i praktiske handlingssituationer.”* Definitionen sammenkæder en vidensdimension, en værdidimension, en evaluand og et anvendelsesmål, men for Dahler-Larsen er den pragmatiske betydning, at evaluering fastsætter en sondring mellem det antastelige og det uantastelige. Evalueringer udfordrer en indsats, en institution eller en praksis ved at spørge, om praksis er god nok. Herved antastes selvfølgeheden i den evaluerede praksis, den praksis som altid kunne være anderledes, med henblik på at informere om og gøre praksis bedre:

”Idealet i evaluering er med andre ord ud fra en relativt uantastelig position at antaste en praksis for gennem informeret beslutning at gøre denne praksis et skridt mere uantastelig.” (Dahler-Larsen 2006, s. 31).

Evalueringer sigter således mod at udvikle og forbedre noget, og det er afgørende for deres legitimitet, at de gør det ud fra et relativt uantasteligt grundlag, specielt hvad angår den anvendte metode.

Begrebet evalueringsskulturer er altså sammensat af to termer med forskellig valør. Mens evalueringer skal være så eksplicite som muligt i deres begrundelser og procedurer, så er kulturer karakteriseret ved det implicite:

”Evaluering udføres med sigte på et formål (en eller anden form for anvendelse), hvorimod kulturer udmærker sig ved at holde noget helligt, som ikke lader sig reducere til noget instrumentelt.” (Dahler-Larsen 2006, s. 36).

Evaluering er baseret på viden, kultur på værdigrundlag; evaluering skal være forståelig for enhver med fornuft, kultur skal være meningsfuld for medlemmerne; evalueringer får legitimitet gennem alment gyldige principper, kultur har legitimitet i sig selv. Evalueringsskulturer kan derfor være udtryk for en række paradokser udtrykt ved sætninger som: Vi skal implicit acceptere det eksplicite, vi skal have en fast tro på evalueringens fornuft, vi skal i vores egen kultur indføre det samfundsmæssigt legitime etc.

En evalueringsskulturer kan således opfattes som en gruppe fælles, implicite holdning til, hvorfor og hvorledes medlemmer af gruppen evaluerer. Man kan tale om evalueringsskulturer på et nationalt niveau og på et lokalt niveau. En lokal evalueringsskulturer, fx på en skole, vil i større eller mindre grad afspejle den nationale evalueringsskulturer – i det omfang en sådan findes. Det vil vi forsøge at svare på i denne undersøgelse.

3. Viden om evaluering i naturfag i folkeskolens ældste klasser

Det foregående afsnit viser, at der om ikke andet er indført et testregime i den danske folkeskole – hvorvidt det bliver til en evalueringskultur vil tiden vise. Der er dog næppe tvivl om, at de mange tests vil få afgørende indflydelse på undervisningen i folkeskolen. Det er sådan set derfor de er indført, og 'Teaching-to the test' er en rimelig og ønskværdig udvikling – hvis testen er fornuftig.

De vejledende eksempler på tests i naturfag, der indtil nu har været offentliggjort, tyder på, at undervisningen vil blive drejet i retning af øget indøvelse af basale færdigheder. Disse er naturligvis nødvendige – især hvis man skal klare sig godt i en national test – men øget fokus på dem kan risikere at ske på bekostning af mere komplekse og kreative kompetencer.

På samme måde har PISA-testene påvirket uddannelsessystemet på en endnu ikke helt gennemskuelig måde. Men det kan være nyttigt at undersøge, hvilken viden der er om evaluering, specielt i relation til naturfagene, og hvilke konsekvenser evalueringer har haft for evaluanderne, for undervisningen, for uddannelsessystemet etc. Vi vil derfor i dette afsnit gennemgå en række danske og udenlandske undersøgelser af evaluering.

Danske undersøgelser af evalueringspraksis i naturfagene

Der er ikke mange undersøgelser af evalueringspraksis i naturfag i folkeskolens ældste klasser og endnu færre, som ser på evaluering i et mere teoretisk perspektiv. Vi vil derfor i dette afsnit gennemgå de seneste danske undersøgelser for i næste afsnit med afsæt i international forskning at give nogle eksempler på anvendelse og konsekvenser af evalueringer i relation til naturfagene.

Med Niels Egelund som projektleder blev der i starten af 2000-tallet gennemført en række undersøgelser for at belyse den svigtende interesse for naturvidenskab og teknik blandt unge. To af de udgivne rapporter belyser evalueringspraksis i naturfag i folkeskolens ældste klasser. Den ene rapport (Broch og Egelund 2002) baseres på interview af 17 lærere, hvoraf de ni underviste i fysik/kemi. Om disse ni læreres brug af evalueringer i undervisningen sammenfattes:

Der er stor forskel på lærernes brug af evalueringer i undervisningen, herunder hvor meget de gør ud af at evaluere undervisningen samt hvilke metoder de gør brug af. Nogle lærere benytter flere forskellige metoder, mens andre udelukkende bruger de tests, der er i bogen. Kun enkelte af lærerne i undersøgelsen gør systematisk brug af evalueringer som en del af undervisningen, og desuden fore-

kommer det os på baggrund af analysen, at en del af lærerne ikke tillægger det særlig stor vægt. (s. 161).

Den anden publikation (Egelund 2002) rapporterer om en spørgeskemaundersøgelse omfattende 611 fysik/kemilærere. I undersøgelsen har man stillet spørgsmålet ”hvor ofte evaluerer du din fysik/kemi-undervisning?”, men ulykkeligvis er de bundne svarmuligheder en sammenblanding af ”hvor ofte” og ”hvornår” der evalueres. Da resultaterne yderligere anføres uden signifikansangivelse, har de samlet set en begrænset anvendelighed. Der er dog en indikation af, at når det gælder hyppigheden af undervisningsevaluering, er den nok nærmere ”sjældent” end ”flere gange i et undervisningsforløb”. Endvidere antyder resultaterne, at ydre ”afregning” er den væsentligste drivkraft i denne evalueringssindsats, idet den primært foregår ”ved afslutningen af skoleåret” og ”i forbindelse med forældresamtaler”. Den internt begrundede evaluering ”ved afslutningen af et undervisningsforløb” har lavest tilslutning blandt de svarmuligheder, som vedrører den tidlige placering.

Et andet spørgsmål i samme undersøgelse kortlægger, hvordan lærerne ”evaluerer undervisningen/elevernes præstationer.” Heller ikke her er spørgsmålets formulering ideel, idet det sammenblander evaluering af undervisning og elevpræstationer – to perspektiver, som meget vel tilgodeses med forskellige metoder.

Lærerne skulle på en firetrinsskala angive, hvor hyppigt de anvendte 6 angivne metoder. Resultaterne var som følger (s. 68), idet ’altid’ tildes værdien 1, ’ofte’ værdien 2, ’sjældent’ værdien 3 og ’aldrig’ værdien 4:

Ved skriftlige opgaver til eleverne	2,24
Ved klassesamtaler	2,38
Udelukkende på baggrund af egne refleksioner	2,50
Ved samtaler med den enkelte elev	2,64
Med udgangspunkt i elevrapporter	2,88
Andre måder	3,04
Med udgangspunkt i elevernes noter eller logbøger	3,18

Lærerne evaluerer altså ofte ved hjælp af skriftlige opgaver og sjældent med udgangspunkt i elevrapporter eller logbøger, dvs. mere procesorienterede former.

Lærernes spørges også om, hvorledes evalueringerne anvendes. Her er det hyppigst forekommende at anvende evaluering til at planlægge det videre undervis-

ning i klassen (1,87), mens det er relativt sjældnere at bruge evalueringerne systematisk i planlægning af undervisningen i fysik/kemi (2,33).

Endelig spørges til lærersamarbejdet. I 12 % af tilfældene er fysik/kemilæreren alene, og det almindeligste er, at der er 3 fysik/kemilærere på en skole. Lærerne vurderer samarbejdet som helt overvejende tilfredsstillende, selv om det er relativt sjældent forekommende og af relativ informel karakter (s. 82). Kun 30 % af lærerne oplyser, at de arbejder med planlægning og/eller evaluering af fysik/kemi-undervisningen i et fagteam med andre fysik/kemi-lærere.

Danmarks Evalueringsinstitut rapporterede i 2004 undersøgelsen "*Løbende evaluering af elevernes udbytte af undervisningen i folkeskolen*" (EVA, 2004). Som titlen antyder, fokuseres der på et snævert område af evaluering – og resultaterne fremlægges kun undtagelsesvis med reference til bestemte fag/fagområder. Alligevel er det både interessant og markant, at lærerne generelt bruger "test/prøver" lige så hyppigt som "lærerelev-samtaler" som grundlag for løbende evaluering (p. 34). Det er også et relevant træk, at procesorienterede evalueringsredskaber (fx logbog, portfolio) anvendes i langt mindre udstrækning, at disse fortrinsvis bruges af kvinder og hyppigst i dansk og anden humanistisk undervisning. Med en mulig relevans for naturfag hedder det:

"Flere matematiklærere på skoler, der generelt satser på portfolio, vurderer, at de forholdsvis let ville kunne arbejde med redskabet i forhold til matematikundervisningen – og de har typisk vanskeligt ved at forklare, hvorfor de ikke allerede gør det." (p. 37).

I denne sammenhæng har vi yderligere fundet det værd at fremdrage et forskningsmæssigt studie, som ikke fokuserede på kortlægning af evalueringspraksis – men søgte at "*inspirere til udvikling af lærerens tænkning i forbindelse med didaktiske vurderinger som en del af en generel, løbende intern evaluering.*" (Weng and Hoff 1999) I 1995 deltog Danmark i den internationale undersøgelse *Performance Assessment in IEA's Third International Mathematics and Science Study* (Harmon and et al 1997), et særligt appendiks til de sædvanlige TIMMS-studier.

Performance Assessment-delen havde som underliggende rationale, at meget væsentlige dele af elevers naturfaglige (og matematiske) kompetencer dårligt indfanges af teoretiske opgaver i et papir-og-blyant test-setup. Alternativet var her at efterprøve elevernes formåen ved at lade dem udføre små praktisk orienterede problemløsningsaktiviteter og evaluere dem på deres nedskrevne overvejelser og resultater i tilknytning hertil. Svararkene indeholdt fortrykte struktureren-

de spørgsmål, men selve svarformatet var åbent. De danske resultater fra denne *performance assessment* er rapporteret i (Weng and Hoff, 1999), da det danske sample ikke fuldt levede op til kriterierne for den internationale rapport.

Studiet var designet, så eleverne gennemførte både det ordinære teoretiske TIMMS-opgavesæt og Performance Assessment-aktiviteterne. Ved sammenkørsel af resultaterne viste det sig meget væsentligt, at der i det danske sample kun var en beskedent korrelation mellem elevernes resultater i de to testsituationer (varierende i intervallet 0.16-0.58, afhængigt af aktivitet). Dette synes umiddelbart at underbygge rationalets forestillinger om, at de forskellige test-setupps giver udfoldelsesmuligheder for forskellige elevkompetencer. Forfatterne fremhæver den nuancering af elevkompetencerne, som opnås – og understreger i en naturlig konsekvens heraf ”*nødvendigheden af at anvende flere typer af opgaver, når kompetencer skal vurderes.*” (p. 232).

Et andet væsentligt resultat var, at piger og drenge præsterede lige godt i den praktiske testning, både nationalt og internationalt. Dette i markant modsætning til TIMMS-hovedundersøgelsen, hvor pigerne klarede sig signifikant dårligst. Det er tankevækkende, at de meget omtalte kønsforskelle i elevpræstationer i naturfag kan fremkaldes/udviskes med et (u)passende valg af evalueringsmetode.

Det anderledes testformat åbner op for og fremhæver ellers oversete kompetencer. Det gør det imidlertid også synligt, når sådanne kompetencer – trods det nye udfoldelsesrum - *ikke* er til stede. F.eks. peger undersøgelsen på, at mange elever har svært ved at planlægge forsøg. Det fremgår også, at deres evne til at reflektere over og formidle resultater er såre beskedent, samt at der mangler argumentation og ræsonnement i deres skriftlige begrundelser. Tillige påpeges det, at eleverne generelt anvender meget få faglige begreber i deres skrevne forklaringer til forsøgene. Det er fristende at tolke det relative fravær af alle disse centrale, faglige delkompetencer som et udtryk for, at disse i for ringe grad vægtes i den daglige undervisning og i lærernes almindelige evalueringspraksis.

Evaluering i relation til naturfagene i et internationalt perspektiv

En række nyere meta-blik på evaluering i relation til naturfag (fx Lawrenz, 2007; Millar, 2005) refererer tilbage til Jenkins' (1976) forståelse af undervisningsmæssig evaluering: ”*Educational evaluation is the process of delineating, obtaining and providing useful information for judging decision alternatives*”. Ophavsmændene til disse metablik fokuserer primært på ”program”-evaluering

(i en bred¹ forståelse heraf), men definitionen er anvendelig også for elevevaluering. Det afgørende i denne sammenhæng er, at definitionen taler om at tilvejebringe nyttig information, men undlader at specificere, hvem der udstikker rammerne for ”nyttig information”, hvordan den relevante viden konstrueres, hvem der beslutter noget på grundlag af informationen og hvorvidt evalueringen er en integreret del af en undervisningsmæssig proces. Langt hen ad vejen vil denne ene definition kunne udspænde Stake’s (1986, loc.cit. i Millar p. 17) otte dimensioner til karakterisering af evalueringstiltag:

- formative / summative
- formal / informal
- case-particular / generalisable
- outcome / process
- judgmental / descriptive
- preordinate / responsive
- holistic / analytic
- Internal / external.

PISA-setuppet fastlægger og adskiller aktørrollerne på en temmelig håndfast vis, størst uklarhed er der omkring, hvem der afgrænser, hvad der er relevant viden. Vekselspillet mellem de politiske opdragsgivere og PISA’s rådgivende Scientific Board er ikke helt transparent. Efter Stake’s karakteriseringssystem må PISA betegnes som summativ, formel, tvivlsomt generaliserbar(?), resultatorienteret, bedømmende, forud fastlagt, holistisk (idet effekten af hele undervisningssystemet og uformel læring vurderes under ét) og eksternt fastlagt evaluering. Disse træk deler PISA selvfølgelig med størstedelen af alle andre evalueringstiltag på stor skala (”Large-Scale-Assessment”, LSA), om end der er eksempler på, at data tilbageføres til respondenterne på en måde, så der også opnås et vist formativt islæt. Interessant er i øvrigt Millar’s redegørelse for, hvorledes man allerede i 1970’erne drabeligt kritiserede evaluering som lod al proces-information ude og udelukkende anskuede klasserummet som en black-box. (Millar, p. 22). Nutidens PISA-koncept kan ses som en eksponent for netop denne kritiserede tendens.

Enhver karakterisering af et konkret evalueringstiltag efter fx Stake’s dimensioner skal selvfølgelig ses i relation til det aktuelle formål med evalueringen. I forlængelse af den anførte Jenkins-definition er evaluering tydeligvis forbundet med nytteværdi og dermed et instrumentalistisk sigte. Evalueringen skal kunne anvendes til noget! Der er givet forskellige bud på kategorisering af dette ”no-

¹ Hos Millar dækker dette over systematisk evaluering af alt fra nationale initiativer (fx læreplaner) over lokale (fx universitets-)kurser til forskningsinformerede undervisningssekvenser (såkaldte ”Teaching sequences”).

get” (Black, 1998 p.35; Dahler-Larsen, 2001 p.215). Black fremhæver således tre evalueringstyper knyttet til hovedformålene:

- Den løbende, formative med læring som formål
- Den summative, med elevsortering, overførsel og certification som formål
- Den summative, med legitimering og afregning i.f.t. offentligheden som formål.

Der er ingen tvivl i litteraturen om, at der generelt evalueres mere end nogensinde – hvilket også gælder for evalueringer af PISA-typen. Fx opsummerer Britton et al (Britton and Schneider 2007) ”*the frequency and types of international comparisons of student understanding of science... are increasing.*”

Hvad angår typerne af evaluering er der modstridende tendenser, bl.a. trives internationale, eksterne evalueringer fint side om side med en opprioritering af deltager-involverede tilgange (”participatory approaches”, se fx Lawrenz, 2007). Der er tegn på en udvikling i anvendelsen af forskellige typer af evaluering, idet Gitomer & Duschl (1998) anfører: ”*Historically in science education, assessment of conceptual components have been separated from process, practical, inquiry and attitudinal components.*” PISA 2006 er åbenlyst et brud på denne adskillelse af begrebsmæssige og holdningsmæssige komponenter, idet undersøgelsen bevidst søger at indfange elevernes kontekstuelle interesse for emner, som de netop har beskæftiget sig opgavemæssigt med..

Britton & Schneider problematiserer den politiske anvendelse af stor-skala-evalueringer, bl.a. ”*attaching increasingly high stakes to assessments can increase the amount of science instruction occurring in classrooms and provide some policy guidance, but the policy implications drawn and resulting science instruction can be at odds with research on some aspect of how students learn.*” (p. 1008). Derudover påpeger de, at evaluering som ikke er afstemt efter standarder for naturfagsundervisning kun vanskeligt kan tjene som rettesnore for udvikling af lærernes undervisning. De understreger også, at evalueringsformer baseret på multiple choice- og korte, åben respons-opgaver kun tillader måling af et meget snævert spænd af mål og kompetencer.

I forvejen er der meget, der antyder, at evalueringsformer og undervisningsmål er usynkrone, fx opsummerer Orpwood (2001): ”*the changes in the goals of science education in the 1960s towards a greater emphasis on inquiry skills were matched some 20 years later with a change in assessment to include performance assessment. Now the new goals of science education are focused on the need to link science to the broader social context, but assessment practices have yet to catch up with this change....the new curriculum emphasis may well be ig-*

nored unless new approaches to assessment are designed and implemented soon.” I Britton og Schneiders diskussion af mulige fordele og ulemper ser man også modstridende konsekvenser af udstrakt testning, idet dette positivt kan bidrage til at fokusere undervisningsindsatsen, men samtidig meget nemt udvikler sig til *”teaching for the test”* (p. 1010). Der er en tillige en række eksempler på, at elevscorer umiddelbart forbedres i kølvandet på testning, men også at dette kun er et overflade-fænomen:

”Test score increases appear to be inflated. If teachers teach only to the test and not to the larger domain that the test is intended to represent, the test score results may represent just that – higher scores on a specific test and not genuine learning that generalizes to other domains of science instruction” (p. 1010).

Sammenhængen mellem testning og elevpræstation i naturfag er kompliceret. Der foreligger således temmelig modstridende empiriske og teoretiske beskrivelser af effekten af *”high stakes-assessment”*, hvor et dårligt evalueringresultat kan få store konsekvenser for den evaluerede. I én variant fremsættes den hypotese, at udsigten til prøver samt stressende omstændigheder ved selve test-situationen, kan lægge et produktivt pres på eleven.

Nu er PISA ikke en high stake-test, idet resultaterne ikke har konsekvenser for de deltagende skoler og elever. I den danske PISA 2003 rapport kan man imidlertid se, at danske elever føler, at PISA-setuppet allerede i den foreliggende version får dem til at anstrenge sig betydeligt (om end yderligere intensivering af vilkårene angiveligt ville kunne give den det sidste nøk). Imidlertid skriver Nichols et al (Nichols, Glass & Berliner 2005) i en opsummering til den amerikanske regering: *”The authors conclude that there is no convincing evidence that the pressure associated with high-stakes testing leads to any important benefits for students’ achievement. They call for a moratorium on policies that force the public education system to rely on high-stakes testing.”*

En modsat rettet teori forudsiger, at eksamenslignende testning vil fremkalde *”test-anxiety”*, hvilket vil få eleverne til at underpræstere. Evidensen for denne model fremgår bl.a. af et omfattende meta-studium (EPPI, 2002), hvor det bl.a. fremgår, at efter indførelse af Nationale tests i England har *”svage”* elever - i modsætning til tidligere - et lavere selvværd (self-esteem) end andre. Mere generelt konkluderes der:

”Most of the evidence presented in the review indicates negative impact of summative assessment on the learning of some students. A range of reasons are cited, including:

- *Lowering the self-esteem of less successful students, which reduces their effort;*
- *Test anxiety, which affects students differentially;*
- *Restricting learning opportunities by teaching which is focused on what is tested, and by teaching methods which favour particular approaches to learning.”*

Disse træk modsvarer meget godt en konklusion fra Nichols et al (2005):

“High-stakes testing pressure is negatively associated with the likelihood that eighth and tenth graders will move into 12th grade. Study results suggest that increases in testing pressure are related to larger numbers of students being held back or dropping out of school.”

Den indlysende pointe må være, at man skal have gode grunde til at gennemføre evalueringer med “høj indsats”. Det koster i engagement, i særdeleshed for de mindst succesrige elever. Overraskende nok finder man ikke en righoldig international litteratur om, hvad der kendetegner naturfagslæreres evalueringspraksis og hvilke holdninger, der ligger til grund for denne. Som det formuleres i en nyere artikel (Rodriguez, 2004):

”Much of the literature regarding classroom assessment exists in the form of professional development-related articles and books.”

Dvs. at der er i vid udstrækning tale om udviklingsorienteret viden, som i bedste fald er opsamlet efter best practice-kriterier. Nogle få ældre forskningsmæssige studier (fx (Stiggins and Bridgeford, 1985)) viser, at amerikanske lærere i grundskolens matematik og naturfagsundervisning i vid udstrækning foretrækker selvgjorte ”objective” tests, typisk traditionelle småopgaver og langt fra procesorienteret Performance Assessment. Det er påfaldende, at denne monochrome evalueringsform dominerer, uanset om evalueringen efter lærernes mening skal bruges til diagnosticering, elevgruppering, karaktergivning eller vidererapportering. Undersøgelsen viser lidet overraskende, at intensiteten af sådan testing tager til fra 2. til 11. klasse. I princippet foreligger der en nyere fond af viden om læreres evaluering i ”Teacher Background Files” til TIMMS 2003. Her findes oplysninger om frekvens og indhold af forskellige typer evaluering. Materialet er så vidt vides aldrig blevet publiceret, men dele af datafilerne er tilgængelige på TIMMS-hjemmesiden. Desværre er samplet på 1090 mangelfuldt deklareret, fx kan nationalitetssammensætningen ikke aflæses.

4. Empirisk undersøgelse af evalueringspraksis i folkeskolens naturfag i 8.-9. klasse

Metode

For at få et dækkende billede af evalueringskulturen i naturfagene i de ældste klasser valgte vi at gennemføre en spørgeskemaundersøgelse blandt et repræsentativt udsnit af danske naturfaglærere for 8.-9. klasse.

Fastlæggelse af spørgeskemaets indhold

De relevante spørgsmål blev indkredset gennem to fokusgruppeinterviews.

Vi fik kontakt til naturfaglærerne på to forstadsskoler, en i København og en i Odense. Lærerne fik før interviewet tilsendt nogle spørgsmål om deres holdninger til evaluering og deres anvendelse af evaluering.

Interviewene fandt sted i april 2005. Interviewene gav nogle meget fyldige og reflekterede overvejelser om brug af evalueringer hos to tilfældigt udvalgte lærergrupper (begge repræsenterende fysik/kemi, biologi og geografi). Disse gjorde det muligt at opstille relevante spørgsmål til feltet.

På baggrund af interviewene blev der udarbejdet et spørgeskema. Dette blev som en første pilot sendt til de interviewede, og på baggrund af deres kommentarer blev der udformet et endeligt skema i Inquisite. Spørgeskemaet er vist i Bilag 2.

Sample og dataindsamling

Vi har været nødt til at gennemføre dataindsamlingen to gange, idet første indsamlingsrunde ikke gav et tilstrækkeligt datagrundlag. Når vi alligevel omtaler begge, er det fordi de illustrerer vanskelighederne ved en elektronisk spørgeskemaundersøgelse af danske skolelærere.

I den første dataindsamling, som foregik i april/maj 2005, udvalgte vi 44 skoler efter de normale procedurer for repræsentativ udvælgelse af skoler. De udvalgte skoler blev kontaktet telefonisk og orienteret om undersøgelsen samt bedt om at oplyse navne og email-adresser på naturfaglærere på skolen, som i skoleåret 2004/2005 underviste i et naturfag i 8. eller 9. klasse. Ved denne første kontakt viste det sig, at 14 skoler ikke opfyldte betingelsen, at der blev undervist i naturfagene i 8. og 9. klasse; disse skoler blev erstattet af matchende skoler, dvs. skoler med samme nummer på en sekundær liste over 44 skoler.

Skolekontakten resulterede i navn og e-mail på 131 lærere repræsenterende 37 skoler. Til disse lærere udsendte vi en mail med en orientering om undersøgelsens formål samt et link til det elektroniske spørgeskema. Det afsløredes herved at kvaliteten af de tilbagemeldte email-oplysninger var noget tvivlsom, idet 46 mails automatisk blev returneret. Af de 85 invitationer, som vi kan formode er nået frem til en respondent, er der 54 som har besvaret spørgeskemaet. Svarprocenten blandt de der faktisk har fået muligheden for at svare er således 63,5 %. Erfaringen fra kontaktfasen er, at det kan være meget vanskeligt at få kontakt med skolen. Det er især meget vanskeligt at få kontakt med skoleledelsen. Mail benyttes tilsyneladende på meget forskellig måde på forskellige skoler. På nogle skoler er der aldrig kommet svar på mail til skolen, selv om der er sendt flere rykkere. På tilsvarende måde er det ikke uden problemer at regne med at kunne benytte sig af lærernes mailadresser. Nogle opgivne mailadresser er tilsyneladende "døde" mens andre er forkerte. På nogle skoler er det naturligt, at kontoret kender lærernes mailadresser, mens det på andre er ganske ukendt, om lærerne overhovedet har en email-adresse.

Det web-baserede spørgeskema, et HTML-skema, som returneres til en server, har ikke givet anledning til tekniske problemer. En enkelt respondent angav, at besvarelsen ikke kunne sendes direkte, men vedlagde i stedet besvarelsen i en mail. Problemet er kendt og skyldes ofte, at respondenter benytter en maskine, hvor der er skruet meget højt op for it-sikkerheden.

Problemet med datamaterialet var imidlertid, at der var for få respondenter til, at vi kunne foretage en tilstrækkelig detaljeret statistisk bearbejdning – mange af svarcellerne havde simpelthen for få gyldige svar. Vi måtte derfor anvende en anden procedure til dataindsamling.

Ved den anden dataindsamling, som forløb i perioden januar-februar 2006, tog vi udgangspunkt i Undervisningsministeriets liste over uddannelsesinstitutioner. Vi udvalgte grundskolerne, 1998 i alt, som efter en nærmere granskning blev reduceret til 1989 skoler.

Alle disse skoler fik tilsendt en mail med opfordring til at give naturfagslærere, som i skoleåret 2005/2006 underviste i fysik/kemi, biologi og/eller geografi i 8. eller 9. klasse, en invitation til det web-baserede spørgeskema.

Den 1. marts 2006 var der modtaget 1159 besvarelser fra 667 skoler ud af de 1235 skoler som vi kunne komme i forbindelse med (119 mails kom retur), og som vi kunne forvente svar fra (415 skoler angiver fx, at de er 7-klassede og derfor udenfor segmentet). Med den valgte indsamlingsprocedure er det er ikke mu-

ligt at angive en svarprocent, ligesom man ikke kan tage for givet, at det faktiske sample er repræsentativt. Men datamaterialet tyder dog på det.

Baggrundsdata om lærersamplet

I en undersøgelse af denne type er det et centralt spørgsmål, om de adspurgte udgør et repræsentativt udsnit – så man med rimelighed kan generalisere resultaterne til en større population. Nedenfor vil vi karakterisere samplet ud fra de væsentligste baggrundsparemetre (køn, alder, linjefagsbaggrund og undervisningserfaring udtrykt i antal år) og søge at vurdere repræsentativiteten. Det sidste er vanskeliggjort af, at der ikke foreligger opgørelser over den samlede bestand af lærere, som i 2005/2006 underviste i mindst et af naturfagene i 8. eller 9. klasse. Det er derfor svært at finde et helt velegnet sammenligningsgrundlag.

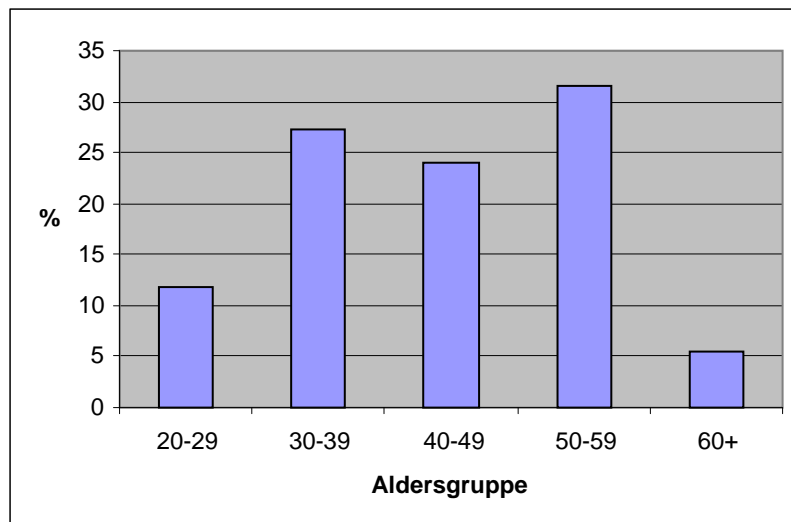
Køn:

Samplet (N=1159) indeholder ca. 65 % mandlige og 35 % kvindelige naturfagsundervisere. Skævheden i kønsfordelingen af naturfagslærerkorpset er ikke overraskende, da piger/kvinder på samtlige niveauer af det danske uddannelses-system er mere negative overfor naturfag eller underrepræsenterede i dem. Til sammenligning består grundskolens *samlede* lærerkorps af dobbelt så mange kvinder som mænd (på tværs af samtlige fag)

(kilde: <http://www.uvm.dk/statistik/grundskole/elevlarer0304-0506.htm?menuid=551005>).

Man kan få *et vist indtryk* af samplets repræsentativitet ved at sammenligne med Egelunds tal for ”fysik/kemilærerne” (Egelund, 2002). Desværre fremgår det ikke, om prædiketet dækker over, at man har en sådan linjefagsbaggrund – eller ”blot” (som her), at man ”*normalt underviser i*” faget. I vores sample er der 2/3 mænd blandt de lærere, som har fysik/kemi-linjefagsbaggrund. Egelund anfører tilsvarende, at der er 78 % mænd blandt fysik/kemilærerne. Den mindre uoverensstemmelse er ikke bekymrende.

Aldersfordeling (aktuelt sample, alle fag):

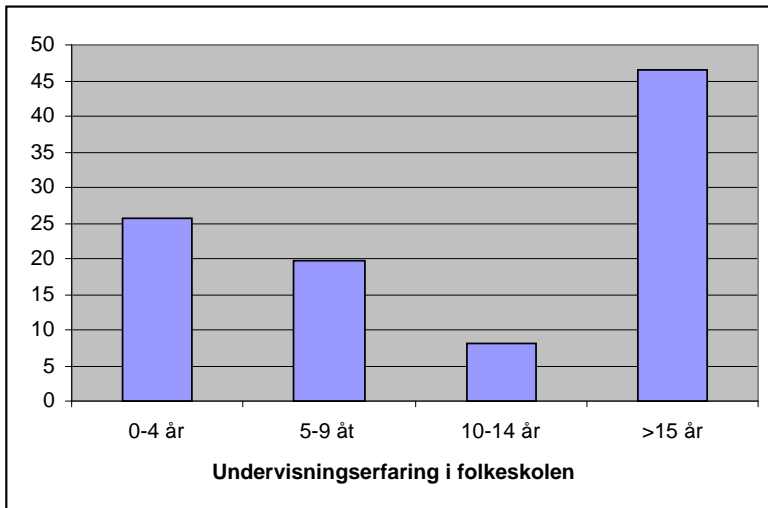


Figur 1: Aldersfordeling i samplet

Figur 1 illustrerer folkeskolens bemandingsproblem, her konkret hvad angår naturfagsundervisningen i de større klasser. Med uændret tilstrømning og fordeling vil godt 30 % af underviserne være gået på pension indenfor de næste 10 år – og kun ca. halvdelen vil blive erstattet af nye lærere. Aldersfordelingen er rimeligt repræsentativ for det samlede danske lærerkorps (jf. ovenstående link til uvm-statistik, hvor en direkte sammenligning kompliceres af, at aldersgrupperne er forskudt med 5 år i.f.t. søjlediagrammet her).

Undervisningserfaring

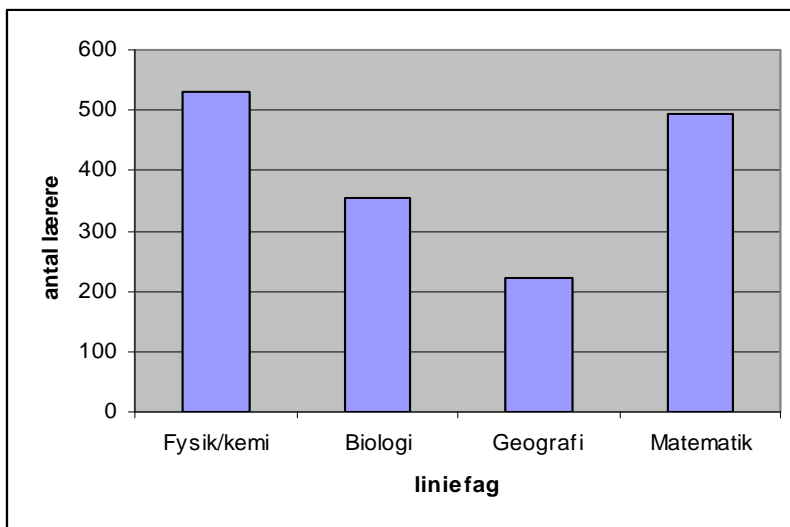
Ud fra aldersfordelingen er det ikke overraskende, at op imod halvdelen af lærerne er endog ganske rutinerede, dvs. tilhører kategorien med 15+ års erfaring inden for folkeskoleundervisning. Mere overraskende er det måske, at der kun er ca. 12 % lærere i yngste kategori, men 25 % med meget begrænset undervisningserfaring. Ganske mange bliver øjensynligt naturfagslærere på et relativt sent tidspunkt i deres voksenliv.



Figur 2: Fordelingen af undervisningserfaring i folkeskolen

Liniefagsbaggrund og -dækning

Det fremgår, at der er flest lærere med fysik/kemi og matematik som baggrund for deres undervisning.



Figur 3: Lærernes naturfaglige linjefagsbaggrund

Analysen af baggrundsdataene godtgør, at naturfagsundervisningen på de ældre klassetrin i høj grad varetages af lærere med en eller anden naturfaglig linjefagsuddannelse. Kun 12 % har ikke en sådan baggrund. Ca. 43 % af lærerne i vores sample har kun et naturfagligt linjefag, knapt 40 % af lærerne har to sådanne linjefag, godt 5 % har tre, mens endelig ca. 0,3 % har alle fire kortlagte linjefag. Fysik/kemi-linjefag i kombination med matematik er den hyppigste linjefagsbaggrund (ca. 22 %).

Alligevel underviser naturfagslærerne hyppigt i et naturfag, hvor de *ikke* har linjefagsbaggrund.

I spørgeskemaet blev lærerne bedt om at angive, hvilke fag de "normalt" underviser i. Resultaterne viser, at fx kun 70 % af de "normale" undervisere i fysik/kemi har den relevante linjefagsbaggrund (Egelunds tilsvarende tal er 60 %, p. 21). Ifølge vores sample undervises der i matematik med en linjefagsdækning på ca. 57 %, i biologi ca. 52 % og i geografi kun med ca. 40 %'s dækning. Netop hvad angår disse tal findes der et rimelig godt sammenligningsgrundlag i en opgørelse fra Uni-C (KL 2006, p.9), som tidsmæssigt og m.h.t. afgrænsning er et nært match til vores. Opgørelsens tal for dækningen med linjefagsbaggrund er: matematik: 57 %, biologi: 57 %, geografi: 40 % og endelig fysik/kemi: 79 %. *Den gode overensstemmelse her er nok vores bedste argument for, at samplet og resultaterne må tilskrives en vis repræsentativitet.*

Lærerne har yderligere forholdt sig til, hvilke(t) fag de "identificerer sig mest med". Godt 6 % identificerer sig øjensynligt ikke med nogen af de anførte fag, mens 7-10 % har en klar identitetsmæssig præference for et enkelt af fagene. Der er ret indlysende en vis korrelation mellem identifikation og linjefagsbaggrund, men alligevel også interessante træk. Man kunne godt have en forventning om, at lærere med linjefagsbaggrund i ét enkelt naturfag vil opbygge en identitet og en entydig identifikation med dette. For lærere med geografi- eller biologibaggrund gør dette sig gældende i ca. halvdelen af tilfældene (53 %, hhv. 48 %). De tilsvarende tal for lærere med matematik- eller fysik/kemibaggrund forekommer overraskende lave: 33 % hhv. 28 %.

Ved nærmere undersøgelse viser det sig, at de lave tal først og fremmest hidrører fra lærere med matematikbaggrund, som gennem professionsudøvelsen *også* udvikler en identifikation med fysik/kemifaget – og omvendt. Den tendens til sammentænkning af disse to fag, som allerede kom til udtryk i lærernes hyppige valg af en linjefagsbaggrund inden for *både* fysik/kemi *og* matematik synes således at blive forstærket af den daglige undervisning og strukturerne på skolerne.

Undersøgelsens spørgsmål og indholdsmæssige afgrænsning

Undersøgelsen er overvejende deskriptiv og har haft et dobbelt sigte, nemlig at besvare specifikke PISA-rettede spørgsmål og samtidig lave en bredere indkredsning af evalueringspraksis og ansatser til evalueringskultur i naturfagene i folkeskolens større klasser.

Til hvert af disse sigter hører der forskellige forskningsspørgsmål og indholdsmæssige fokuseringer.

a) Operationaliseringen af det PISA-rettede sigte.

Følgende forskningsspørgsmål sammenfatter denne del af undersøgelsen: *Er dansk evalueringspraksis i naturfag i rimelig overensstemmelse med PISA's test-koncept?*

Indholdsmæssigt søger denne del af undersøgelsen at afdække, i hvilken udstrækning lærerne anvender PISA-lignende evalueringsformater (individuel, paper-and-pencil, vægt på MC-opgaver), om de samme typer af kompetencer evalueres, om kontekster inddrages på sammenlignelig måde, hvilken betydning anvendelse af korrekt fagsprog tillægges i den daglige evaluering, o.s.v.. Disse indholdskomponenter er i spørgeskemaet overvejende formuleret som lukkede, Likert skala-spørgsmål.

En underliggende hypotese knyttet til det PISA-rettede forskningsspørgsmål har været, at danske elever muligvis underpræsterer i PISA, fordi de ikke er fortrolige med en sådan type testning. Fx er der først nu indført afsluttende prøver i folkeskolens biologi- og geografiundervisning, mens fysik/kemi traditionelt har arbejdet hen imod en afsluttende prøve, som var praktisk/performance-orienteret. Der er altså grund til at forvente et problematisk mismatch.

b) Indkredsning af den bredere evalueringspraksis i naturfag.

Vi har her udvalgt et antal grundlæggende spørgsmål, som vanskeligt lader sig sammenfatte under ét:

- *Hvilke begrundelser har lærerne for deres evaluering?*
- *Hvad karakteriserer i en bredere forstand naturfagenes evalueringspraksis?*

- *Hvilke mønstre viser der sig i evalueringspraksis (fagspecifikke træk, den typiske lærers evalueringsrepertoire)*
- *Er der indikationer af, at den enkelte lærers evalueringspraksis er forankret i en lokal skolebaseret evalueringskultur?*

De første delspørgsmål adresseres via specifikke items (åbne og/eller lukkede) i spørgeskemaet, andre kan først besvares efter en mere sammensat analyse.

Vores gennemgang af undersøgelsens resultater vil følge ovenstående struktur.

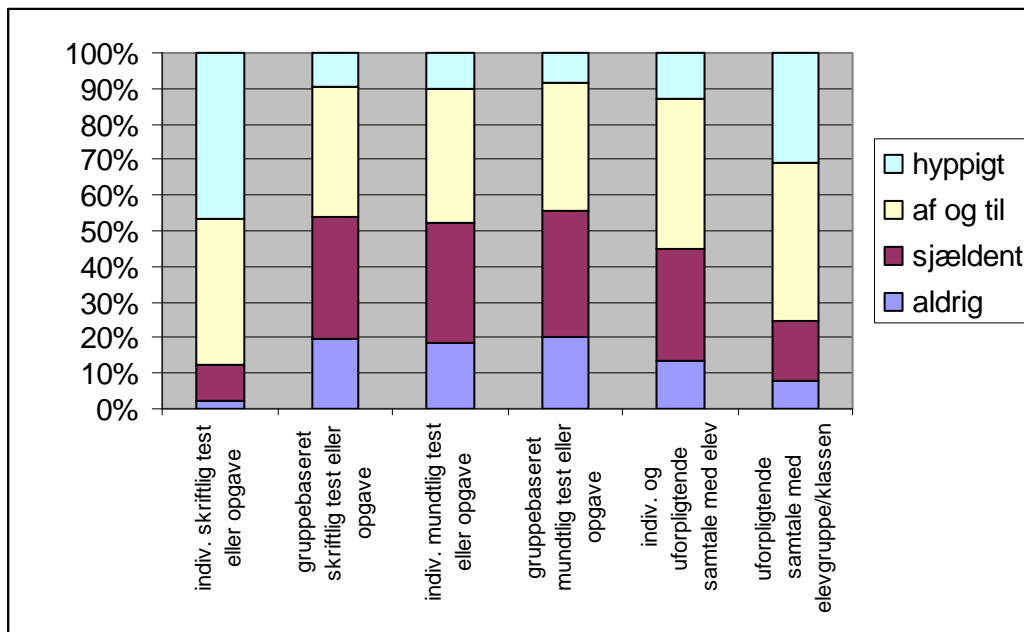
Relation til dansk evalueringspraksis i naturfag i PISA's testkoncept

Anvendelsen af PISA-relevante evalueringsformer

Vi har tidligere på forskellig vis karakteriseret PISA's test-setup. Nogle af de væsentligste træk er den *formelle* resultatorientering, den *individuelle* afregning og det *skriftlige* testformat. For at få et indtryk af, hvorledes den danske praksis matcher disse træk, har vi stillet lærerne spørgsmålet ”*Hvor ofte benytter du følgende evalueringsform.... i undervisningen?*”

Svarfordelingerne er vist i Figur 4. Heraf fremgår det, at hyppigheden af samtlige evalueringsformer generelt er ganske høj (”af og til” eller ”hyppigt”) hos store lærergrupper. *Det PISA-lignende format med individuelle, skriftlige tests viser sig at være den markant hyppigste evalueringsform.* Om dette træk er udtryk for en nyorientering foranlediget af de senere års ”oprustning” på testområdet og/eller forudgående PISA-runder, kan vi ikke afgøre. Pointen er imidlertid tydelig nok: *Danske elever er bestemt ikke uvante med en PISA-lignende testform.*

Umiddelbart ser det paradoksalt ud, når den næsthyppest evalueringsform samtidig er *uforpligtende samtale med elevgruppe/klassen*. Denne evalueringstype repræsenterer på stik modsat vis det uformelle, det kollektive og det verbale. En nærmere analyse, som også omfatter lærernes åbne svar, peger dog på, at der er tale om evaluering med to vidt forskellige funktioner. Hvor den første i høj grad er summativ og tjener som grundlag for standpunktsgivning, forældre-konsultation m.m., udgør den anden først og fremmest en sondering og justeringsmulighed i den daglige undervisning. Man kan sige, at de to foretrukne former begge er overkommeligt administrerbare ift. en helklasse-situation, og at de *tilsammen* leverer den nødvendige information og den detaljeringsgrad, som læreren har brug for.



Figur 4: Hvor ofte benytter du følgende evalueringsform i undervisningen?

De kvantitative besvarelser præsenteres af pædagogiske grunde først, idet de giver et meget godt første overblik over tendenser i lærersvarene. De har selvfølgelig nogle begrænsninger i form af de af os fastlagte spørgsmål og svarkategorier. Derfor er det meget godt at kunne krydsklippe til svarene fra det åbne spørgsmål *Beskriv kort hvordan du hyppigst evaluerer din undervisning*. Dette spørgsmål blev stillet forud for gruppen med lukkede spørgsmål i vores survey.

De åbne svar udgør et meget fyldigt materiale, som frem for alt viser den store variation i de anvendte evalueringsformer. Det er umuligt at opgøre svarene gennem en enkelt, udtømmende opdeling, idet flere forskellige formater og typer blandes. Nogle angiver former som mundtlig og skriftlig, andre taler om tests og prøver, andre igen beskriver konkrete metoder, fx begrebskort eller logbog, og nogle skelner mellem individuel og gruppebaseret evaluering.

Vi har derfor optalt fordelingen for forskellige opdelinger.

Ca. 9 % angiver som den hyppigste form som værende ”kun samtaler”, fx i form af:

”Gennem samtale med eleverne”

”Ved klassesamtale. Hvis det omfatter et større projekt vil det som regel være mere gruppevise og en karakter”

”Mundtlig evaluering”

”Spørgsmål og samtale”.

En lille halvdel af lærerne, ca. 43 %, anvender som den hyppigste evalueringsform *kun skriftlig evaluering*:

"Forskellige former for skriftlige prøver"

"Som oftest evalueres eleverne, ved at de udfærdiger en rapport om emnet, hvor de bl.a. skal tage stilling til egne forsøgsresultater med udgangspunkt i den gennemgåede teori"

"Skriftlige spørgsmål som jeg selv fremstiller. Eleven må bruge alle hjælpemidler"

"Gennem logbøger og skriftlige opgaver".

Op mod halvdelen af lærerne, ca. 48 %, angiver *"en kombination af samtale og skriftligt produkt"*, typisk en prøve eller en rapport, fx

"Samtale og test"

"Ved at lade eleverne besvare spørgsmål mundtligt i klassen. At uddele skriftlige prøver"

"Mundtlig overhøring og skriftlige rapporter"

"Klassesamtaler, prøver efter hvert forløb"

"Vi snakker om udførte forsøg. Eleverne får skriftlige prøver".

Af de, der anvender skriftlig evaluering, skriver mange, at det sker i form af tests. I alt 56 % af lærerne angiver, at tests indgår som deres mest anvendte evalueringsform.

Relativt få lærere anfører nyere og formativt spidsstillede evalueringsformer som deres mest anvendte evalueringsform. Det gælder således kun for ca. 6 % af lærerne. Tilmed beskrives de oftest i kombination med mere traditionelle former, hvilket er tydeligt i nedenstående eksempler:

"Jeg bruger logbog i alle fagene. De har desuden en portefølje i alle deres fag. De afleverer rapporter om hvert emne og små opgaver. Mundtligt skal de forklare emnerne og finde forsøg der underbygger teorien."

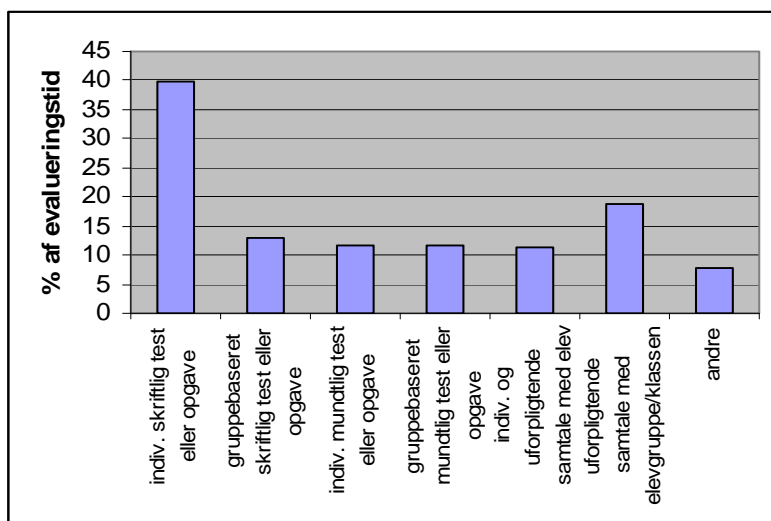
"Eleverne samler op på forløb ved at lave en kort præsentation af emnet til deres portefølje. Derudover anvendes multiple choice test."

"Test. Fælles begrebkort. Eleven skriver rapport. Plancher. Fremlæggelser. Logbog."

"Begrebkort, hvor eleverne sætter fagtermene sammen til en helhed. Snak på klassen, tests."

"I form af prøver og i logbog samt mundtlig evaluering."

Billedet af de individuelle, skriftlige tests og klassesamtalerne som de dominerende evalueringsaktiviteter underbygges således af de åbne lærersvar. Dette indtryk ændres heller ikke væsentligt, såfremt man beder lærerne angive det relative tidsforbrug, frem for hyppigheden af de forskellige evalueringsaktiviteter. Den klareste ændring er dog, at PISA-matchende, individuelle, skriftlige tests blot træder endnu tydeligere frem.

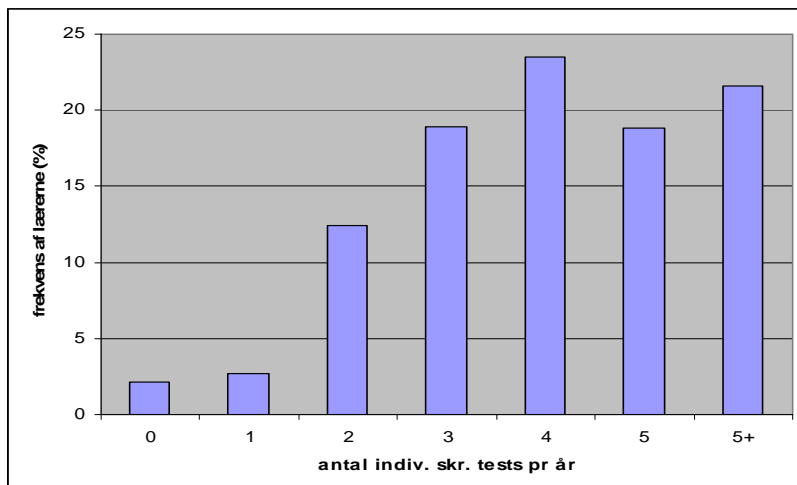


Figur 5: Relativt tidsforbrug anvendt på evaluering²

I forbindelse med Fig.5 er det værd at bemærke, at den angiver, hvorledes de forskellige evalueringsaktiviteter tidsmæssigt vægtes relativt *i forhold til* hinanden. Hvor meget tid, der i absolut mål går på evaluering, fremgår ikke af den kvantitative undersøgelse. Man kan få et vist (men usikkert) indtryk af dette via vores indledende lærerinterviews. Her spurgte vi ind til, hvor stor en del af faktiske undervisningstid lærerne anvendte til ”direkte tests” (svarende til søjlen yderst til venstre i Fig.5). Den øvre grænse i svarene var ”under 10 %”. Den nedre grænse blev markeret af en lærer, som udtalte: ”..det er under 5 % - for det er jo ikke der læringen finder sted!”.

Som en triangulering af de tidligere mål har vi endelig spurgt om antallet af individuelle skriftlige tests pr. år i en typisk naturfagsklasse. Billedet (Fig.6) viser, at eleverne typisk oplever dette 4 gange pr. år og fag. Det er altså helt konsistent med det foregående.

² NB: Summen af %-tallene overstiger 100, da lærerne lejlighedsvist har tildelt mere end 100 %, bl.a. pga. kategorien ”andre”.)

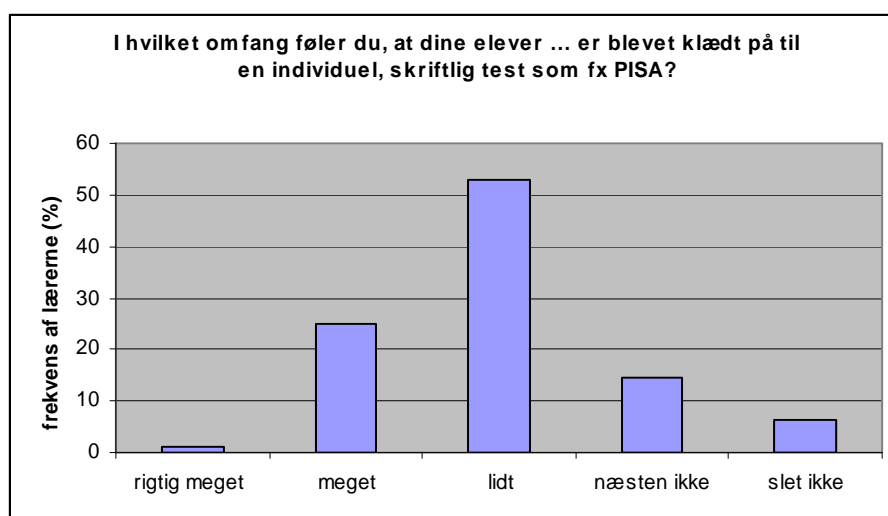


Figur 6: Antallet af individuelle skriftlige tests pr. år

En nærmere analyse godtgør, at hyppigheden af denne type testning *ikke* signifikant afhænger af lærernes linjefagsbaggrund. Der er (mindst) to forskellige udlægninger af denne observation: I den ene har naturfagslærerne ved at snakke sammen og undervise i en række forskellige naturfag skabt en *fælles og distinkt evalueringsspraksis for naturfagene under ét*. Alternativt kan det ses som et udtryk for, at en lærers evalueringsspraksis dybest set *ikke* afhænger af fag, men i langt højere grad er betinget af personlige præferencer og/eller lokale *skole*påvirkninger.

Føler lærerne, at de har klædt deres elever på til en individuel skriftlig test, som fx PISA?

I forlængelse af ovenstående har vi direkte spurgt lærerne, om de føler, at de har ”klædt eleverne på” til PISA test-lignende omstændigheder. Det typiske lærersvar er her, at det har de gjort i en udstrækning svarende til ”*lidt*” (Fig.7)

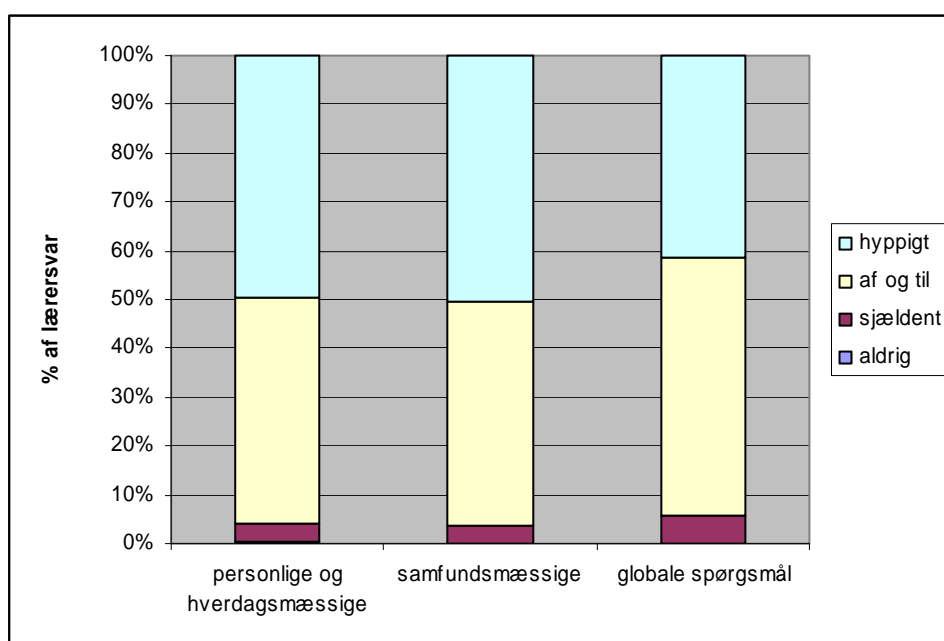


Figur 7: Har du ”klædt eleverne på” til PISA-test?

Det er dog vigtigt at fastslå, at lærernes svar her væsentligst forholder sig til *test-formen*. PISA-opgavernes konkrete karakter (fx tekstmængde), indhold, samt de operationaliserede kompetencer hemmeligholdes i en udstrækning, så det er svært for lærerne at orientere undervisningen derefter. Det er nærliggende at antage, at lærernes moderate tilslutning til spørgsmålet netop skyldes, at de ikke *indholdsmæssigt* – via ”*teaching for the test*” – har kunnet forberede eleverne til PISA. Det er imidlertid et vilkår, som de danske deltagere i PISA deler med alle andre.

Hvilke anvendelsessammenhænge indgår den naturvidenskabelige viden i, når den evalueres?

PISA-opgaverne er indlejret i en kontekst, men ”*It is not possible to be more prescriptive of items across contexts*” (ScFor (0407)1 p. 33). Alligevel er det relevant at få et indblik i, hvilke kontekster danske naturfagslærere inddrager i deres undervisningsmæssige evalueringer. Vi har her spurgt, om naturfagslærernes evaluering inddrager de tre overordnede kontekst-kategorier, som PISA’s teoretiske ”Framework” opererer med (Fig. 8).



Figur 8: Inddrager evalueringen de tre overordnede PISA kontekst-kategorier?

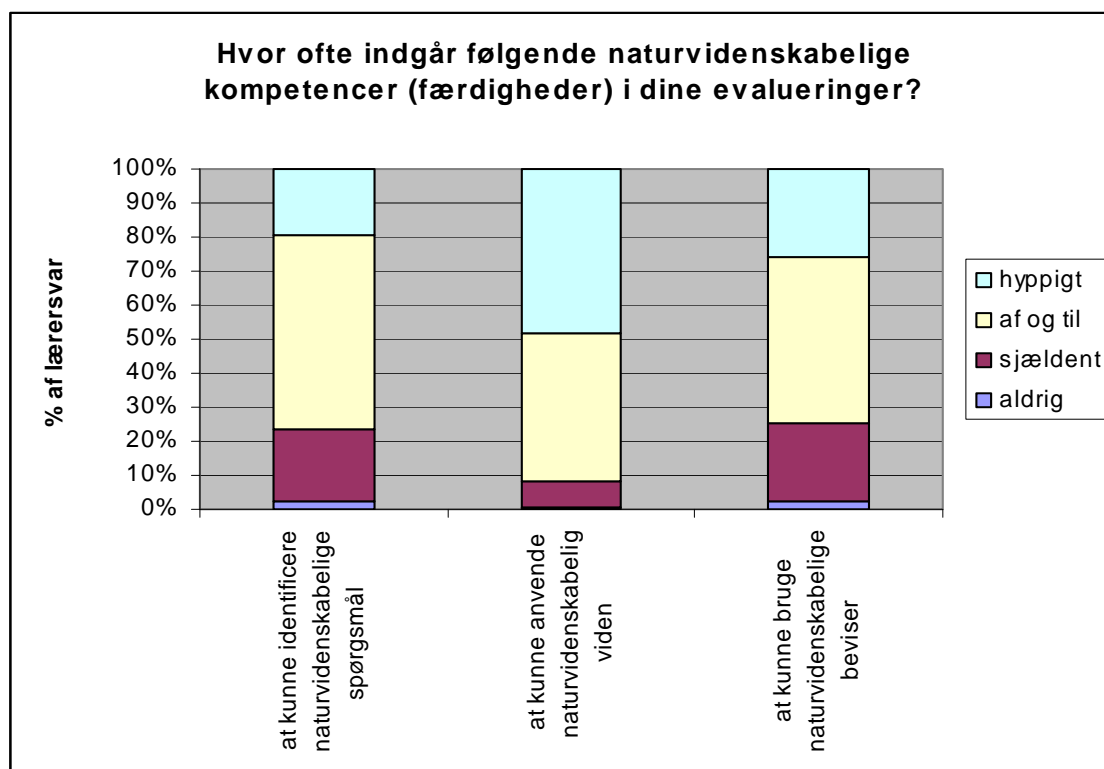
Lærerne tilkendegiver klart, at deres evaluering inddrager de PISA-relevante kategorier, formentlig fordi disse langt hen ad vejen svarer til prioriteringerne i *Fælles Mål for hhv. Natur/teknik og Fysik/kemi* (”nære omverden”, ”fjerne omverden” og ”menneskets samspil med naturen” hhv. ”hverdag og samfund”). Der er ikke væsentlig forskel på, hvor hyppigt de tre kontekster evalueres. Især den

hyppige inddragelse af globale spørgsmål i evalueringen synes forenelig med PISA-opgaverne. I udvælgelsen af disse søger man at minimere kulturelt bias – hvorfor velegnede opgaver har en indbygget orientering mod det fælles og det universale, hvortil det globale hører.

Evalueres PISA-lignende kompetencer?

Som tidligere omtalt er et antal naturfaglige kompetencer ekspliciteret i PISA's teoretiske rammesætning. Ideelt set tjener opgavernes enkelte underspørgsmål som operationalisering af evnen til at *identificere naturvidenskabelige spørgsmål*, evnen til at *anvende naturvidenskabelig viden* og evnen til at *bruge naturvidenskabelige beviser*. Det har været PISA's intention at bruge 25-30 % af test-tiden på den første kompetence og 35-40 % på de øvrige.

I denne undersøgelse har vi spurgt til netop disse kompetencer, men i en udfoldet formulering, hvor de forskellige kompetencer eksemplificeres. Svarene fremgår af Fig.9, og det ses, at samtlige kompetencer evalueres "hyppigt" eller "af og til" hos mindst 75 % af lærerne.



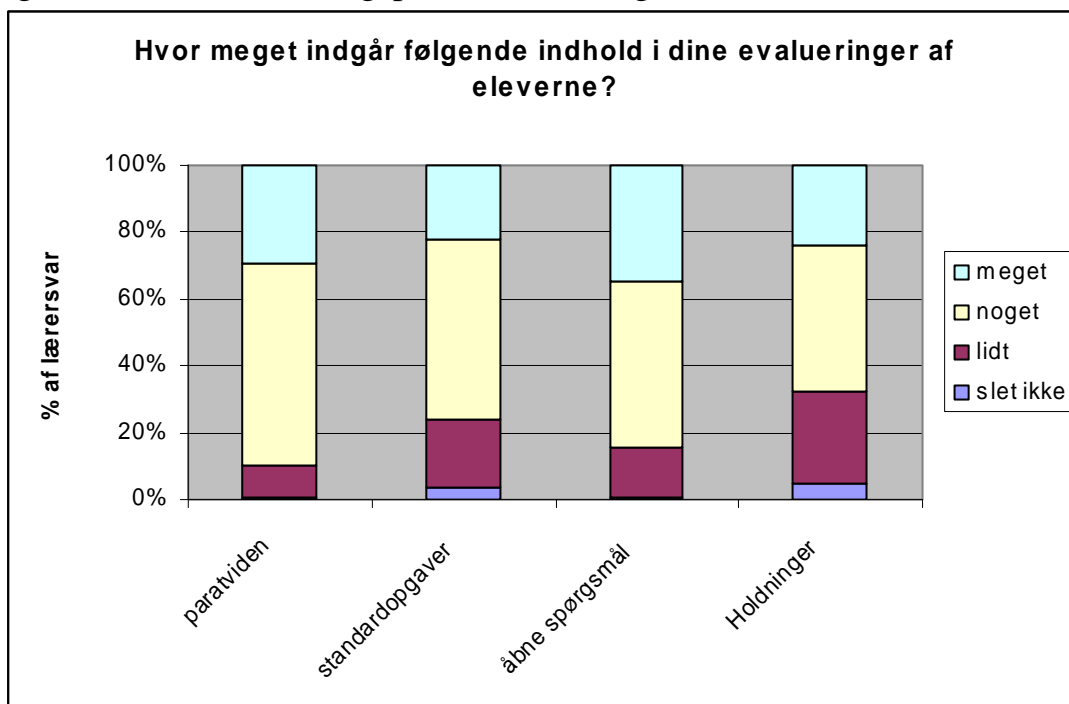
Figur 9: Eksempler på kompetencer, der evalueres

Indholds- og opgavetyper, som indgår i den typiske naturfagsevaluering

PISA-konsortiet har en hensigtserklæring om at lade mindst halvdelen af opgaverne være af multiple choice-typen. Dette letter rettetarbejdet (og dermed den økonomiske testningsbyrde) og muliggør en pålidelig scoring – samtidig med, at det begrænser, hvad man validt kan udmåle.

Åbne spørgsmål indebærer, at eleven selv skal formulere et svar. En del af de på overfladen åbne spørgsmål lukkes imidlertid af PISA's scoringsprocedure, idet man ofte fastlægger scoringskriterierne så rigtigt, at kun bestemte "trigger"-ord belønnes, mens en samlet vurdering, som måske ville tilgodese en mere kompleks forståelse frafaldes. Den langt overvejende del af PISA-opgaverne er de facto af den lukkede type (ét bestemt statement, bestemte trigger-ord).

I den danske skolekontekst har vi fundet det relevant at se mængden af evaluering rettet mod *Paratviden* og håndtering af *Standardopgaver* som indikatorer for brugen af lukkede opgavetyper. Vurderet ud fra Fig.10 anvendes både lukkede og åbne spørgsmål hyppigt, med en samlet tendens til dominans af mere lukkede formater. Heller ikke her kan man øjne nogen modstrid mellem PISA og den danske evalueringspraksis i naturfag.



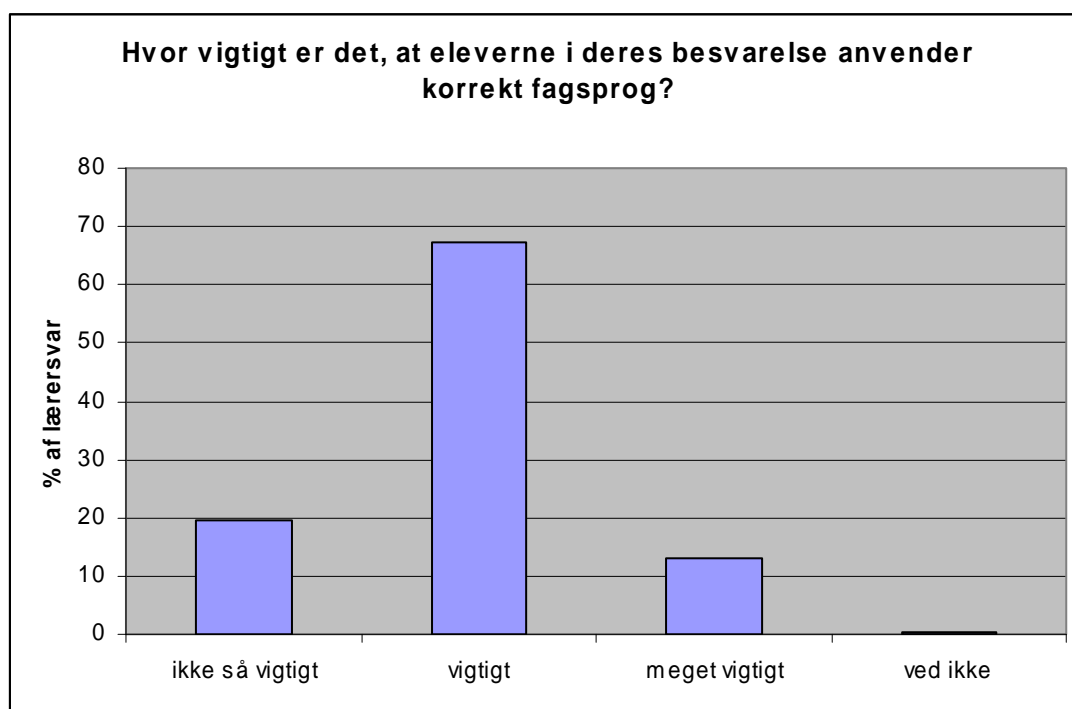
Figur 10: Hvad evalueres der på?

Uden at det i øvrigt har nogen indflydelse på de danske testresultater, kan man notere sig, at ca. 2/3 af naturfagsunderviserne forsøger at indfange elevholdnin-

ger i deres evaluering. I PISA-setuppet har man en hensigtserklæring om at lade det holdningsprægede, *affektive domæne* fylde 17 % af testtiden.

Betydningen af korrekt fagsprog i elevsvar

Med den tidligere omtalte temmelig rigide scoring af de åbne spørgsmål er det ikke nok at kende essensen af en opgaves problemstilling, man skal tillige anvende bestemte fagspecifikke termer i besvarelsen. I slående kontrast hertil nævnes fagsproget overhovedet ikke i de formelle trinnål for Fysik/kemi's 9. eller 10. klassetrin. Det er derfor særdeles relevant at se på, hvorledes dette aspekt alligevel udfoldes i lærernes *evalueringsspraksis*. Fig.11 viser lærernes vægtning af korrekt fagsprog i elevbesvarelser i evalueringssammenhæng.



Figur 11: Vægtning af korrekt fagsprog

Svarene kan formentlig tolkes derhen, at fagsproget nok opfattes som ”vigtigt”, men at andre aspekter meget vel kan være *endnu* vigtigere. I forlængelse af denne tolkning svarede én af lærerne fra de indledende interviews: ”*For mig betyder det ikke det store, om de siger ’elektronspin’ – eller ’de der, der gør sådan rundt om kernen’*”.

Opsummering:

Er dansk evalueringspraksis i naturfag i rimelig overensstemmelse med PISA's testkoncept?

På det foreliggende grundlag må konklusionen være, at evalueringspraksis i dansk naturfag synes at være endog ganske forenelig med PISA's testformat. Undersøgelsen godtgør således:

- 1) at individuel og skriftlig testning er den hyppigst anvendte evalueringsform
- 2) at PISA-relevante kontekster og kompetencer indgår i evalueringen hos praktisk taget alle lærere
- 3) at lukkede og åbne opgaveformater anvendes i PISA-sammenligneligt omfang
- 4) at anvendelsen af korrekt fagsprog i elevbesvarelser vægtes rimelig højt.

Eneste lille forbehold i datamaterialet kan aflæses i lærernes moderate tro på at have "klædt eleverne på" til en individuel, skriftlig test som PISA. Dette ses som et udtryk for, at lærerne kun har kendskab til de offentliggjorte, overfladiske træk ved PISA og derfor ikke føler, at de har kunnet forberede eleverne på vanlig, indholdsorienteret vis.

I tilgift til denne undersøgelses mere generelle afdækning, ville det have været interessant at få lærernes tilbagemelding på den *operationaliserede* version af PISA, dvs. få deres kommentarer til en booklets helt konkrete opgavekontekster, kompetenceudfoldninger og scoringskriterier. Den tilsyneladende overensstemmelse i de overordnede formater kan godt blive gradbøjet noget, når de håndfaste udmøntninger studeres. Desværre har krav om hemmeligholdelse af PISA-opgaverne forhindret en sådan mere dybgående undersøgelse.

Indkredsning af danske læreres evalueringspraksis i naturfag

Lærernes vigtigste begrundelser for at evaluere

Kortlægningen af lærernes begrundelser er foregået med både et åbent og et antal lukkede Likert-skala-spørgsmål. Det åbne spørgsmål: "*Hvad er din vigtigste begrundelse for at evaluere?*" blev bevidst placeret først. Svarformatet og placeringen sikrede plads til lærernes egne formuleringer samt et indblik i de begrundelser for evaluering, som lærerne spontant har opmærksomhed på og prioriterer. I de efterfølgende kvantitative spørgsmål blev deres opmærksomhed i stedet *henledt* på begrundelsesaspekter, som *vi* havde formuleret og valgt. Derved har der meget vel kunnet ske en vis forskydning i prioriteringerne.

Analysen af lærernes åbne begrundelser har vist, at de i meget høj grad kan kategoriseres efter:

- *Hvad der er evalueringens genstand* – er det eleverne eller deres undervisning?
- *Hvor evalueringens fokus ligger* – er det på læring eller på affektive, oplevelsesorienterede forhold?
- *Hvilket sigte evalueringen har* – er der tale om en summativ afregning eller om et formativt, fremadrettet perspektiv?
- *Hvem der "ejer" evalueringen* (udformning, resultater) – er det læreren som beholder, fortolker og beslutter i forlængelse af evalueringen – eller bliver eleverne involveret i evalueringen (fx via medbestemmelse, når den udformes, eller ved at få detaljerig indsigt i dens resultater – med mulighed for at kvalificere fremtidige handlinger)?

Evalueringens genstand: Her viser lærernes svar klart, at *evaluering først og fremmest handler om eleverne*. Blandt de dechiffreerbare svar begrunder ca. 93 % af lærerne deres evaluering med henvisning til eleverne. Et par typiske lærersvar indenfor denne kategori kunne være:

"For at se om eleverne har lært det jeg forventede"

"For at have føling med, om alle får det rigtige udbytte af undervisningen."

Ca. 40 % af lærerne anfører dog *tillige*, at de foretager evaluering for at få tilbagemelding på deres undervisning. Kun 7 % begrunder *alene* deres evaluering med ønsket om undervisningsevaluering. Eksempler på lærersvar indenfor denne kategori kunne være:

"For at kunne forbedre undervisningen."

"For at tilpasse undervisningen og drage erfaringer af de ting der gik galt, så det kan gøres anderledes næste gang."

Evalueringens fokus: De foregående svareksempler demonstrerer samtidig, at ikke alle lærersvar udtrykker et entydigt evalueringssfokus. I situationer som denne, kan man ganske enkelt ikke aflæse, hvorvidt hensigten med forbedringer og tilpasninger er at forbedre elevernes faglige udbytte eller deres oplevelse af undervisningen. Kun ca. 70 % af lærersvarene lader sig således sikkert indplace-re. Alligevel er det påfaldende, at ca. 93 % af disse *udelukkende* fokuserer på at *evaluere læringsudbyttet*, fx via formuleringer som:

[vigtigste begrundelse:] *"For at have føling med om alle får det rigtige udbytte af undervisningen."*

[vigtigste begrundelse:] ” For at få overblik over, hvor eleverne befinder sig fagligt - er de med, samt for evt. at kunne ændre på min undervisningsform undervejs, hvis for mange ikke er med.”

Praktisk taget alle lærere, som ikke *udelukkende* går efter læringsudbyttet, går efter *både* læringsudbytte og affektive tilbagemeldinger. Som et typisk eksempel kan fremdrages:

”At undersøge om eleverne har opnået forståelse for det gennemgåede stof.”

”At undersøge elevernes oplevelse af undervisningsformen.”

Man kan derfor med stor rimelighed fastslå, at *læringsaspektet er det absolutte fokus, uanset om det er eleverne eller undervisningen der evalueres.*

Evalueringens sigte: Her er tendensen mindre entydig: Godt 20 % af lærerne omtaler *udelukkende* et fremadrettet sigte med deres evaluering, mens ca. 43 % *udelukkende* angiver et summativt sigte. De resterende ca. 37 % kombinerer disse perspektiver. Eksempler på de tre typer af lærersvar:

[formativ variant1]: ”Se om eleverne får noget ud af undervisningen. Vurdere, om min undervisning kan gøres bedre.”

[formativ variant2]: ”At sikre mig, at eleverne bliver bevidste om, hvordan deres udbytte bliver af undervisningen.”

[summativ variant]: ”For at se, om stoffet er indlært, og for at kunne bedømme eleverne.”

[kombinationsvariant1]: ”At finde ud af, om eleverne har lært det de skal ifølge ”Faghæftet for geografi”, og for at kunne gøre min undervisning endnu bedre.”

[kombinationsvariant2]: ”At eleverne får samlet op på indholdet og får indsigt i målene for undervisningen. At jeg får en fornemmelse af elevernes udbytte.”

Det summative sigte træder tydeligst frem i begrundelserne, bl.a. illustreret af, at ca. 80 % af lærerne i en eller anden udstrækning udtrykker en summativ tænkning. Dette underbygger den tidligere analyse, som indikerede et relativt match mellem naturfagslærernes evalueringspraksis og PISA test-setuppet. Samtidig giver knapt 60 % af lærerne i en eller anden form udtryk for en formativ evalueringstænkning. Man kan derfor ikke betegne dette aspekt som uanseligt eller uudfoldet. Det fremgår endvidere af citat-bidderne, at lærerne tænker deres evaluering formativt både i tilknytning til undervisningen og til eleverne. Disse to tendenser fremstår nogenlunde lige tydelige af materialet.

Elevernes ejerskab og involvering i evalueringen er rimeligvis en forudsætning for, at den kan bidrage til elevernes udvikling fremadrettet. Her fremgår det af lærersvarene, at for hver gang én lærer åbner op for elevernes medejerskab til evalueringens resultaterne, vælger ca. tre lærere at forbeholde sig retten til selv at tolke og agere på disse. Der synes derfor at være et stort og uudnyttet potentiale her. Svarene fra den foregående kategori indeholder allerede eksempler på hver af de to elevinvolveringsmodeller.

Lærernes begrundelser – opsummering på de åbne svar

De åbne svar udtrykker selvfølgelig en stor mangfoldighed – og alligevel tegner der sig et billede af en faglig evalueringspraksis, som i dominerende grad handler om elever, deres faglige læring, først og fremmest i et summativt/afregningsorienteret sigte, og hvor eleverne i ringe udstrækning inddrages i evalueringen.

Med udgangspunkt i trends i tidens evalueringforskning vil man kunne problematisere, at den produktive, formative elevevaluering kun prioriteres af knap 30 % af lærerne, samt at eleverne i meget ringe grad involveres i evalueringens præmisser og produkter. Omvendt må man konstatere, at der er tale om et noget nær ”perfect match” til PISA’s hidtidige praksis.

Lærernes begrundelser – lukkede spørgsmål.

Efter det åbne spørgsmål blev lærerne bedt om at angive deres vægtning af en række mulige begrundelser (E2-E5) på en 5-punkts Likert-skala.

E2: Jeg bliver klogere på hvad den enkelte elev har lært i undervisningen

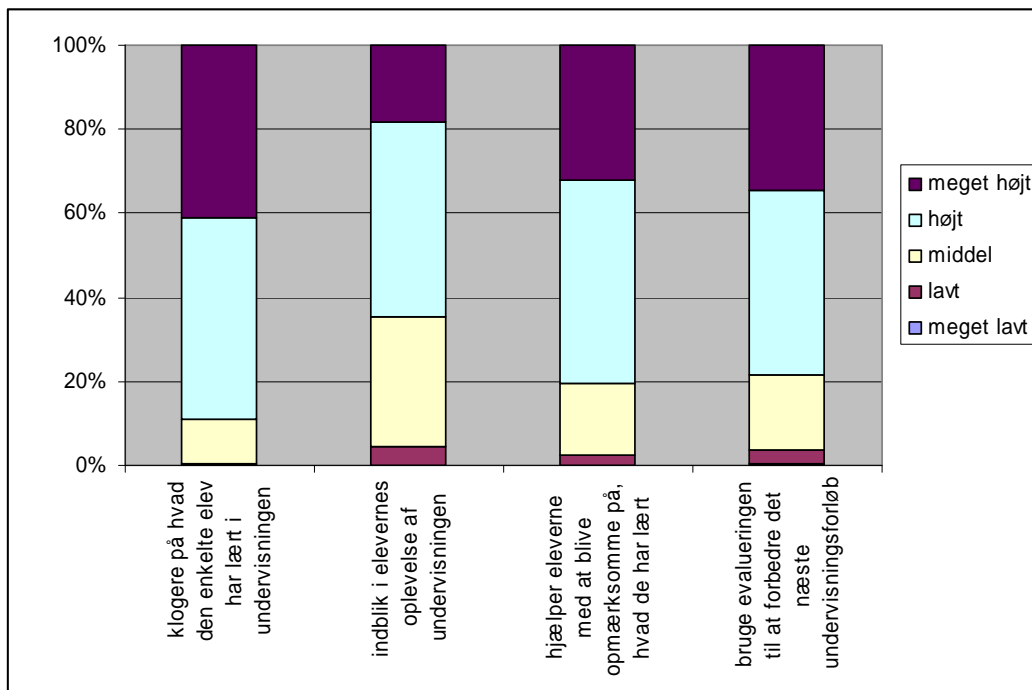
E3: Jeg får indblik i elevernes oplevelse af undervisningen

E4: Jeg hjælper eleverne med at blive opmærksomme på, hvad de har lært

E5: Jeg kan bruge evalueringen til at forbedre det næste undervisningsforløb med klassen

Disse spørgsmål er formuleret forud for etableringen af kategorier for de åbne svar. Alligevel vil man ret nemt kunne indplacere dem i dette system.

Svarfordelingerne for de fire lukkede begrundelsesspørgsmål ses på Fig.12. Det fremgår, at alle begrundelser vægtes ganske højt (middelværdier svarende til de to øverste kategorier ’meget højt’ eller ’højt’). Det er dog signifikant, at begrundelsen E2 (elever som evalueringens genstand, læring som fokusaspekt og lærerejerskab af resultaterne) vægtes højest. Tilsvarende er det signifikant, at evaluering mindst af alt har til formål at give indblik i elevernes *oplevelse* af undervisningen (E3: elever som genstand, affektivt fokusaspekt, lærerejerskab).



Figur 12: Lærernes begrundelser for at evaluere

Resultaterne fra de lukkede spørgsmål understøtter således hovedtræk fra analysen af de åbne – om end trækkene fremstår lidt mindre skarptskårne.

Hvornår evalueres der – undervejs eller afslutningsvist?

Gennem analysen af lærernes begrundelser for evaluering fik vi et indtryk af, hvorledes de vægter summative hhv. formative aspekter. Summativ evaluering har afregningskarakter og er derfor tilbageskuende: Hvad har eleven *lært* eller hvordan er det *gået*? Formativ evaluering er en del af en læringsproces og fremadrettet. Det er fristende at lave afregningen ved et forløbs afslutning, og det fremadrettede undervejs i den løbende proces. Men principielt er der intet til hinder for at indsamle summative delresultater undervejs og tage en formativ slut-evaluering med henblik på næste forløb. Derfor siger disse prædikater reelt intet om, *hvornår* evaluering placeres i den pædagogiske tilrettelæggelse af et undervisningsforløb.

Vi har ønsket at belyse den tidsmæssige placering af naturfagsevaluering i et typisk forløb. Til dette formål har vi indføjet spørgsmål om løbende, hhv. afslutningsvis evaluering.

Graden af løbende evaluering indikeres via en stillingtagen til udsagnet ”I det enkelte undervisningsforløb evaluerer jeg løbende eleverne, så de hver især bliver klar over deres standpunkt....”

Svarskalaen for dette udsagn var beklageligvis ikke optimal, idet to svarkategorier ved en fejl fremstod som sammenfaldende. Ved at slå svarprocenterne fra disse sammen kan man imidlertid fastslå, at 72 % af lærerne faktisk foretager en eller flere faglige evalueringer undervejs i et forløb. Rigtig mange lærere indsamler altså løbende data, som i princippet vil kunne bruges såvel med et summativt som et formativt sigte. De foregående analyser sandsynliggør, at det første vil dominere.

Graden af afslutningsvis evaluering indikeres ved udsagnet ”Jeg evaluerer elevernes viden efter afslutningen af et undervisningsforløb”. Resultaterne (Fig.13) viser først og fremmest, at evaluering er en del af hverdagen i de undersøgte klassetrin: Hele 96 % af lærerne angiver, at de ”altid” eller ”som oftest” afslutter et forløb med en evaluering.

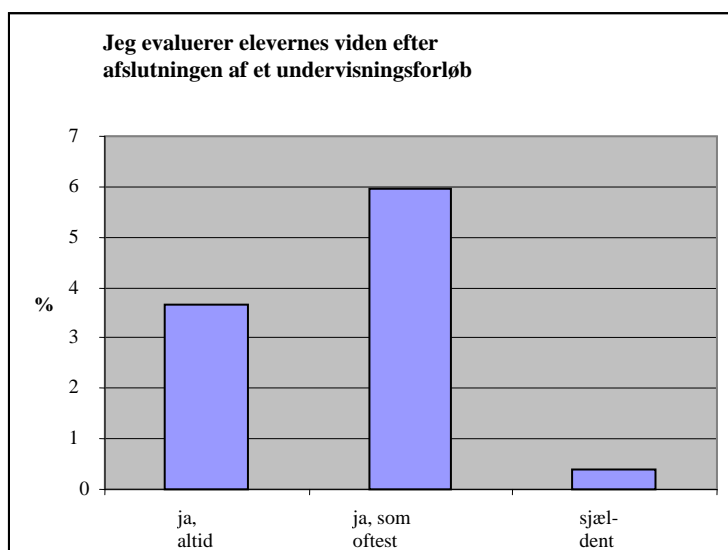


Fig. 13: Graden af afslutningsvis evaluering

Er eleverne bekendte med kriterierne for den skriftlige evaluering?

Der er ingen tvivl om, at kompetent besvarelse af en opgave forudsætter, at man har kendskab til de kriterier, som besvarelsen efterfølgende vil blive vurderet efter. Kun derved får eleverne mulighed for at agere metakognitivt og disponere deres indsats. Det er derfor relevant at finde ud af, om naturfagslærerne i almindelighed ekspliciterer deres evalueringskriterier for eleverne. Lærernes selvrapport til denne undersøgelse giver anledning til at tro, at noget sådant er ganske almindeligt (på de undersøgte højere klassetrin), jf. Fig.14.

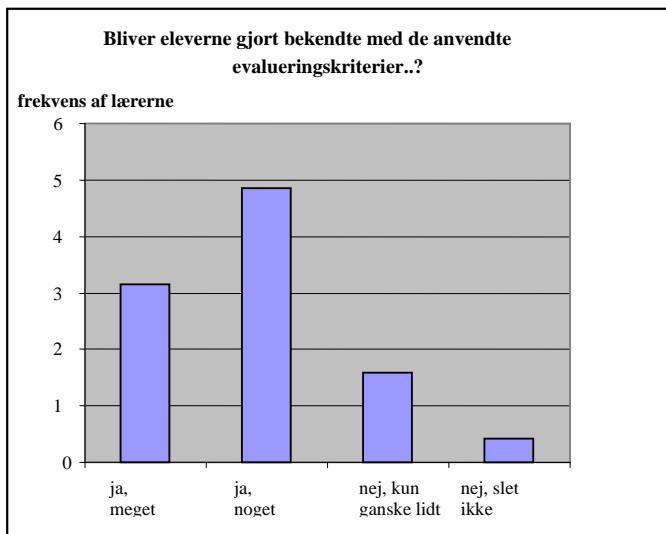


Fig.14: Elevernes kendskab til evalueringskriterier

Anvendelsen af feedback til eleverne i kølvandet på evaluering

En evaluering formative elevpotentialer kan kun udfoldes, såfremt eleven får en tilbagemelding i en brugbar form. Der findes en omfattende pædagogisk og fagdidaktisk litteratur om den rette *formative feedback*. I den hidtidige form har PISA i DK været administreret således, at hverken elever eller lærere har fået brugbar feedback på deres præstationer, et forhold som bl.a. problematiseres i OECD-rapporten om dansk "Educational R&D" (Hargreaves et al., 2004)). Nærværende undersøgelse af danske naturfagslæreres brug af evaluering viser, at de (efter egen opfattelse) i meget stor udstrækning giver en eller anden form for feedback i forlængelse af evalueringer. Over 87 % giver det "altid" eller i hvert fald "hyppigt", jf. Fig.15.

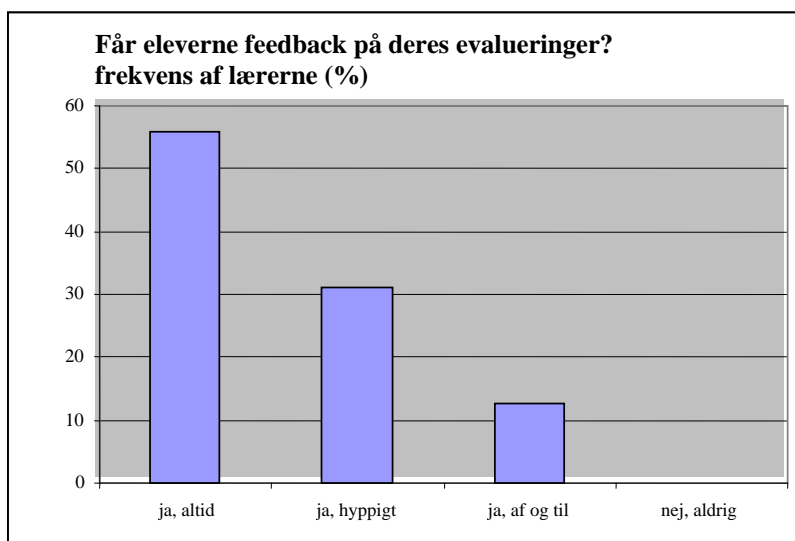


Fig.15: Får eleverne feedback på evaluering?

Karakterer er isoleret at betragte som en summativ evalueringsform, mens gode råd, anvisninger og andre fremadrettede kommentarer er eksempler på formative feedback-former. Af undersøgelsen her fremgår det, at både skriftlig og mundtlig tilbagemelding i form af kommentarer forekommer stort set lige så tit som tilbagemelding med karakter. Mundtlige kommentarer er endda signifikant mest udbredt (Fig.16). Ud fra det foreliggende har vi ikke mulighed for at afgøre *hvor* formative kommentarerne er. Kommentarer som fx ”*Fint*”, giver måske eleven lidt selvværd, men kvalificerer ikke til næste læringsopgave.

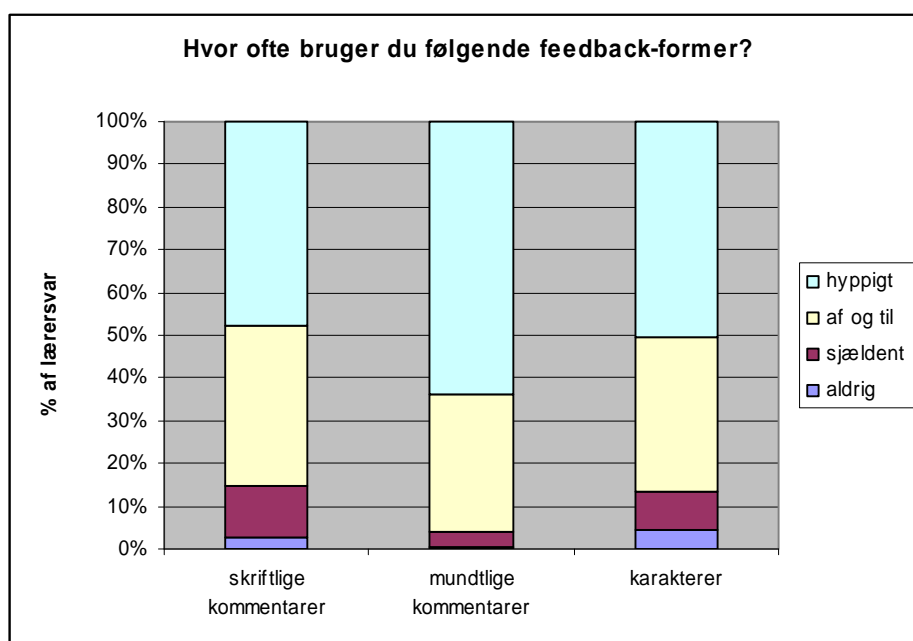


Fig.16: Feedback-former

Hvor meget fylder den praktiske dimension i den samlede evaluering af eleverne?

Af praktiske og økonomiske årsager har PISA valgt ikke at inddrage det praktiske arbejde i deres test-setup³. Det praktisk-eksperimentelle arbejde fylder imidlertid ganske meget i folkeskolens læreplaner og er en integreret del af den bundne mundtlige eksamen i faget fysik/kemi (BEK nr 863 af 05/07/2007).

Vel vidende at der ofte er et gab mellem intenderede og implementerede læreplaner har det generel interesse at fastslå, om de formelle papirers høje vægtning af praktisk-eksperimentelt arbejde afspejles i naturfagslærernes evalueringspraksis. Ud fra en PISA-betragtning er spørgsmålet interessant, idet en kraftig praktisk vægtning af dette aspekt kunne indikere, at væsentlige dele af danske ele-

³ Tidligere har de internationale TIMMS-undersøgelser ellers demonstreret, at det *kan* lade sig gøre at gennemføre sådan ”large-scale performance-testing”

vers kompetenceprofil og normale evalueringsgrundlag forbliver usynlige i PISA-testningen. Fig.17 viser, at lærerne faktisk tillægger den praktiske dimension en betragtelig vægt i deres evaluering. Her er altså tale om et forhold, som ikke forringer den danske PISA-præstation, men måske *forvrænger* billedet af elevernes samlede formåen – ved ganske enkelt at undertrykke dele af dette billede (jf. dog den tidligere omtale af danske elevers præstation i den praktisk orienterede TIMMS *Performance Assessment* (Weng and Hoff, 1999))

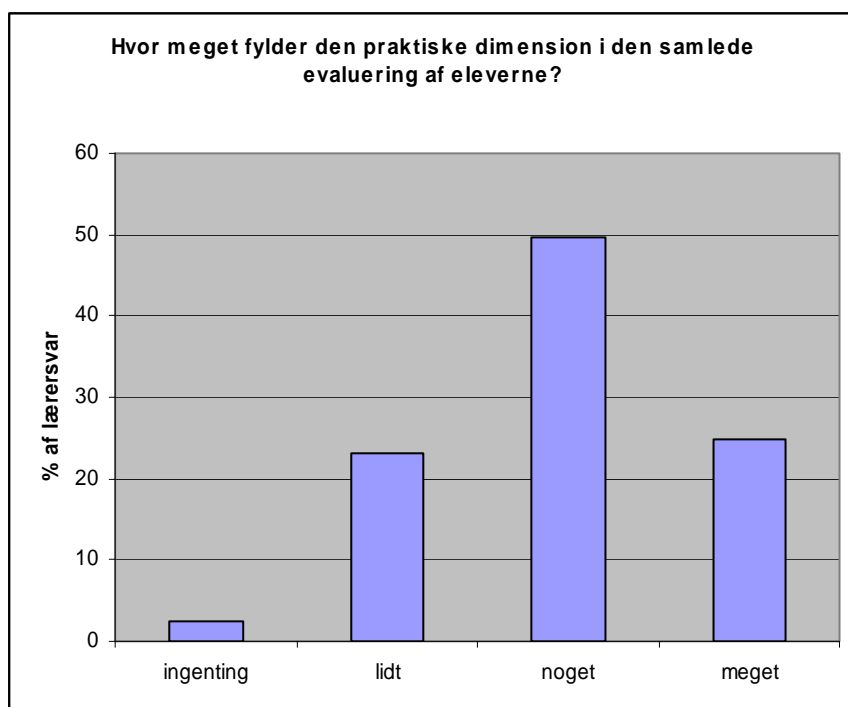


Fig.17: Den praktiske dimension i evalueringen

I naturfag, hvor den afsluttende prøve har en praktisk-eksperimentel dimension, trækker dette tydeligvis spor baglæns til den daglige undervisning og evaluering. Dette kom fx kraftfuldt til udtryk i ét af vore indledende lærerinterviews: *”Hele vores daglige undervisningsform lægger op til den eksisterende prøveform, hvor de har mulighed for at snakke om et større emne bygget op omkring et forsøg.”*

Mønstre i lærernes brug af evaluering

Hidtil har vi i analysen brugt betegnelsen *evalueringsspraksis* for den enkelte lærers omgang med evaluering. Det giver kun mening at tale om en *evalueringsskultur*, såfremt der foreligger et mere eller mindre *fælles* sæt af grundantagelser/værdier, normer, forestillinger og sædvanlige handlinger knyttet til evaluering. I nogle tilfælde er der tale om et fælles sæt på lokalt skoleniveau, hvorfor det giver mening at tale om *skolebaseret* evalueringsskultur. I andre tilfælde kan man forestille sig et fælles normsæt etableret udelukkende blandt lærere inden

for ét eller flere naturfag (lokalt eller på tværs af skoler). Her vil man meningsfuldt kunne snakke om en *naturfaglig* evalueringskultur.

Den enkelte lærer skaber sin evalueringspraksis gennem deltagelse i krydsfeltet mellem den lokale skolekultur, forskellige faglige subkulturer og en løbende fortolkning af det centralt stillede regelgrundlag. I Folkeskolelovens §13 fastlægges de generelle og fælles rammer for evalueringsindsatsen i folkeskolen, som altså indebærer en vis overskridelse af det lokale og det fagspecifikke. Paragraffen foreskriver først og fremmest standpunktsgivning og nationale tests – men i stk 2 hedder det tillige:

”Som led i undervisningen skal der løbende foretages evaluering af elevernes udbytte heraf, herunder af elevens tilegnelse af kundskaber og færdigheder i fag og emner set i forhold til trin- og slutmål, jf. § 10. Evalueringen skal danne grundlag for vejledning af den enkelte elev og for den videre planlægning og tilrettelæggelse af undervisningen, jf. § 18, og for underretning af forældrene om elevens udbytte af undervisningen...”

Den formelle intention er således, at de enkelte fag evalueres både summativt og formativt – og med det dobbelte sigte at forbedre elevernes udbytte og undervisningens kvalitet. Langt hen ad vejen viser denne undersøgelse, at naturfagslærernes evalueringspraksis er i overensstemmelse med dette sigte. Måske er evalueringspraksis i folkeskolen ikke helt så svag, som hævdet af Hargreaves et al? (jf. p.5).

Eksistensen af gennemgående mønstre i naturfagslærernes evalueringsadfærd på tværs af skoler udtrykker de centrale bestemmelsers gennemslag og/eller styrken af faglige traditioner og faglige naturfagsnetværk ift. de lokale skolekulturer. Da vi ikke har sammenlignelig empiri fra andre folkeskolefag, er det her ikke muligt at afgøre, om evt. gennemgående mønstre er almene eller fagspecifikke. I det foregående er der omtalt en række enkelttræk, som er så udtalte hos naturfagslærerne på tværs af samplets mange skoler, at de må forstås som enten almene for folkeskolen eller som udtryk for en faggruppespecifik evalueringskultur.

Ud over disse enkelttræk har vi også iagttaget enkelte, mere komplekse mønstre. Blandt disse vil vi især fremhæve:

- *Naturfagslærerne har i almindelighed et ganske aktivt evalueringsrepertoire.*

Vi har tidligere redegjort for, hvor hyppigt forskellige evalueringsformer anvendes (Fig.4). En nærmere analyse godtgør, at over halvdelen af lærerne anvender *mindst 4* (af 6 anførte) evalueringstyper ”hyppigt” eller ”af og til”. Selvrappor-terne viser således ikke bare, at der evalueres ganske hyppigt, men også at en meget stor gruppe af lærere har en ganske varieret evalueringspraksis. Kun 20 % af lærerne har et evalueringsrepertoire med to eller færre af evalueringstyper i regelmæssig brug.

- *Naturfagslærernes almindeligste evalueringsorienteringer.*

Tidligere fremlagde vi analysen af lærernes åbne begrundelser for evaluering på en måde, så man kunne se svarfrekvenserne i forhold til fire *uafhængige* kategoriseringsdimensioner.

Vi har lavet en alternativ analyse, hvor vi tager udgangspunkt i, hvordan de fire kategorier bruges *sammen*. Man kan sige, at kategorierne tilsammen udspænder 16 forskellige sæt/mønstre af *evalueringsorienteringer*. Vi har derpå kortlagt, hvor hyppigt de enkelte lærere bekender sig til hvert af disse mønstre. Den suverænt hyppigste enkeltstående evalueringsorientering (ca. 20 % af samtlige forekomster) handler om, at evaluering udelukkende retter sig mod elever, har et entydigt læringsfokus, et rendyrket summativt sigte og undlader at involvere elever i evalueringens præmisser og resultater.

Yderligere 40 % overlapper med denne type, idet de hist og her har tilføjet et ekstra aspekt i en af kategorierne (dette er muligt, idet de åbne svar fx ofte indeholdt *både* summative og formative aspekter).

- *Naturfagslærernes evalueringspraksis er kun i ringe grad kønsspecifik.*

I undersøgelsens empiri kan man godt finde enkeltspørgsmål, hvor der er signifikant forskel på (signifikansniveau $p < 0.05$), hvad grupperne af hhv. kvindelige og mandlige lærere har svaret. Forskellene er imidlertid hverken voldsomme eller iøjnefaldende systematiske.

- *Undervisningserfaring synes at have betydning for en lærers evalueringspraksis.*

Lærernes undervisningserfaring indgår som én baggrundsparemeter i undersøgelsen. Den statistiske analyse viser, at undervisningserfaring på signifikant vis kobler til dele af evalueringspraksis. Analysen tager udgangspunkt i de fire erfaringsintervaller, som er angivet i spørgeskemaet, mens selve opsummeringen her

skelner kun mellem *de erfarne* (10 år eller mere i folkeskoleundervisning) og *de mindre erfarne* (under 10 års erfaring i folkeskolen). Undersøgelsen viser, at de erfarne i højere grad

- anvender individuelle, skriftlige tests
- styrer uden om gruppebaserede, mundtlige tests
- indgår i individuel og uforpligtende evalueringssamtale med elever
- styrer uden om karakterer
- tester paratviden.

Det ville være dybt interessant at vide, om disse træk er udtryk for tids-/generations-typiske holdninger og præferencer eller udtrykker en tilpasningsproces, hvor evalueringen udvikles gennem et årtis lærererfaringer i feltet. I sidstnævnte tilfælde vil man kunne opfatte dem som karakteristiske træk ved "bæredygtige evaluering" i naturfag. Desværre har vi ikke noget belæg for at foretrække den ene frem for den anden tolkning.

Har skolen en fælles holdning til evaluering?

Den enkelte lærers evalueringspraksis vil efter al rimelighed være mest påvirket af lokale skoleforhold, såfremt der er en ekspliciteret og mere eller mindre fælles holdning til evaluering. Vi har direkte spurgt lærerne, om de opfatter, at noget sådant er tilfældet på deres skole. Svarfordelingen fremgår af Fig.18.

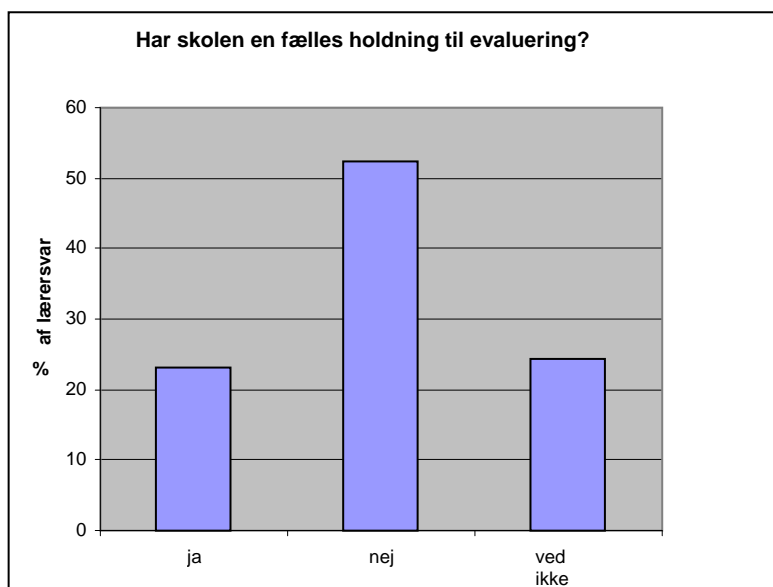


Fig. 18: Har skolen en fælles holdning til evaluering?

Man kan i hvert fald fastslå, at for hver lærer, som oplever en fælles skolebaseret holdning til evaluering, er der mindst to som *ikke* har denne oplevelse. Omkring "ved ikke"-besvarelsene er det fristende at tænke, at "Hvis der virkelig var en

fælles holdning, ville de også vide det". I en sådan tolkning hører ¾ af lærerne til på skoler, hvor der ikke er etableret en fælles evalueringskultur. På lokalt niveau er der således lang vej til realiseringen af undervisningsministeriets ambition om en gennemgribende evalueringskultur.

Lærerne i vore indledende interviews peger på evaluering som en individualiseret, hhv. et team-inspireret foreteelse:

"Hvordan vi evaluerer, hænger jo også sammen med den enkelte lærers undervisningsform. Der har jo ikke været nogle fælles krav."

" [Er der en evalueringskultur her på skolen?]: Nej, det er meget afhængigt af de teams, der sammensættes. Det er ikke ensartet – og ikke lige godt."

I princippet vil undersøgelser som denne kunne afgøre, hvor meget lærernes evalueringspraksis er præget af det lokale skoleniveau, hhv. det skoleovergribende (nationale) naturfagsdidaktiske niveau. Dvs. i princippet bør man kunne udtale sig om, hvorvidt en evt. evalueringskultur er skole- eller fag-baseret. Desværre forudsætter en meningsfuld analyse af denne slags ("nested" eller "hierarkisk lineær model"), at der er et rimeligt antal enheder på hvert analytisk niveau. Da der i vores sample højst forekommer 5 lærere fra én og samme skole, er vi afskåret fra at forfølge dette interessante aspekt.

5. Konklusioner

Denne undersøgelse haft et dobbelt sigte: Dels specifikt at undersøge, om evalueringspraksis i danske naturfag er forenelig med det test-format som PISA 2006 Science-undersøgelsen betjener sig af – og dels mere generelt at bidrage til karakterisering af den nuværende evalueringspraksis med henblik på at kvalificere udviklingen af denne.

På det PISA-rettede forskningsspørgsmål ”*Er dansk evalueringspraksis i naturfag i rimelig overensstemmelse med PISA’s test-koncept?*” må konklusionen på denne undersøgelse være et temmelig klart JA!

1159 naturfagslærere i folkeskolens 8. og 9. klasser tegner således via deres spørgeskemasvar et billede af en evalueringspraksis, som i vid udstrækning anvender PISA-lignende evalueringsformater, fx er individuel og skriftlig testning det hyppigst anvendte organiserings- og evalueringsformat. På det overordnede niveau angiver i hvert fald 75 % af lærerne regelmæssigt at teste kompetencer svarende til PISA. Efter lærernes oplysninger inddrager den typiske lærer PISA-relevante kontekster i sine evalueringsopgaver - og der synes at være en god overensstemmelse mellem lærernes og PISA’s vægtning af opgaver med åbne, hhv. lukkede svar-formater. Endelig anser lærerne det for rimelig vigtigt, at eleverne anvender korrekt fagsprog i opgaverne. Muligvis ville en endnu større vægtning af dette aspekt være i overensstemmelse med de ganske rigide krav til fagsprogsbrug, som PISA lægger til grund for sin scoring.

Når dette er sagt, er det selvfølgelig relevant også at fremdrage de begrænsninger, som nærværende undersøgelse har. Først og fremmest baserer den sig på lærernes selv-rapporterede evalueringsadfærd – ikke på observationer af den faktiske adfærd. Her ville en form for triangulering af resultaterne have været nyttig. Indholdsmæssigt er undersøgelsen begrænset af, at vi ikke har kunnet bede lærerne kommentere på overensstemmelsen (spørgsmålstyper, kompetencer og kontekstanvendelse) mellem deres egne evalueringsopgaver og det helt konkrete PISA 2006-sæt af opgaver. Her har PISA-konsortiets behov for at genbruge/hemmeligholde størstedelen af de brugbare opgaver stået i vejen for et mere dybdegående check af overensstemmelsen. Det er tankevækkende – og muligvis symptomatisk - at lærerne i undersøgelsen kun udtrykker moderat tiltro til at have ”klædt eleverne på” til en individuel, skriftlig test som PISA. Hvis man som lærer kun har kendskab til de offentliggjorte, overfladiske træk ved PISA, kan det være svært at tro på, at man har forberedt sine elever på bedste vis.

Undersøgelsen giver ikke meget belæg for den indledende hypotese om, at danske elever underpræsterer i PISA, som følge af manglende fortrolighed med PISA's testformat. Omvendt er der heller ikke meget i undersøgelsen, som antyder, at præstationen er blæst op via "teaching for the test". Meyerhöfer (2007) diskuterer studier af "test-wiseness" og "Testfähigkeit", som påviser hvorledes træning og øget metakognitiv opmærksomhed på teststrategier forbedrer elevernes evne til at score i tests. Man kan diskutere, om sådanne tiltag åbner op for, at elevernes formåen kommer til et reelt og dækkende udtryk, eller om det i bund og grund bidrager til, at denne misrepræsenteres. Givet er det, at sådanne tilsyneladende "forbedringer" ikke afspejler ændringer i elevernes faglige formåen. I undersøgelsen her har vi ikke gjort noget forsøg på at kortlægge brugen af de metakognitive strategier, som Meyerhöfer fremdrager – men vores lærerinterviews afdækkede ingen tendens til eksplicitering af sådanne.

Ét er, at danske elever efter alt at dømmes er rimeligt fortrolige med PISA-lignende evaluering. Noget andet er, at evalueringspraksis i naturfagene faktisk synes mere righoldig end PISA-konceptet, idet både formative og praktiske aspekter kommer klart til syne i den indsamlede empiri.

H. H. Knoop stiller i en PISA-høringsammenhæng spørgsmålet "*Er PISA i stand til at måle folkeskolens styrker?*" (Knoop 2005). I denne sammenhæng er det væsentligt at pointere, at dette spørgsmål kun besvares *delvist* af undersøgelsen her. Spørgsmålet omfatter både aspekter af, *hvad* PISA måler og *hvordan* dette måles – og forudsætter, at dette sammenholdes med, *hvad* der er folkeskolens styrke og *hvordan* disse måles i folkeskolens hverdag. I sin egen diskussion forholder Knoop sig udelukkende til det første aspekt (p. 162), ligesom VAP-rapport 1 i øvrigt gjorde det. Nærværende undersøgelse forholder sig derimod udelukkende til overensstemmelse mht., *hvordan* evt. elevstyrker evalueres.

Karakteriseringen af den bredere evalueringspraksis i naturfag

For det første er det blevet meget tydeligt gennem projektets litteraturstudier, at der aktuelt ikke findes helt fyldestgørende undersøgelser af evalueringspraksis i naturfagene i Danmark. Nærværende undersøgelse må anses for et skridt i den rigtige retning, om end den har de begrænsninger, som en kvantitativ undersøgelse baseret på læreres selvrapport nu engang må have. Forhåbentlig vil det kommende strategiske forskningsprogram for uddannelsesforskning give mulighed for en uddybende og kvalitativ forskning, som også rykker indenfor i naturfagsundervisningen og tættere på den *implementerede* evalueringspraksis.

Vi har fundet, at den typiske naturfagslærer har et vist repertoire af aktive evalueringsformer - med individuel skriftlig testning og uforpligtende samtale med

klassen som de dominerende former. Næsten alle lærere tilkendegiver at have et elevorienteret sigte med deres evaluering, men samtidig foretager knap halvdelen evaluering med deres undervisning som begrundelse. Der er et udtalt fokus på læring og læringspotentialer (93 %), når der evalueres. Fire ud af fem lærere udtrykker en summativ tænkning omkring deres evaluering, mens tre ud af fem anfører en formativ tænkning. Undersøgelsen viser også, at de fleste lærere evaluerer løbende og efter det enkelte forløb. En meget stor del af naturfagslærerne er således i overensstemmelse med de intentioner og den generelle rammesætning, som er formuleret i Folkeskolelovens §13. Hensigtsmæssigheden af den formative evaluering styrkes af, at lærerne ofte indvier eleverne i kriterierne for evaluering, at eleverne så godt som altid får feedback i kølvandet på evaluering – og at feedback hyppigst gives i form af mundtlige eller skriftlige kommentarer.

Undersøgelsen af evt. mønstre i lærernes evalueringspraksis har afdækket, at denne ikke er signifikant præget af deres linjefagsbaggrund, og at den kun i ringe grad er kønsspecifik. Blandt flere andre tolkninger, kan dette isoleret forstås inden for rammerne af en fælles naturfaglig evalueringsskiltur. Imidlertid synes evalueringsskiltur at være forskellig for mere eller mindre erfarne naturfagslærere – eller uskelneligt herfra: for lærere repræsenterende forskellige tiders lærer-socialisering. Dette svækker indtrykket af en fælles og overgribende naturfaglig evalueringsskiltur. Direkte adspurgt svarer 2/3 af lærerne ”nej” til spørgsmålet om, at der foreligger en ”fælles holdning til evaluering” på deres skole. Desværre har undersøgelsens datagrundlag ikke gjort det muligt at forfølge dette i dybden, men det antyder i hvert fald, at en evt. evalueringsskiltur i mindre grad skal forstås lokalt end nationalt.

6. Litteratur

- Andersen, N. O., Busch, H., Horst, S., Andersen, A. M., Dalgaard, I., Dragsted, S., and Norrild, P. (2006). *Fremtidens Naturfag i Folkeskole - rapport fra udvalget til forberedelse af en handlingsplan for naturfagene i folkeskolen*. København: Undervisningsministeriet.
- Black, P. (1998). *Testing - Friend or Foe*. London: Falmer Press.
- Britton, E.D. and Schneider, S.A. (2007). Large-Scale Assessments in Science Education. In Abell, S. and Lederman, N. (eds.), *Handbook of Research on Science Education*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers, p. 1007-1040.
- Broch, T. and N. Egelund (2002). *Et lærerperspektiv på natur/teknik- og fysik/kemiundervisningen – en kvalitativ analyse*. København: Danmarks Pædagogiske Universitet.
- Dahler-Larsen, P. (2001). Anvendelse af evaluering - historien om et begreb, der udvider sig. In Dahler-Larsen, P. and Krogstrup, H.K. (eds.), *Tendenser i evaluering*. Odense: Odense Universitetsforlag, p. 211-220.
- Dahler-Larsen, P. (2006). *Evalueringskultur. Et begreb bliver til*. Odense: Syddansk Universitetsforlag.
- Dolin, J., H. Busch, et al. (2006). En sammenlignende analyse af PISA 2006 science testens grundlag og de danske målkategorier i naturfagene. *Rapport nr. 1 for projektet Validering Af PISA science*. Odense: Center for Naturvidenskabernes og Matematikkens Didaktik, Syddansk Universitet.
- Egelund, N. (2002). *Fysik/kemilæreres syn på deres undervisningssituation i danske skoler - en kvantitativ undersøgelse*. København: Danmarks Pædagogiske Universitet.
- EVA, Danmarks Evalueringsinstitut (2004): *"Løbende evaluering af elevernes udbytte af undervisningen i folkeskolen"*, København: Danmarks Evalueringsinstitut.
- Gitomer, G. & Duschl, R. (1998). Emerging Issues and Practices in Science Assessment. In B. Fraser and K. Tobin, (Eds.), *International Handbook of Science Education*. pp. 791-810. Dordrecht: Kluwer Academic Publisher.
- Hargreaves, D., El-Khawas, E., and Ekholm, M. (2004). *National Review on Educational R&D: Examiners Report on Denmark*. EDU/CERI/CD(2004)10. OECD.
- Harlen, W. and Deakin, R. (2002). A systematic review of the impact of summative assessment and tests on students' motivation for learning. *Research Evidence in Education Library*. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

- Harmon, M. et al (1997). *Performance Assessment in IEA's Third International Mathematics and Science Study*. Chestnut Hill: Center for the Study of Testing, Evaluation, and Educational Policy, Boston College.
- Jenkins, D. (1976). *Open University Course E203, Curriculum Design and Development. Unit 19: Curriculum Evaluation*. Milton Keynes: The Open University.
- Johansen, J. and Langager, S. (2002). *Andre mål, nye evalueringsveje - forandringer til skolen, udfordringer for eleverne*. København: Danmarks Pædagogiske Universitet.
- KL, Finansministeriet, Undervisningsministeriet (2006). *Rapport fra arbejdsgruppen om efteruddannelse af lærere og skoleledere*. København: Undervisningsministeriet.
- Knoop, H.H. (2005). PISA-undersøgelsen og Folkeskolens ”målsætning” In ”PISA-undersøgelsen og det danske uddannelsessystem - Resume og redigeret udskrift af høring d. 12. september 2005. København: Teknologirådet, p.56-62.
- Lawrenz, F. (2007). Review of Science Education Program Evaluation. In Abell, S. and Lederman, N. (eds.), *Handbook of Research on Science Education*. Mahwah, New Jersey: Lawrence Erlbaum Associates, p. 943-963.
- Meyerhöfer, W. (2007). Testfähigkeit – was ist das? In Hopmann, S.; Brinek, G. And Retzl, M. (eds.), *PISA zufolge PISA - PISA According to PISA*. Wien: LIT VERLAG GmbH & Co. KG, p. 57-92.
- Millar, R. (2005). Evaluating educational programmes: Issues and perspectives. In Bennett, J., Holman, J., Millar, R., and Waddington, D. (eds.), *Evaluation as a tool for improving science education*. New York: Waxmann Publishing Co., p. 15-32.
- Nichols, S., Glass, G., and Berliner, D. (2005). High Stakes Testing and Student Achievement: Problems for the No Child Left Behind Act. Tempe, AZ: Educational Policy Research Unit, *EPSL-0509-105-EPRU*.
- Orpwood, G. (2001). The Role of Assessment in Science Curriculum Reform. *Assessment in Education* 8:135-151.
- Rodriguez, M.C. (2004). The Role of Classroom Assessment in Student Performance on TIMMS. *Applied Measurement in Education* 17:1-24.
- Stiggins, R.J. and Bridgeford, N.J. (1985). The ecology of Classroom Assessment. *Journal of Educational Measurement* 22:271-286.
- Waddington, D., P. Nentwig, et al., Eds. (2007). *Standards in Science Education*, Waxmann Verlag.
- Weng, P. and Hoff, A. (1999). *Evaluering i matematik og naturvidenskabelige fag i folkeskolen*. København: Danmarks Pædagogiske Universitet.

7. English summary

This is the second report from the VAP-project (Validation of PISA Science). The overall aim of the project is to validate the PISA 2006 science test within a Danish context. The project set-up appears in the diagram on page 52(a ~ the concept of literacy (in PISA and elsewhere); b ~ the science goals in the Danish educational system; c ~ the PISA test (knowledge of and about science, format etc.) and the Danish test results; d ~ the Danish evaluation system in science; e ~ the VAP test).

The first step in the VAP-project was to compare the PISA Science Framework (OECD 2007) with the general understanding of scientific literacy within the science education community and with the general aims of Danish science education in the compulsory school (the Common Goals). The first VAP report gave the results of this comparison (Relation 1 in the diagram).

The next step will examine the question of whether the PISA science test (i.e. the items) actually measures what it intends to (according to the Framework) (Relation 2), and to what extent the test corresponds to Danish science educational goals (Relation 3). Report No. 3 will deal with these issues.

Relation 4 considers the PISA test format in relation to the assessment culture in the Danish compulsory school. Interviews with science teachers and a nationwide survey provides a basis for assessing the importance of the differences between the PISA testing system and everyday assessment and test practice in science education in Denmark. This second report gives the results of this comparison.

These three studies are the foundation of the real validation question: Are the PISA science test results an adequate expression of the students' competencies in science? One-hundred and twenty students were selected among those who took the PISA 2006 test. Some weeks later these same students were subjected to a specially designed VAP-test. Specially educated assistants (student teachers) assessed the students (one at a time) with regard to two selected items, which they had been exposed to in the PISA test. The conversations involved relevant artefacts, took 30 minutes, and followed a rather tight scheme for reasons of reliability. All of the talks were videotaped and scored afterwards. The results from this oral, socio-culturally oriented assessment are compared with the paper-and-pencil PISA test (Relation 6). In addition to assessing the students' knowledge and competencies related to the PISA item content, the conversations also included the subject specific knowledge demanded by the Danish

curriculum, while still within the item domain. Finally, the students were paired together to do a practical lab activity (carrying into practice one of the PISA items), which was videotaped and scored. (Relation 7).

Before we present a summary of this second report, we will review the results from the first report.

The results of the comparative analysis

The PISA Science Framework and the Danish course of study (Common Goals) are two very different systems of describing desired outcomes from science education. They emphasize different aspects and they use different terminologies - which complicates comparison. But the report concludes:

- *The literacy definition* in the PISA 2006 Science Framework is quite different from the literacy definition in 2003, bringing the 2006 definition in accordance with the internationally used terminology. The items are, apart from the attitudinal questions, completely alike in the 2003 and the 2006 tests.
- *The contexts* for assessment items in the Framework align well with the recommended contexts in Common Goals.
- The PISA *concept of competency* is **not** in accordance with the way competency is used in the Danish educational system.
- The correspondence within *the affective domain* is **not** very good. Both weigh interest in science, but the Common Goals do not mention support for scientific inquiry (which is a part of the PISA attitudinal area) but instead give high priority to responsibility towards resources and environments (which is a part of the PISA attitudinal area in the Framework, but not in the items).
- *The knowledge of science* categories have a different weight in the PISA science test compared to the Danish compulsory school. PISA gives more importance to Biology and Geoscience, while the Danish school gives higher weight to Physics and Chemistry.
- The alignment in *content* varies from subject to subject. For Biology the correspondence is nearly perfect, in the Geosciences the correspondence is very high, but in Physics/Chemistry it is very difficult to compare due to different description formats.
- The PISA *knowledge about science* has a relatively weak position in the Danish curriculum. Aspects of knowledge about science are part of the described working methods and ways of thinking in the Common Goals for different subjects, but they do not form a systematic, self-contained system.

- *The practical dimension*, which is a major part of the Common Goals, is not included in the PISA Framework, although the theoretical aspects are.
- *In general*, the Danish course of study is stated in broader, more general terms than the PISA Framework. Phrases like ‘emphasis on understanding of connections’, ‘the essential points’, ‘involve the perspectives for’ etc. are quite widespread in the Common Goals. This promotes teaching with emphasis on processes, overall views, problem orientation etc. where the teachers adapt the Common Goal curriculum to the concrete class and students. The result is a more heterogeneous science teaching, without concern for standards.

The assessment culture in the Danish lower secondary school

The aim of this study was twofold:

1. to investigate whether assessment practice in science subjects in Danish grades 8 and 9 (year 14-16) is in accordance with the test format used in the PISA 2006 Science test, and
2. to characterise the present assessment practice in science subjects in order to give a basis for future improvement.

Based on two focus group interviews with science teachers and a nationwide, representative survey involving 1159 teachers teaching science in years 8 and 9 in 2005/2006, the short answer to the first question is YES. The assessment practice in science in the Danish lower secondary school is to a large degree in accordance with the PISA test format. Individual, written tests are the most commonly used test format in science subjects and 75 % of the teachers indicate that such tests are used regularly to test the PISA competencies. Teachers typically include PISA relevant contexts in their assessments and the teachers seem to weigh the balance between open/closed answer formats in accordance with the PISA weighting. Danish science teachers also give moderate emphasis to their students’ use of correct technical language – though a heavier weight on this aspect would be more in accordance with the relatively rigid demands on this in the PISA scoring handbook.

In the light of these results, it is thought-provoking that the teachers only express moderate confidence in having prepared the students adequately for the PISA test. An explanation could be the lack of evidence in our study for any ‘teaching to the test’ or rehearsal of PISA-like test items (which the concealment of most of the PISA items makes impossible). The teachers do not teach in order to enhance a PISA performance. They teach for more varied goals and they simply use a much broader palette of assessment tools than the PISA formats – things this study clearly shows. PISA has had relatively little impact on science teach-

ing in Denmark - at least in 2005/2006, a situation that seems to be changing thanks to the increasing attention politicians are giving to PISA rankings.

We have found that the typical science teacher holds a broad repertoire of active assessment forms, with a combination of individual written tests and not-grade orientated talks with the whole class as the dominant forms.

Most of the teachers expressed a student oriented aim with their assessments. They have a marked focus on learning and the learning potential of their assessments as well as with formative and summative approaches, and they evaluate both during and after teaching sequences. They often make the criteria for an assessment open to the students, and the students always receive feedback on the tests and assessments.

The study did not uncover any patterns in assessment and evaluation practice due to gender or science subject. Two thirds of the teachers answered “no” as to whether the school has “a common attitude toward assessment” which indicates that an assessment culture, if any, is rather a national phenomenon than a local one.

Bilag 1

Validering af PISA Science

VAP-projektet forestås af lektor, ph.d., Jens Dolin, Institut for Naturfagernes Didaktik, Københavns Universitet, lektor Henrik Busch, Københavns Universitet, og lektor, ph.d. Lars Brian Krogh, Institut for Videnskabsstudier, Aarhus Universitet.

Det overordnede formål med VAP-projektet er at placere PISA 2006 science undersøgelsen i en dansk kontekst. Som en forudsætning herfor undersøges PISA 2006 science dels på sine egne præmisser og dels i forhold til de danske mål for naturfagene. De forskellige elementer i en sådan undersøgelse og deres indbyrdes sammenhænge er illustreret i diagrammet. Gennem projektet svares på nedenstående spørgsmål (hvor tal og bogstaver i parentes refererer til diagrammet).

Hvad vil PISA måle sammenlignet med andre literacymål og med de danske mål for naturfagene?

Som svar på første delspørgsmål undersøges PISAs literacybegreb (a) og sammenlignes med andre internationalt anvendte literacybegreber.

1. I andet delspørgsmål sammenlignes PISAs literacybegreb (a) med kompetence-/dannelsesmål i det danske uddannelsessystem som de er formuleret i Fælles Mål (b). En sådan analyse vil kunne perspektivere PISAs ”fit for life”-opfattelse med det danske dannelses-/kompetencebegreb. Metodisk vil der være tale om tekstanalyse. Dette spørgsmål behandles i *Delrapport 1*.

Måler testen det den vil, og hvorledes svarer den til kravene i Fælles Mål?

2. Her analyseres PISA-test setup’et (c) i forhold til det opstillede literacy begreb (a). Denne interne validitet siger noget om testens begrænsninger på testens egne præmisser. Også her vil metoden være tekstanalyse.

3. Ved at sammenholde PISA-testen (c) med de samlede mål, der stilles for de danske uddannelser (b) kan man afgrænse hvilke aspekter af de danske naturfagsuddannelser PISA-testen reelt kan udsige noget om, og hvilke de ikke kan. Igen tekstanalyse. Disse to spørgsmål behandles i *Delrapport 3*.

Er målemetoden rimelig ift. den gængse evaluering i det danske uddannelsessystem?

4. Her må test-setup'et (c) sammenlignes med de måder, man traditionelt evaluerer på i danske skoler (d). Dette vil kunne forklare resultater grundet i forskellige testkulturer. En sådan analyse forudsætter et kendskab til den danske evalueringskultur i naturfagene, som kun kan tilvejebringes via en empirisk undersøgelse af (d).

5. Et interessant aspekt er at sammenholde den eksisterende evalueringskultur med de officielt opstillede mål. Tester vi i Danmark det, vi er interesseret i at eleverne skal kunne? Dette vil også kunne sige noget om hvorvidt danske evalueringsresultater er et rimeligt udtryk for opfyldelsesgrad af danske mål. Relation 5 ligger uden for VAP-projektet, hvorimod relation 4 behandles i *Delrapport 2*.

Er PISA-resultaterne udtryk for det eleverne kan inden for det testede område?

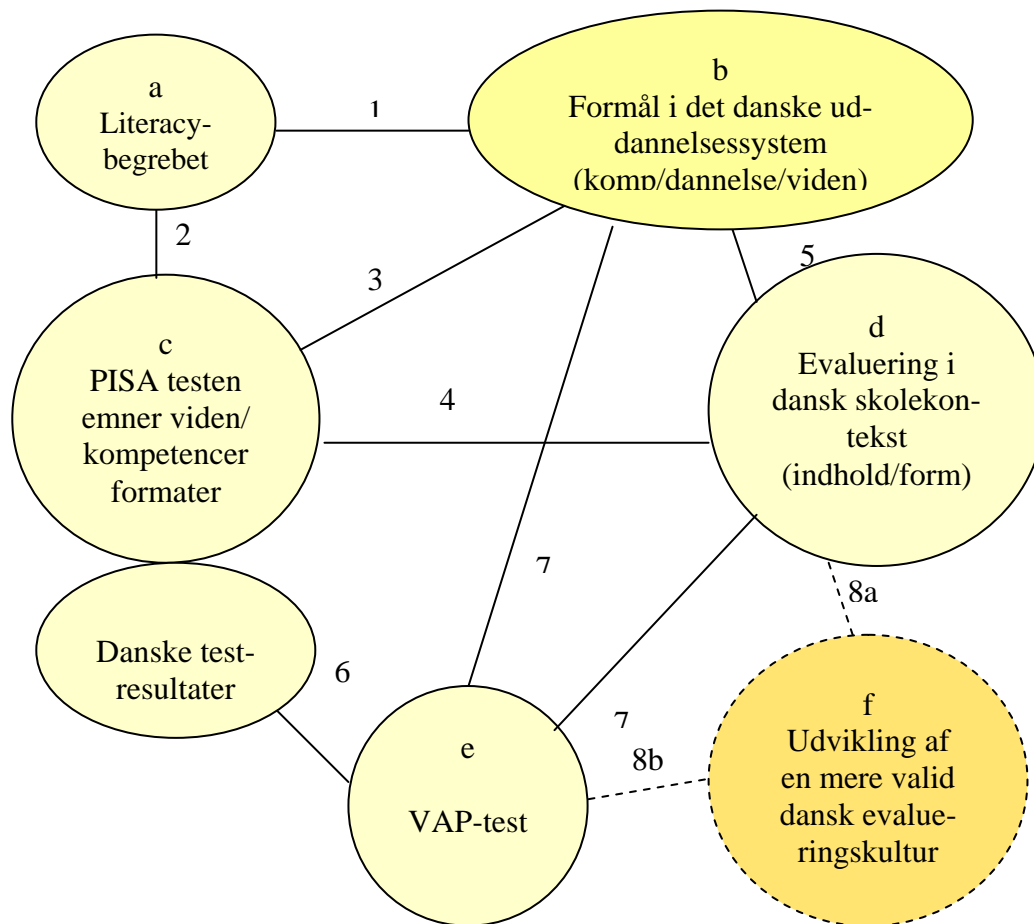
6. Dette er en "traditionel" validering af testen, og et af forskningsprojektets centrale spørgsmål. Når test-tekniske forhold filtreres fra, kan spørgsmålet kun besvares ved en feltundersøgelse, dvs. empirisk. Dette gøres ved at ud-sætte de PISA-testede elever for en mere omfattende evaluering (e). Denne VAP-evaluering består af en række faser (fx samtale), som i højere grad anses for at evaluere det (i PISA) ønskede.

7. Ved samme evaluering (e) undersøges, i hvilket omfang eleverne opfylder de mål, som ikke indgår i PISA-testen (fx praktiske færdigheder). Disse spørgsmål behandles i *Delrapport 4*.

Udvikling af en mere valid evalueringskultur i de naturfaglige uddannelser

8. Dette kunne være et spin-off af projektet, som ville kræve en selvstændig indsats. Hvorledes kan en sammenstilling af det, som forskningsprojektet viser at eleverne rent faktisk ikke kan ifølge PISA (6), og som den danske evalueringskultur ikke indfanger (7), bruges til udvikling af en evalueringsform og -kultur (f), som i højere grad end den nuværende evaluerer de opstillede formål (b) med uddannelserne?

Diagram over elementer og sammenhænge i VAP



Bilag 2 Lærerspørgeskema om evaluering i naturfag

LÆRERSPØRGESKEMA OM EVALUERING I NATURFAG

Dette spørgeskema henvender sig til lærere, som i skoleåret 2005/2006 underviser i fysik/kemi, biologi og/eller geografi i 8. eller 9. klasse.

Forskere fra Syddansk Universitet, Danmarks Pædagogiske Universitet og Århus Universitet er i gang med at undersøge PISA-resultaternes gyldighed for danske forhold. I den forbindelse vil vi bede dig om at besvare nedenstående spørgsmål vedrørende din egen måde at evaluere din undervisning i naturfag på. Svarene skal altså kun være baseret på din undervisning i de tre naturfag.

Hovedparten af skemaet udfyldes med markering i relevante felter, og besvarelsen vil tage ca. 5-10 minutter.

På forhånd tak
Jens Dolin

Baggrundsspørgsmål:

Din skoles postnummer:

Køn
 mand kvinde

Alder
 20-29 30-39 40-49 50-59 60->

Hvor mange år har du undervist i folkeskolen?
 0-4 år 5-9 år 10-14 år 15-> år

	Fysik/kemi	Biologi	Geografi	Matematik
Jeg har linjefag i følgende fag	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Jeg underviser normalt i følgende fag	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Jeg identificerer mig mest med følgende fag	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Hvilke emner har klassen arbejdet med eller skal arbejde med i dette skoleår i følgende fag?

fysik/kemi	8. klasse	<input type="text"/>
fysik/kemi	9. klasse	<input type="text"/>
biologi	8. klasse	<input type="text"/>
biologi	9. klasse	<input type="text"/>
geografi	8. klasse	<input type="text"/>
geografi	9. klasse	<input type="text"/>

Begrundelser for evaluering

Hvad er din vigtigste begrundelse for at evaluere? (svar med 1-2 korte sætninger)

Forestil dig, at du har gennemført et undervisningsforløb med en af dine klasser. Hvor meget vægter du typisk følgende grunde til at evaluere et bestemt undervisningsforløb?

	meget højt	højt	middel	lav	meget lavt
Jeg bliver klogere på hvad den enkelte elev har lært i undervisningen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Jeg får indblik i elevernes oplevelse af undervisningen (var den interessant, svær ...)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Jeg hjælper eleverne med at blive opmærksomme på, hvad de har lært	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Jeg kan bruge evalueringen til at forbedre det næste undervisningsforløb med klassen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Anden grund: <input type="text"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Anden grund: <input type="text"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Evalueringsform

Beskriv kort hvordan du hyppigst evaluerer din undervisning?:

I de enkelte undervisningsforløb evaluerer jeg løbende eleverne, så de hver især bliver klar over deres standpunkt:

ja, flere gange gennem forløbet
 ja, enkelte gange i forløbet
 nej, ikke i selve forløbet

Jeg evaluerer elevernes viden efter afslutningen af et undervisningsforløb:

ja, altid
 ja, som oftest
 sjældent

Ved "test" forstår vi i det følgende evaluering af elevernes viden, hvor resultatet kan angives med et tal.

Hvor ofte benytter du følgende evalueringsformer i din undervisning?

	hyppigt	af og til	sjældent	aldrig
Individuel skriftlig test eller opgave	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Gruppebaseret skriftlig test eller opgave	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Individuel mundtlig test eller opgave	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Gruppebaseret mundtlig test eller opgave	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Individuel og uforpligtende samtale med elev	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Uforpligtende samtale med elevgruppe/klassen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Hvis du opdeler al den tid, du bruger på evaluering, på følgende kategorier, prøv så at angive hvor stor en andel af den samlede evalueringstid, de forskellige former optager:

	Andel
Individuel skriftlig test eller opgave	<input type="text"/> %
Gruppebaseret skriftlig test eller opgave	<input type="text"/> %
Individuel mundtlig test eller opgave	<input type="text"/> %
Gruppebaseret mundtlig test eller opgave	<input type="text"/> %
Individuel og uforpligtende samtale med elev	<input type="text"/> %

Uforpligtende samtale med elevgruppe/klassen	<input type="text"/> %
Andre	<input type="text"/> %
Ialt	100%

Hvor mange gange giver du i gennemsnit dine elever i hver af dine naturfagsklasser skriftlige, individuelle test i løbet af et skoleår?

<input type="checkbox"/> 0 gange	<input type="checkbox"/> 1 gang	<input type="checkbox"/> 2 gange	<input type="checkbox"/> 3 gange	<input type="checkbox"/> 4 gange	<input type="checkbox"/> 5 gange	<input type="checkbox"/> mere end 5 gange
----------------------------------	---------------------------------	----------------------------------	----------------------------------	----------------------------------	----------------------------------	---

Bliver eleverne gjort bekendte med de anvendte evalueringskriterier for skriftlige tests (lige meget tid på alle opgaver, giv et svar selv om du er i tvivl etc.)?

<input type="checkbox"/> ja, meget	<input type="checkbox"/> ja, noget	<input type="checkbox"/> nej, kun ganske lidt	<input type="checkbox"/> nej, slet ikke
------------------------------------	------------------------------------	---	---

I hvilket omfang føler du at dine elever gennem din undervisning er blevet klædt på til at klare en to-timers individuel, skriftlig test som fx PISA?

<input type="checkbox"/> rigtig meget	<input type="checkbox"/> meget	<input type="checkbox"/> lidt	<input type="checkbox"/> næsten ikke	<input type="checkbox"/> slet ikke
---------------------------------------	--------------------------------	-------------------------------	--------------------------------------	------------------------------------

Uddybende kommentar

Får eleverne feedback på deres evalueringer?

<input type="checkbox"/> ja, altid	<input type="checkbox"/> ja, hyppigt	<input type="checkbox"/> ja, af og til	<input type="checkbox"/> nej, aldrig	<input type="checkbox"/> ved ikke
------------------------------------	--------------------------------------	--	--------------------------------------	-----------------------------------

Hvor ofte bruger du følgende feed-back former?

	hyppigt	af og til	sjældent	aldrig
Skriftlige kommentarer	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Mundtlige kommentarer	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Karakterer	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Andet: <input type="text"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Indhold i evaluering af eleverne

Hvor ofte indgår følgende <i>anvendelsessammenhæng</i> af naturvidenskab i dine evalueringer?				
	hyppigt	af og til	sjældent	aldrig
Personlige og hverdagsmæssige	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Samfundsmæssige	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Globale spørgsmål	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Hvor ofte indgår følgende naturvidenskabelige kompetencer (færdigheder) i dine evalueringer?				
	hyppigt	af og til	sjældent	aldrig
At kunne identificere naturvidenskabelige spørgsmål F.eks. at skulle - genkende hvilke spørgsmål man kan undersøge videnskabeligt, kunne - identificere naturvidenskabelige nøgleord som hjælp til informationssøgning og kunne - genkende centrale træk ved en naturvidenskabelig undersøgelse	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
At kunne anvende naturvidenskabelig viden F.eks. ved at - beskrive og forklare naturvidenskabelige fænomener og forudsige ændringer, - identificere passende beskrivelser, forklaringer og forudsigelser eller - anvende naturvidenskabelig viden og viden om naturvidenskab i givne situationer	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
At kunne bruge naturvidenskabelige beviser F.eks. ved at - tolke naturvidenskabeligt bevismateriale og drage konklusioner, - argumentere for eller imod konklusioner og identificere forudsætninger for givne konklusioner eller - kommunikere konklusioner og de beviser og ræsonnementer de er baseret på.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Hvor meget indgår følgende indhold i dine evalueringer af eleverne?:				
	meget	noget	lidt	slet ikke
Paratviden	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Standardopgaver (fx fra lærebog)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Åbne spørgsmål	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Holdninger	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Andet: <input type="text"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Andet: <input type="text"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Hvor meget fylder den praktiske dimension i den samlede evaluering af eleverne?

<input type="checkbox"/> meget	<input type="checkbox"/> noget	<input type="checkbox"/> lidt	<input type="checkbox"/> ingenting
--------------------------------	--------------------------------	-------------------------------	------------------------------------

Hvor vigtigt er det, at eleverne i deres besvarelse anvender korrekt fagsprog?

<input type="checkbox"/> meget vigtigt	<input type="checkbox"/> vigtigt	<input type="checkbox"/> ikke så vigtigt	<input type="checkbox"/> ligegyldigt
--	----------------------------------	--	--------------------------------------

Skolekultur

Har skolen en fælles holdning til evaluering?

<input type="checkbox"/> ja	<input type="checkbox"/> nej	<input type="checkbox"/> ved ikke
-----------------------------	------------------------------	-----------------------------------

Hvis ja, hvorledes kommer det til udtryk:

Når du er færdig med at udfylde spørgeskemaet, så tryk på

INDs skriftserie

- Nr. 1: Kandidater i matematik-, fysik- og kemifagene fra Københavns Universitet – De gik videre Produktion og beskæftigelse 1985-1999 (2002)
- Nr. 2: Bachelorer – ej blot til pynt (2002)
- Nr. 3: Studieårgangene 1999-2000 på geografi (2002)
- Nr. 4: Faglige forskelle og tidlige tendenser Det Naturvidenskabelige Fakultet, 1985-2001 (2002)
- Nr. 5: Studieforløbsundersøgelser i naturvidenskab - en antologi (2003)
- Nr. 6: Kammeratlæring på førsteårskursus i mekanik (2003)
- Nr. 7: If reform of science education is the answer - what were the questions? (2003)
- Nr. 8: Gennemførelsesprocenter på kandidatuddannelser på Det Naturvidenskabelige Fakultet (2003)
- Nr. 9: Universitetsstuderendes tilgang til eksperimentelt udstyr – kønslige og faglige aspekter (2003)
- Nr. 10: De studerendes motivation bag valget af speciale (2003)
- Nr. 11: Om studiestarterne 2002 på nanoteknologistudiet på Københavns Universitet (2004)
- Nr. 12: Kandidater i biologi, geografi, geologi og idræt fra Københavns Universitet – De gik også videre. Produktion og beskæftigelse, 1985 – 2002 (2005)
- Nr. 13: De studerendes udbytte af kollokvier og forelæsninger på faget jordbundsbiologi (2005)
- Nr. 14: Projekt Rum til Bedre Læring (2005)
- Nr. 15: Didactics of Mathematics – the French Way. Texts from a Nordic Ph.D.-Course at the University of Copenhagen (2005)
- Nr. 16: Evaluering af Matematik C på stx og htx – erfaringer fra det første år efter gymnasireformen (2007)
- Nr. 17: Den naturfaglige evalueringskultur i folkeskolen (2008)**

Dette er den anden rapport i forskningsprojektet Validering af PISA Science. Det overordnede formål med projektet er at placere PISA 2006 scienceundersøgelsen i en dansk kontekst.

Formålet med denne rapport er at vurdere, hvorvidt PISAs testformat og tilgang til testning svarer til den evalueringskultur, der er fremherskende i naturfagene i de ældste klasser i den danske folkeskole.

Rapporten indledes med en kortfattet oversigt over nogle centrale udviklingstræk i evalueringsfeltet – op til og under indflydelse af PISA. Selve begrebet evalueringskultur gøres her til genstand for overvejelser.

Herefter følger en gennemgang af en række undersøgelser af evaluering i naturfagene i den danske folkeskole. Disse perspektiveres i forhold til udenlandske forskningsresultater.

Rapportens centrale afsnit udgøres af empiriske undersøgelser af, hvorledes danske naturfagslærere rent faktisk anvender evaluering i deres naturfagsundervisning i 8. og 9. klasse. Denne praksis relateres til PISA testformatet.



**INSTITUT FOR NATURFAGENES DIDAKTIK
KØBENHAVNS UNIVERSITET**