

Konsekvenser af evaluering i matematikundervisning

Kristine Jess

Københavns Dag- og Aftenseminarium

Diskussionen om test/evaluering i matematik i folkeskolen er blevet meget aktuel i Danmark, hvor vi er på vej ind i en evalueringskultur. Erfaringer fra udlandet og fra testningens historie viser imidlertid at vejen frem er fyldt med mulige faldgruber. Efter en begrebsafklaring ses der på testningens oprindelse og på negative virkninger af test/evaluering. Det påvises derefter at hensigtsmæssige evalueringsformer kan støtte reformer af matematikundervisningen på grundskoleniveau. Den vigtigste faktor i denne forbindelse er den ofte undervurderede tilbagevirkende effekt af test/evaluering der både kan have positive og negative konsekvenser for undervisningen.

Introduktion

Test og evaluering har ikke været særlig udbredt i grundskolens matematikundervisning i Danmark, men i fremtiden skal der testes/evalueres mere. Det fremgår af regeringsgrundlaget "Nye mål" for den nuværende VK-regering¹, og Undervisningsministeriet har i sommeren 2005 udbudt et evalueringsprojekt, "Fremme af evalueringskulturen i folkeskolen", til 30-35 mio. i licitation². Projektet angår fag i grundskolen – herunder matematik, hvor der skal testes efter 3. og 6. klasse. I september 2005 blev der indgået aftale³ mellem regeringspartierne, Socialdemokraterne og Dansk Folkeparti om at indføre obligatoriske test i udvalgte fag på bestemte klassetrin.

Der er i uddannelsespolitiske kredse store forventninger til den forøgede grad af testning/evaluering, og netop derfor er det nu væsentligt at skærpe opmærksomheden på forskningsresultater og erfaringer på feltet. Det er min opfattelse at valget af det faglige indhold der ønskes testet, og valget af test- og evalueringsformer har så stor

1 http://www.stm.dk/publikationer/reggrund05/index.htm#Fornyelse_af_folkeskolen_

2 <http://presse.uvm.dk/nyt/pm/pm290605.htm?menuid=641015>

3 <http://www.uvm.dk/05/aftaletekst.doc>

indflydelse på udviklingen af matematikundervisningen at emnet må søges belyst og diskuteret. Det er min hensigt at bidrage hertil.

For at få et indblik i hvordan test og evaluering har udviklet sig til de udbredte former som vi i dag ser, vil jeg først kaste et blik tilbage i tiden og se på oprindelsen til testning. Derefter vil jeg redegøre for mulige konsekvenser af evaluering, og i mangel af erfaringer fra danske forhold vil jeg gå tæt på den seneste udvikling især i USA for at belyse hvad man kan lære herfra. Jeg kommer ind på hvilke krav der fra forskelligt hold er stillet til evaluering, og fremhæver at evaluering kan bidrage til en positiv udvikling af matematikundervisningen. Men først er det nødvendigt med en præcisering af min anvendelse af begreberne test og evaluering.

Begreberne test og evaluering

I artiklen anvendes, som det allerede fremgår, begreberne “test” og “evaluering”. Jeg anvender evaluering som en overordnet betegnelse der kan antage mange forskellige former. En af formerne er kvantitative test hvor et fællestræk er at det matematiske indhold er opsplittet i korte opgaver som altid har et entydigt resultat. Man ser kun på om facit er korrekt, og opgaveløsningsprocessen tages ikke i betragtning. Formen kan derfor ikke evaluere hele det brede spektrum af kompetencer der indgår i en matematisk faglighed, som det f.eks. er beskrevet i “Kompetencer og Matematiklæring” (KOM-rapporten), (Niss & Jensen, 2002). Færdighedsprøven i grundskolen efter 9. klasse er et eksempel på en kvantitativ test.

Evaluering dækker foruden test de mange forskellige måder man kan evaluere tilegnelsen af matematik på, lige fra en skriftlig eksamen hvor der redegøres for løsningsprocesser, til logbog, portfolio og projektopgave. Evaluering kan hermed indeholde en kvalitativ vurdering. Det er oplagt at der er meget større chance for at yde eleven eller den studerende retfærdighed ved at undgå den snævre testform. På samme måde kan læreren få langt mere kvalificerede oplysninger når løsningsprocessen også inddrages i evalueringen. Via den afdækkes det at der alt efter problemstillingens karakter kan være flere veje til et rigtigt resultat, og at forskellige resultater baseret på velunderbygget argumentation kan accepteres. Den mundtlige afgangsprøve efter 9. klasse er et eksempel på en evalueringsform hvor løsnings- og argumentationsprocessen også inddrages.

Der skelnes almindeligvis mellem to kategorier af evaluering: en summativ og en formativ. Den summative evaluering anvendes ved afslutningen af et undervisningsforløb for at finde frem til i hvilken grad eleven har nået de mål der var opstillet for forløbet. Hensigten med denne er at fremskaffe en status for elevens matematiske formåen på det givne tidspunkt. Betegnelsen “summativ” peger på at det med denne form ikke er hensigten at give handleanvisninger for fremtidig undervisning men at evaluere den videnstilegnelse der har fundet sted. Begge de ovennævnte afgangsprøver er summative.

Med den formative form er det hensigten at afdække elevens læringspotentialer for dermed at finde det bedste afsæt for den fremtidige undervisning. Læreren forsøger at fastlægge hvilken viden eleven har tilegnet sig, og som der kan bygges videre på. Ideelt set tilrettelægger læreren den kommende undervisning i overensstemmelse med den indsigt der er opnået.

På trods af at der findes så mange forskellige evalueringsformer, indtager de kvantitative testformer en dominerende rolle i mange lande. Findes der en forklaring på at netop denne form er blevet så udbredt?

Testningens oprindelse

Allerede i 1905 blev der udviklet test i aritmetik til elever i 1. til 7. klasse af en franskmand ved navn Vaney. Sammen med kollegaen Binet anvendte han disse test til at identificere retarderede børn (Kilpatrick, 1993, s. 33). Med udviklingen af den såkaldte Binet-Simon-skala i 1905 og 1911 og med englænderen Spearmans test fra første fjerdedel af århundredet var intelligenstaget fastslået som et nyt instrument til psykologisk undersøgelse. Disse nye metoder vandt hurtigt indpas i USA, se f.eks. (Thomson, 1969) og (Torpe, 1972).

Samtidig opstod der ifølge Shepard i begyndelsen af 1900-tallet en effektivitetsbevægelse i USA der byggede på en opfattelse af at problemer vedrørende industriel udvikling og urbanisering kunne løses ved hjælp af passende videnskabelige metoder. I overensstemmelse hermed mente man at videnskabeligt baserede principper om ledelse – med henblik på at maksimere produktionen – kunne overføres med tilsvarende succes til skoler. Undgåelsen af spild indgik også i kravet om effektivitet, og derfor skulle det enkelte individ have en uddannelse der var i overensstemmelse med vedkommendes (mentale) kapacitet, og kun elever der blev fundet egnede til en videregående uddannelse, fik undervisning i akademiske fag i skolen. De nye testmetoder fandt anvendelse ved udvælgelsen (Shepard, 2000, s. 4-5).

Effektivitetstankerne kom også til at præge undervisningen med ideer om "teacher-proof"-læseplaner, som foreskriver undervisningsmetoder der skal følges præcist og omhyggeligt af læreren. I Danmark vandt disse tanker også indpas; det fremgår med al tydelighed af figur 1 hvor der er gengivet en side fra Friis-Petersen, Gehl & Jessens metodiske håndbog fra 1922: "Regneundervisning i det femte skoleår". Til hver time er der opstillet en detaljeret beskrivelse af hvad læreren skal gøre. Samtidig ses en opsplitning af stoffet der kan danne grundlag for kvantitativ testning.

Denne opfattelse af undervisning er tydeligt præget af den behavioristiske teori (Thomson, 1969 s. 143). Behaviorismen har haft en række konsekvenser for måden at tilrettelægge undervisning og testning på: (i) Læring antages at ske gennem akkumulering af små dele af viden, (ii) testning skal foretages hyppigt for at sikre at stoffet

45. Time. (Tabel 5- og 15-Stykket).

Multiplikation med 10 og 100. Ved Timens Begyndelse skriver Læreren en Række Opgaver op paa Skoletavlen til Indøvelse af Multiplikation med 10 og 100 ved Flytning af Kommaet, f. Eks.

10 · 2,75 Kr. 10 · 0,06 m; 10 · 1,73 hl; 10 · 2,507 kg
100 · 8,25 Kr.; 100 · 2,365 m; 100 · 0,275 kg.

Hovedregning.

1) $3 \cdot 0,2 + 2 \cdot 0,3$; $5 \cdot 1,5 - 2 \cdot 0,7$; $10 \cdot 0,5 - 5 \cdot 0,5$
 $5 \cdot 0,4 + 3 \cdot 0,4$; $6 \cdot 1,5 - 7 \cdot 0,2$; $10 \cdot 1,5 - 4 \cdot 1,5$

2) $8 \cdot 0,15$ Kr. — $4 \cdot 0,15$ Kr.; $3 \cdot 0,15 + 5 \cdot 0,15 - 7 \cdot 0,15$
 $5 \cdot 1,15$ Kr. — $2 \cdot 1,16$ Kr.; $4 \cdot 0,03 + 5 \cdot 0,04 - 2 \cdot 0,05$

3) Omsæt til Meter: 8 m 3 dm 6 cm; 1 m 7 dm 5 cm;
5 m 3 dm; 3 dm 7 cm; 10 m 5 cm; 20 m 5 dm.

Tavleregning.

I Nr. 4—7 regnes der med Benævnelsen m.
Nr. 4 regnes paa Skoletavlen.

4) 15 · 3 m 6 dm 7 cm	8) 65 · 9,27 Kr. + 5 · 0,69 Kr.
5) 15 · 5 m 6 dm 7 cm	9) 74 · 9,49 Kr. + 6 · 0,79 Kr.
6) 14 · 14 m 4 dm 3 cm	10) 84 · 9,55 Kr. + 4 · 1,45 Kr.
7) 12 · 33 m 6 dm 7 cm	11) 97 · 9,30 Kr. + 6 · 1,15 Kr.

Resultater: 4) 55,05 m 5) 85,05 m 6) 202,02 m 7) 404,04 m
8) 606 Kr. 9) 707 Kr 10) 808 Kr. 11) 909 Kr.

Figur 1. En side fra Friis-Petersen, Gehl & Jessens metodiske håndbog fra 1922: "Regneundervisning i det femte skoleår".

beherskes før man går videre til nyt stof, og (iii) der udvikles en tendens til at man kun lærer det der testes i (Shepard, 2000, s. 5).

Testning og matematik

I perioden fra 1908 til 1916 udviklede Thorndike og hans studerende test i aritmetik og i andre fag (Kilpatrick, 1993, s. 34). Thorndike blev allerede i 1918 betegnet som faderen til "scientific measurement". Han og hans studerende bidrog til udviklingen og dominansen af "objektive" test som har været det mest karakteristiske træk ved testning i USA i hele 1900-tallet (Shepard, 2000, s. 5).

Det er fremhævet i bl.a. (Keitel, under udgivelse) at matematik lige fra de tidlige år har spillet en vigtig rolle i testning af både intelligens og præstationer. Dette gælder dels i indholdet af de pågældende test og dels i måden de konstrueres på.

Samme sted fremhæves det også at udviklingen af de oprindelige test skete i nær tilknytning til den fremherskende ideologi om effektivitet. Blandt andet derfor fik aritmetiske færdigheder og logisk tænkning en fremtrædende plads i testene. Evnen til at ræsonnere logisk blev mere eller mindre sat lig med intelligens; derfor var ma-

tematiske testopgaver særligt egnede til testning. Det var desuden en fordel at vælge matematikopgaver i test fordi der ikke var tvivl om rigtigt og forkert. Traditionel matematikundervisning var på en måde forbillede for organisering af test med dens fragmentering af faget i opgaver. Det synlige nærvær af matematik i konstruktionen og i indholdet af test, samt i vurderingen af testresultater, gav et skær af videnskabelig objektivitet (Keitel, under udgivelse).

Jeg har peget på hvordan testningen i sin barndom var tæt knyttet til intelligens-testning, hvordan den var under stærk indflydelse af behaviorismen, og hvordan den via et matematisk familieskab undgik at blive stillet til regnskab for sin berettigelse og objektivitet. I forlængelse heraf er det tankevækkende at se på den test- og evalueringkultur der dominerer mange steder i verden. For selvom den har udviklet sig og er blevet beriget med mange nyskabelser, så findes der 100 år gamle tendenser i de færdighedsprægede test. Enkle delspørgsmål med tiltro til objektive svar der kan sammenlignes og sammenlignes, går igen. Mogens Niss skrev tilbage i 1993:

Selv om behavioristiske positioner er blevet mindre markante blandt undervisere i matematik i de seneste år, så er konstruktionen af testopgaver stadig under indflydelse af den behavioristiske filosofi. (Niss, 1993, s. 20, min oversættelse)

Jeg undrer mig over at der gennem 100 år er bevaret så mange elementer af den oprindelige form i de testmaterialer der er i brug i skolen i dag. Man kan pege på følgende grunde til at kvantitativ testning har fået den store udbredelse (Linn, 2000, s. 4):

- Testning er relativt billig sammenholdt med kvalitetsudvikling i form af mere undervisningstid eller efteruddannelse af lærere.
- Testning kan let pålægges fra eksternt hold.
- Testning kan hurtigt blive indført og anvendt – f.eks. inden for en valgperiode.
- Testresultater er synlige og kan offentliggøres i pressen.

Der kan således være oplagte grunde til fra central side at indføre kvantitative testformer, og desværre er det en ret udbredt opfattelse at det aldrig kan gøre skade at teste i matematik. Derfor mener jeg at det er væsentligt at belyse hvilke konsekvenser forskellige test- og evalueringsformer har for undervisningen.

Evaluering har konsekvenser

På mange af de efteruddannelseskurser jeg har afholdt om evaluering i matematikundervisningen, har jeg mødt den opfattelse at det er uproblematisk at give eleverne en skriftlig test hvor de skal udføre elementære regneopgaver eller blot skrive resultater. Så enkelt forholder det sig imidlertid ikke. De evalueringsformer der bliver anvendt,

og det indhold der er mål for evalueringen, påvirker både elevens og lærerens opfattelse af hvad der er vigtigt i undervisningen, og dermed også elevens læreprocesser og lærerens valg af indhold og undervisningsform.

Men hvordan finder denne påvirkning sted? I det følgende vil jeg først sætte fokus på en tilbagevirkende effekt af test/evaluering på undervisningen og derefter på nogle negative konsekvenser af test/evaluering.

Tilbagevirkende effekt

Det er blevet påpeget at ikke blot den formative men også den summative evaluering har en tilbagevirkende effekt på undervisningen:

Hvad der bliver evalueret, bestemmer [også] hvad der skal undervises i. Det der bliver værdsat i evalueringen, fungerer som mål i undervisningen. ... Heraf følger at hvis vi ikke kan evaluere det vi værdsætter, vil indholdet i undervisningen indsnævres og afspejle den mangelfulde evaluering. (Clarke, 1996, s. 329-330, min oversættelse)

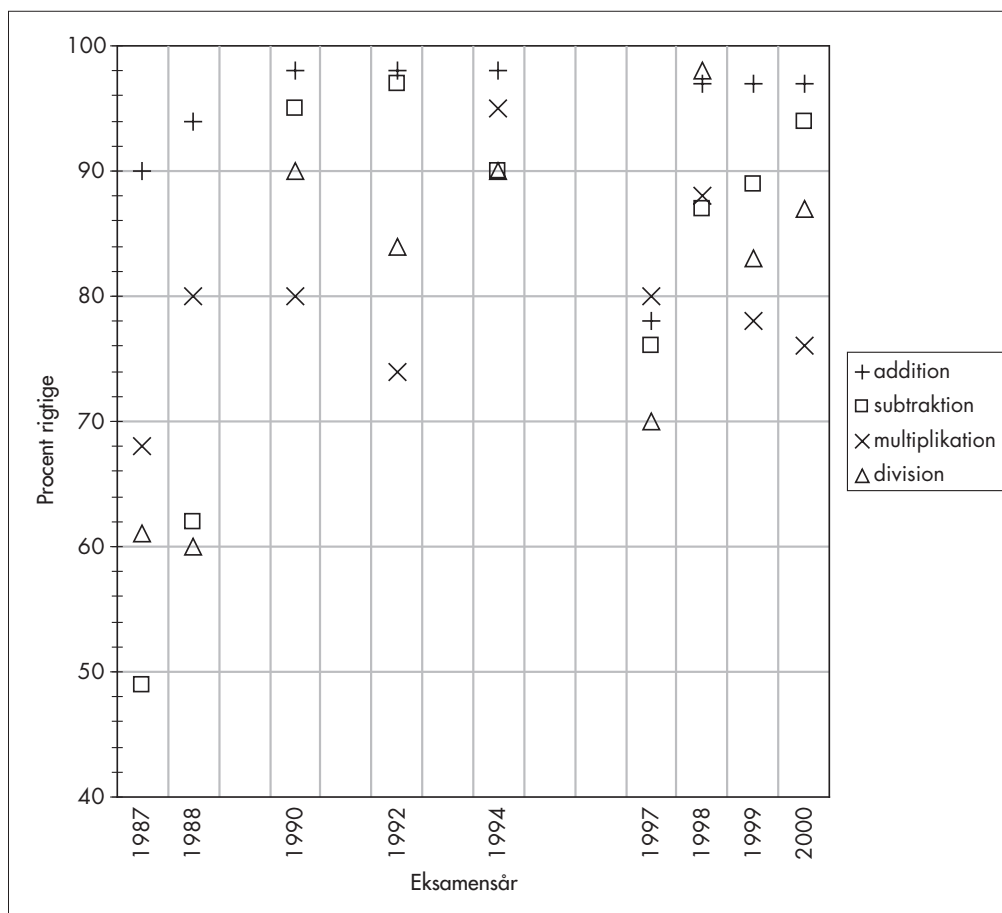
Lignende konklusioner optræder hos (Niss, 1993; Ruthven; 1994; Linn, 2000). Desuden påpeges det at ud over at have indflydelse på hvad der skal læres, hvordan det skal læres, og hvad det betyder at være elev, så former testning også elevens selvopfattelse og dermed dennes fremtidsmuligheder (William m.fl., 2004, s. 50).

Eksistensen af den tilbagevirkende effekt af test på undervisningen blev klart dokumenteret i en stor undersøgelse i Australien i 2000. Her blev summative evalueringsformer og matematikundervisningen i de to sidste år af "secondary school" sammenlignet i de to mest folkerige stater. De obligatoriske summative evalueringer viste sig at have en gennemgribende indflydelse på lærernes opfattelse af hvad det var vigtigt at undervise i. Forfatterne konkluderer at test har en fundamental indflydelse på valg af indhold i undervisningen således at det der er vanskeligt at teste, bliver udeladt i undervisningen (Barnes m.fl., 2000).

På den hjemlige arena findes et godt eksempel på en tilbagevirkende effekt i forhold til afgangsprøverne i matematik. I årene op til 1995 indgik der en færdighedsprøve i de afsluttende prøver efter 9. klasse. Med faghæftet for matematik fra 1995 var det hensigten at færdighedsprøven skulle afskaffes, men dette skete aldrig, for den 27. januar 1997 blev det politisk bestemt at den skulle "genindføres". Der var imidlertid et kompliceret forspil til denne endelige beslutning. I den første del af januar 1997 forlød det således fra Undervisningsministeriet at den annoncerede afskaffelse var annulleret, mens denne udmelding ca. en uge herefter blev trukket tilbage. Endnu en uge derefter forelå den endelige beslutning om at færdighedsprøven skulle bibeholdes og gennemføres i sommeren 1997. I den periode afholdt jeg

hver uge efteruddannelse på en skole. Her erfarede jeg hvordan matematiklærerne forholdt sig til i første omgang meddelelsen om færdighedsprøvens (gen)indførelse. Reaktionen var at nu blev de nødt til at "træne eleverne op" til denne. Ugen efter, da prøven var ude igen, gav lærerne udtryk for lettelse over at de nu kunne foretage sig noget mere fornuftigt. Og så ugen efter, ja, så blev de altså alligevel nødt til at træne færdigheder.

Gav det planlagte ophør af færdighedsprøven sig udslag i elevernes besvarelse? En opgørelse over procentdelen af korrekte besvarelser af de fire første opgaver i færdighedsregning ved afgangsprøven i 9. klasse i nogle af årene 1987-2000 kan kaste lys over dette spørgsmål. Figur 2 er baseret på materiale fra Undervisningsministeriet.



Figur 2. Færdighed i elementær regning. Der indgår kun data for de år hvor et årstal er angivet.

Graferne i figur 2 viser procentdelen af elever med korrekte besvarelser af de fire første elementære regneopgaver i færdighedsprøven efter 9. klasse. Hver opgave bestod af et regnestykke i en af de fire regningsarter, og der måtte ikke anvendes hjælpemidler. Fra 1987 og frem til og med 1994 ses en stigning i antal rigtige besvarelser inden for alle fire regningsarter. Herefter kommer et brat fald i 1997 hvor færdighedsprøven først blev annonceret genindført i januar. Udsvingene i resultaterne er markante og kan formentlig tilskrives tilbagevirkningseffekten i forbindelse med den planlagte afskaffelse af færdighedsprøven. Det skal dog bemærkes at de fire første opgaver netop i sommeren 1997 er sat ind i en kontekst, men sværhedsgraden er ellers den samme⁴. I årene efter steg addition og subtraktion til "normalt" niveau, og der var nogen stigning i division og både lidt frem- og tilbagegang i multiplikation, hvilket kan tilskrives lidt vanskeligere opgaver i denne regningsart. Med forbehold for den kontekstuelle ramme i 1997 illustrerer figur 2 tilbagevirkningseffekten på den danske matematikundervisning.

Negative konsekvenser af evaluering

Det er yderst forståeligt hvis læreren i sin undervisning vægter det som evalueringsformen måler, og bruger en stor del af tiden på at træne specifikke færdigheder hvis evalueringen angår disse. Umiddelbart kan dette synes rigtigt da det f.eks. er nyttigt at kunne regne hurtigt og præcist. Men ser man igen på figur 2 og tager i betragtning at resultaterne ikke alene afspejler eksamenstræning men også er et udtryk for hvad eleverne har lært sig igennem ni skoleår, kan man tolke dykket i 1997 som et tegn på at eleverne hurtigt glemmer de færdigheder de har erhvervet gennem skoletiden, hvis de ikke vedligeholdes gennem øvelser i skoleregi frem til eksamensdagen.

Ensidig færdighedstræning har dog andre konsekvenser. For det første er der begrænset undervisningstid til rådighed, og hvis vægten lægges på færdighedstræning, kan det ikke undgås at andre væsentlige områder forsømmes. Det kan dreje sig om de matematiske kompetencer der er beskrevet i KOM-projektet (Niss & Jensen, 2002), eller om de områder der er nævnt i de to faghæfter for matematik (Faghæfte 12, 1995) og (Fælles Mål, 2003), f.eks. undersøgende og eksperimenterende virksomhed, problemløsning og formulering af egne problemstillinger samt kommunikation i, med og om matematik. Hermed indsnævres undervisningen i matematik, og den lever således ikke op til det der lægges vægt på i den aktuelle danske læreplan for faget.

4 F.eks. lyder additionsopgaven i 1994: " $4168 + 1972 = \underline{\quad}$ ". I 1997 er der vist billeder af henholdsvis kartofler og et salathoved med tilhørende tekst: "Salathoved pr stk. 9,95" og "Kartofler 1 kg 7,95". Opgaven lyder: "1 kg kartofler og 1 salathoved koster i alt $\underline{\quad}$ ". Bemærk i øvrigt at det ikke har været muligt at få resultater for alle årene i perioden.

Som nævnt i min introduktion skal der fremover testes i matematik på 3. og 6. klassetrin. Det fremgår af Undervisningsministeriets udbudsmateriale, "Oversigt over test, fag, profilområder, udvalgte trinmål, trinmålsenheder"⁵, at det gennemgående bliver de let målbare dele af trinmålene der skal testes. "Kommunikation og problemløsning" udgør ét af i alt fire centrale kundskabs- og færdighedsområder i Fælles Mål fra 2003, men dette område er fravalgt efter 6. klassetrin, og kun to ud af syv delmål i "Kommunikation og problemløsning" skal testes efter 3. klassetrin. Jeg finder denne begrænsning i testmål meget betænkelig, for tilbagevirkningseffekten af testningen vil formodentlig føre til forøget fokus på færdigheder på bekostning af problemløsning hvilket ikke er i overensstemmelse med faghæftet Fælles Mål (2003). Fra erhvervslivets side er der også fremsat kritik af fravalget af problemløsning som mål for testning efter 6. klassetrin; det fremgår af en artikel, "Erhvervsliv kritiserer skoletest", i Dagbladet Information d. 16. september 2005.

En anden konsekvens af færdighedsfokuseringen kommer til udtryk i elevudtalelser som: "Blot sig mig om jeg skal gange eller dividere, så skal jeg nok regne det". Når færdighedstræning indtager en dominerende rolle i matematikundervisningen, får eleverne let en opfattelse af at det at lære matematik er ensbetydende med at kunne gennemføre regneprocedurer. Eleverne får via det de bliver testet/evalueret i, påvirket deres opfattelse af hvad der er væsentligt i matematikundervisningen.

En sådan opfattelse er hæmmende for en undervisning hvor eleverne skal være aktive og selv tage initiativ til undersøgende virksomhed. Undervisningen kan f.eks. være tilrettelagt så eleverne skal eksperimentere med konkrete materialer med henblik på at støtte begrebsdannelsen inden for et bestemt område. I dette tilfælde vil mange matematiklærere nikke genkendende til spørgsmål som: "Hvornår skal vi have matematik?" Her er det underforstået at man skal regne stykker i bogen før det af eleverne opfattes som rigtig matematik.

Det er nødvendigt at inddrage udenlandske erfaringer hvis vi skal forberede os bedst muligt på en kommende dansk oprustning inden for test/evaluering. Derfor rettes opmærksomheden nu mod USA og England hvorfra inspirationen til vores nuværende fokusering på test i høj grad er hentet.

Erfaringer fra USA og England

I England får det konsekvenser hvis skolernes testresultater ikke er tilstrækkeligt gode. Det fører f.eks. til besøg af regeringens inspektører så ofte som hver tredje måned, og hvis efterfølgende forbedringer ikke sker hurtigt nok, kan skolen blive lukket og lærerne miste deres job (William m.fl., 2004, s. 54).

5 Lokaliseret oktober 2005 på <http://presse.uvm.dk/nyt/pm/07.pdf>

Den nuværende stærke optagethed af testning i USA begyndte ifølge Rotberg i starten af 1980'erne efter rapporter om tilbagegang i USA's uddannelsessystem. Det beskrives også hvordan den meget omfattende testaktivitet, den såkaldte high-stakes-testning⁶, skaber forringelser af undervisningen fordi der kan gå uger og endog måneder med testforberedelse hvorved det øvrige pensum bliver forsømt, og udenadslære bliver opmuntret. En anden følge af den dominerende testning er at det er svært at fastholde gode lærere på de klassetrin hvor testningen foregår, og skoleledere forlader deres job i stort antal hvorefter der er meget få ansøgere til de ledige stillinger. Testningen bevirker også at skoler lader elever forblive i klassen lige under den der skal testes, hvilket forhøjer scoringen men også forhøjer antallet af dem der forlader skolen i utide. (Rotberg, 2001, s. 170-171). Denne meget kritiske holdning til testning er Rotberg ikke ene om. Der er mange matematikdidaktikere som har udtrykt sig på linje hermed (Ruthven, 1994; Clarke, 1996; Black & Wiliam, 1998; Morgan, 2000; Woodrow, 2003).

Testningen er blevet forøget i USA efter årtusindeskiftet. I 2001 blev der vedtaget en ny lov, "No Child Left Behind", der har som mål at alle elever i offentlige skoler skal nå frem til at score "proficient"⁷ i læsning og matematik i 2014. Vejen til målet omfatter landsdækkende testning i læsning og matematik af alle elever fra 3. til 8. klassetrin (Linn m.fl., 2002). Hver stat skal konstruere "an accountability system" der indebærer at succes bliver belønnet med ekstra ressourcer, og fiasko bliver straffet med nedskæringer i de økonomiske midler og andre ulemper for lærere, skoleledere og skoler samt muligvis firing af dem der anses for at være ansvarlige.

I slutningen af oktober 2004 præsenterede Robert Sternberg en række grunde til at "No Child Left Behind Act" efter hans mening slår fejl. De vigtigste er:

- At skolerne skal stå til regnskab for resultater af test der ikke opfylder en acceptabel standard, og skoler med elever fra hjem med lav socioøkonomisk status, med mange svage elever og med elever der har engelsk som deres andet sprog, bliver straffet.
- At testningen dels opmuntrer skolerne til at få deres svage elever ud af skolen, hvilket sker i større og større omfang, og dels opmuntrer til snyd, som også forekommer. Desuden bidrager testning til at multiple-choice- og short-answer-test bliver betragtet som en slags universalmiddel mod uddannelsessystemets problemer.
- At undervisningen bliver ændret til testforberedende kurser hvor skolerne lægger vægt på udenadslære frem for på forståelse.

6 "High-stakes-testing" betegner en omfattende testning med væsentlige konsekvenser for elever, lærere, skoleledere eller skoler. Vi har endnu ikke et passende dansk ord for dette begreb.

7 *Proficient* betyder oversat til dansk 'at være dygtig, kyndig, velbevandret eller ekspert på et område'. I denne sammenhæng tillægges det betydningen 'at opnå meget gode resultater'.

I september 2005 blev der offentliggjort resultater af en undersøgelse: "High-Stakes Testing and Students Achievement: Problems for the No Child Left Behind Act" (Nichols m.fl., 2005). Forfatterne har studeret testdata fra "the National Assessment of Educational Progress" (NAEP) fra 25 stater i USA for at undersøge om det pres som high-stakes-testning frembringer, forbedrer elevernes præstationer.

Undersøgelsen viste:

- At stater med en stor del af elever fra minoritetsgrupper implementerer "accountability systems" som bevirker at der lægges et større pres på eleverne her end i andre stater. Dette tyder på at problemer der stammer fra high-stakes-testning, vil ramme elever fra minoritetsgrupper uforholdsmæssigt hårdt.
- At presset fra high-stake-testning er omvendt proportionalt med sandsynligheden for at eleverne vil fortsætte i 12. klasse. Resultaterne peger på at der er en sammenhæng mellem forøget pres fra testning, og det at et større antal elever bliver forhindret i at gå videre til næste klasse eller forlader skolen.
- At forøget pres fra testning ikke har produceret noget fremskridt i læsning på 4. og 8. klassetrin (de to trin som undersøgelsen koncentrerede sig om).
- At der i matematik er en svag sammenhæng mellem tidlig forøgelse i pres fra testning og præstationerne i NAEP-testen i 4. klasse. Imidlertid påpeger forfatterne at dette kan skyldes færdighedstræning og "teaching to the test". Denne fortolkning bliver støttet af manglende beviser på at tidligt pres gennem testning forbedrer præstationerne i matematik på 8. klassetrin.

Forfatterne konkluderer at der ikke er nogen tydelige beviser på at det pres som high-stakes-testning frembringer, fører til nogen forbedringer i elevernes præstationer. De foreslår at man stopper den politik der tvinger det offentlige uddannelsessystem til at tro på og være afhængig af high-stakes-testning.

Erfaringerne fra udlandet giver stof til eftertanke i en tid hvor vi i Danmark er på vej til at udvikle en test-/evalueringskultur. Den omfattende indsats i USA som følge af "No Child Left Behind"-loven har ifølge sidstnævnte store undersøgelser ikke ført til nogen positive resultater. Erfaringerne fra USA kan motivere spørgsmålet: Kan evaluering overhovedet komme undervisningen og dermed eleverne til gavn?

Evaluering kan fremme reformer

Flere udenlandske forskere som Barnes m.fl. (2000) har påpeget at ændringer i evalueringsformer kan fremme reformer i matematikundervisningen, og Cooney m.fl. fremhæver at "... evaluation can be construed not only as an object of reform but as

an instrument of reform as well.” (Cooney m.fl., 1993, s. 239). Reformer forstås her som ændringer i matematikundervisningen i lighed med nytænkningen i faghæftet fra 1995, der går igen i Fælles Mål fra 2003.

Hvilke karakteristika har evalueringsformer der kan fremme reformer?

I 1995 opstillede *National Council of Teachers of Mathematics* i USA (NCTM) seks standarder for eksemplarisk evaluering i matematik, og de er gentaget i *Principles and Standards for School Mathematics 2000*.

De nævner at evaluering skal:

1. reflektere den matematik som alle elever behøver at kunne og være i stand til at anvende,
2. fremme matematiske læreprocesser,
3. bidrage til lighed,
4. være en åben proces,
5. fremme gyldige konklusioner,
6. være en sammenhængende proces. (NCTM, 2000, s. 22)

Disse seks standarder er i god overensstemmelse med følgende fem “guiding principles” fremsat af Jan de Lange fra Freudenthal-Instituttet i Holland:

1. Evaluering skal være en integreret del af læreprocessen således at test/evaluering forbedrer læreprocessen.
2. Evaluering skal give eleverne mulighed for at vise hvad de kan, i stedet for det de ikke kan. Vi kalder dette “positiv testning/evaluering”.
3. Evaluering skal kunne måle alle mål.
4. Kvaliteten af evalueringsformer skal ikke dikteres af muligheder for objektiv scoring.
5. Evaluering skal være tilstrækkelig praktisk så den kan passe ind i skolens hverdag. (de Lange, 1993, s. 199, min oversættelse)

Punkt 1 hos NCTM om at evaluering skal reflektere al den matematik som elever ved og kan bruge, svarer til punkt 3 hos de Lange, og begge indeholder således en advarsel om for snævre evalueringer. Punkt 2 hos de Lange afspejler en holdning til elevens kunnen idet det er denne der værdsættes, og ikke mangler der skal afsløres. I de Langes punkt 4 understreges det at kvaliteten af en evalueringsform ikke må bero på om den kan måles objektivt. Clarke går endnu længere idet han skriver at man kan sætte spørgsmålstegn ved rimeligheden i selve det at udtrykke en vurdering

af en elevs matematiske formåen i form af noget så snævert som en karakter. Han udtrykker det således:

Skoler og skolesystemer kan ikke længere foregive at et endimensionalt tal eller en karakter dækkende og nyttigt kan karakterisere en elevs/studerendes matematiske kunnen. (Clarke, 1996, s. 359, min oversættelse)

Han understreger at en kvalitativ vurdering baseret på langt mere end en enkelt test vil yde eleven større retfærdighed, og at det vil give et mere fyldestgørende billede af elevens matematiske kompetencer hvis der anvendes forskellige former for evaluering.

Punkt 2 hos NCTM og punkt 1 hos de Lange anbefaler formativ evaluering da der skal evalueres med henblik på at give eleverne optimale muligheder for læring. Hos NCTM lægges der meget vægt på formativ evaluering, og det pointeres at evaluering skal være meget mere end blot en test i slutningen af et forløb for at se hvordan eleverne klarer sig under specielle betingelser. I stedet for skal evaluering være en integreret del i undervisningen til gavn for eleverne. De lægger også vægt på at læreren er i stand til at tolke de informationer der fremkommer ved evaluering, og forsøger at følge elevens tankegang i processen i stedet for blot at se på rigtigt og forkert (NCTM, 2000, s. 22-24).

Freudenthal-Instituttet i Holland og NCTM i USA er begge i brede matematikundervisningskredse anerkendte for arbejde inden for matematikdidaktikken og kommer altså med overensstemmende bud på hvad der skal til for at evaluering kan få en konstruktiv betydning for undervisningen.

Men findes der konkrete eksempler på at test/evaluering har fremmet reformer? Herhjemme har Paola Valero beskrevet en ung matematiklærer i 8. klasse som vælger bestemte problemstillinger i undervisningen fordi hun er bekymret for hvordan eleverne skal klare sig til den mundtlige afgangsprøve efter 9. og 10. klasse. Læreren forsøger derfor at finde spændende problemstillinger – f.eks. hvordan man kan bestemme det størst mulige volumen af en kasse med given overflade – og arbejder målrettet med at få eleverne til at kommunikere om deres metoder (Valero, 2002, s. 241). Denne summative evaluering får således indflydelse både på hendes valg af undervisningsmateriale og elevernes arbejdsformer allerede i 8. klasse.

Denne beskrivelse stemmer godt overens med erfaringer jeg har fra de mange efteruddannelseskurser jeg har afholdt. Langt de fleste matematiklærere har den opfattelse at den mundtlige afgangsprøve efter henholdsvis 9. og 10. klasse har haft en endog meget stor og positiv indflydelse på den forudgående undervisning. De fremhæver især den forøgede opmærksomhed på arbejdet med matematiske problemstillinger og samarbejdet om løsning af disse og på at eleverne selv skal formulere problemstillinger

som de efterfølgende skal behandle. Hermed bliver eleverne bedre til at kommunikere om matematiske forhold, hvilket igen har en positiv effekt på deres læreprocesser. Disse ændringer er i overensstemmelse med faghæftet fra 1995 og et eksempel på at en evalueringsform har fungeret som et instrument til at fremme implementeringen af de ret omfattende nyskabelser der dengang blev lanceret.

Det er især i forhold til den formative evaluering at den enkelte lærer har mulighed for at vælge evalueringsredskaber der støtter en kvalificering af undervisningen. Der findes i dag et rigt udbud af evalueringsformer egnet til formativ evaluering, såsom logbog, portfolio, begrebskort, det at formulere opgaver til kammerater og samtale med den enkelte elev.

I denne sammenhæng vil jeg nævne en evalueringsform der blev udviklet med den hensigt at indføre formativ evaluering i matematik i alle 1.-5.-klasser i Brøndby Kommune 1997-99. I løbet af de to år blev der udviklet et materiale som består af ark med enkle regneopgaver, åbne opgaver og problemløsningsopgaver. På hvert ark er der afsat plads til at eleverne ved hjælp af tegninger, ord og/eller matematiske symboler kan redegøre for deres løsningsmetoder (Andersen og Jess, 2000). Hermed får læreren mulighed for at få indsigt i elevens læreproces og i hvordan den enkelte elev tænker når denne arbejder med en matematisk problemstilling – en indsigt der kan hjælpe læreren til at afdække potentielle læringsmuligheder og opdage uhensigtsmæssige løsningsstrategier m.m.

Der foreligger en undersøgelse af Brøndbyprojektet hvor der er sat fokus på om matematiklærernes deltagelse i projektet har ført til ændringer eller intentioner om ændringer i deres praksis (Jess, 2004). Den dominerende effekt har været en stærkt forøget opmærksomhed på sprogets betydning for læring af matematik, hvilket formentlig kan tilskrives udformningen af evalueringsmaterialet idet eleverne skal mere end blot skrive et resultat, og dette bygger på sproglig formåen. I mange kommentarer har lærerne udtrykt intentioner om "at lære eleverne at sætte ord på", hvilket er i overensstemmelse med faghæftet fra 1995.

Evalueringsformen har givet anledning til at flere lærere har fået et mere nuanceret syn på elevernes matematiske formåen. Nogle af de elever som lærerne havde anset for at være dygtige, viste sig at have svært ved at redegøre for deres løsningsstrategier, hvilket måske er udtryk for at disse elever blot er gode procedureregnere. Der var også elever som overraskede læreren ved uventet at kunne løse opgaver og redegøre for hvordan. Disse elever har tilsyneladende fået en ny chance for at demonstrere matematisk kompetence fordi kravet om at redegøre for deres løsningsstrategi – for små elevers vedkommende oftest i form af tegning – giver dem et redskab som støtter dem. Det er f.eks. blevet legalt at tegne, og det kan udnyttes i løsningsprocessen og bliver ydermere værdsat af læreren i den pågældende sammenhæng.

Nogle lærere i Brøndbyprojektet beskriver at de i den daglige undervisning er blevet mere opmærksomme på at bede eleverne om at forklare deres løsningsstrategier. Der er også blevet sat mere fokus på indholdet i undervisningen, og lærebøgernes styrende funktion er i nogen grad blevet formindsket. Men det fremgår af undersøgelsen at hensigten med at afdække den enkelte elevs læringspotentialer kun blev nået i begrænset omfang. Og det udarbejdede materiale har haft en styrende effekt, hvilket bl.a. kom til udtryk i bemærkninger som: "Materialet har delvist styret undervisningen".

Alt i alt førte denne evalueringsstrategi til forøget opmærksomhed på sprogets betydning for elevernes læreprocesser og på forståelse frem for automatiserede processer, hvilket må betragtes som et fremskridt i forhold til realiseringen af reformbestræbelserne i faghæftet fra 1995. Samtidig er det et eksempel på at evaluering har en tilbagevirkende effekt.

Konklusion

Jeg har i denne artikel forsøgt at formidle viden og erfaring om testning/evaluering i matematikundervisningen på grundskoleniveau. Min hovedkonklusion er at testning/evaluering kan lede undervisningen i to forskellige retninger. Man kan pege på uhensigtsmæssige tilbagevirkende effekter på undervisningen, men man kan også pege på muligheden for at understøtte reformer. Det placerer et ansvar hos dem der har indflydelse på hvilke der vælges – især hvis en bestemt type test/evaluering gøres obligatorisk.

For den enkelte lærer er det en stor udfordring at gennemskue konsekvenserne ved brug af forskellige typer evaluering. Bliver man som matematiklærer pålagt at anvende bestemte former, og her vil der især være tale om de summative, er det uhyre væsentligt at forholde sig kritisk til i hvilken grad denne form skal have indflydelse på ens undervisning. Det vil være en fordel at diskutere disse forhold med fagkolleger for at få skærpet opmærksomheden på eventuelle faldgruber.

I nogle lande offentliggør man testresultater. Det sker formodentlig ud fra en opfattelse af at konkurrence bidrager til at folk yder deres bedste. Men de menneskelige omkostninger er store, og offentliggørelse i sig selv kan medføre et pres der frister til snyderi og anden utilsigtet adfærd, som det blev dokumenteret med erfaringer fra USA. Det er erfaringer man bør lære af i Danmark. Jeg har ikke set dokumentation for at offentliggørelse virkelig fremmer kvaliteten af undervisningen, og for øjeblikket peger det uddannelsespolitiske flertal i Danmark da også kun på begrænset offentliggørelse i lukkede kredse.

Linn har gennem hele sin karriere beskæftiget sig med test og evaluering. I en artikel fra 2000 skriver han at han gerne ville kunne konkludere at den megen brug af test gennem 50 år havde forbedret uddannelsessystemet og elevernes udbytte dramatisk, men han må konstatere: "Unfortunately, that is not my conclusion." (Linn, 2000, s. 14).

Vi skulle nødig nå til samme uheldige resultat i Danmark om 20 år. Det centrale i min konklusion er derfor at når man ønsker obligatoriske nationale test fra politisk side, så skal de have en form og et indhold der fremmer fagets formål og læreplan gennem en positiv tilbagevirkende effekt på undervisningen. Udvikling af sådanne evalueringsformer kræver faglig indsigt, tid, omtanke og økonomiske midler. At det forholder sig sådan, skulle gerne nå ud over de professionelle cirkler – en opfattelse som jeg deler med Schoenfeld:

Den gode nyhed er at der eksisterer Standard-baserede evalueringsformer af høj kvalitet, og at sådanne evalueringsformer kan anvendes produktivt igangsættende for forandringer. ... De gode evalueringsformer tenderer til at være dyrere end de andre er, og de er vanskeligere at forstå for folk i al almindelighed. Vi står foran et problem der både angår offentlig uddannelse og public relation. (Schoenfeld, 2002, s. 22, min oversættelse)

For mig har det været afskrækkende at følge udviklingen vestpå, og jeg håber at politikere og skolemyndigheder i Danmark vil fastholde den opfattelse at man får det bedste skolesystem hvis man bygger det på tillid til den professionelle lærer frem for at lade sig smitte af de forhold der hersker i USA og England, der i korthed bygger på en opfattelse af at kontrol og konkurrence fremmer kvaliteten i skolerne.

Referencer

- Andersen, M. & Jess, K. (2000). *Evaluering i matematikundervisning*. København: Alinea.
- Barnes, M., Clarke, D.J. & Stephens, M. (2000). Assessment: The engine of systematic curricular reform? *Journal of Curriculum Studies*, 32(5), s. 623-650.
- Clarke, D. (1996). Assessment. I: A.J. Bishop, K. Clements, C. Keitel, J. Kilpatrick & C. Laborde (red.), *International Handbook of Mathematics Education* (s. 327-370). Dordrecht: Kluwer Academic Publishers.
- Cooney, T., Badger, E. & Wilson, M. (1993). Assessment, understanding mathematics, and distinguishing visions from mirages. I: *Assessment in the mathematics classroom. Yearbook* (s. 239-247). National Council of Teachers of Mathematics.
- Faghæfte 12, Matematik*. (1995). Undervisningsministeriet, Folkeskoleafdelingen.
- Fælles Mål, Faghæfte 12, Matematik*. (2003). Undervisningsministeriet, Uddannelsesstyrelsens håndbogsserie nr. 10, 2003. Grundskolen.
- Jess, K. (2004). Formativ Evaluering i Matematikundervisningen – Ændringer i praksis. *Nordisk Matematikdidaktik (NOMAD)*, 9(4), s. 15-47.
- Keitel, C. (under udgivelse). The shaping of mathematics education through testing. I: M. Niss (red.), *Proceedings of the Tenth International Congress on Mathematical Education*.

- Kilpatrick, J. (1993). The chain and the arrow: From the history of mathematics assessment. I: *Investigations into assessment in mathematics education* (s. 31-46). Dordrecht: Kluwer Academic Publishers.
- de Lange, J. (1993). Assessment in problem-oriented curricula. I: *Assessment in the mathematics classroom. Yearbook* (s. 197-208). National Council of Teachers of Mathematics.
- Linn, R.L., Baker, E.L. & Betebenner, D.W. (2002). Accountability systems: Implications of requirements of the No Child Left Behind Act of 2001. *Educational Researcher*, 31(6), s. 3-16.
- Linn, R.L. (2000). Assessments and accountability. *Educational Researcher*, 29(2), s. 4-16.
- Morgan, C. (2000). Discourses of assessment – Discourses of mathematics. I: J. Filipe & M. Santos (red.), *Mathematics Education and Society. Proceedings of the Second International Mathematics Education and Society Conference (MES2) 29th – 31st March 2000* (s. 58-76). Lissabon: Centro de Investigação Faculdade de Ciências da Universidade de Lisboa.
- NCTM. (1995). *Assessment standards for school mathematics*. Reston, VA, National Council of Teachers of Mathematics.
- NCTM. (2000). *Principles and standards for school mathematics*. Reston, VA, National Council of Teachers of Mathematics.
- Nichols, S.L., Glass, G.V. & Berliner, D.C. (2005). High-stakes testing and student achievement: Problems for the No Child Left Behind Act. *Education Policy Research Unit (EPRU), September*. Lokaliseret 30. september 2005 på: <http://www.asu.edu/educ/eps1/EPRU/documents/EP1-0509-105-EPRU.pdf>
- Niss, M. (1993). Assessment in mathematics education and its effects: An introduction to investigations into assessment in mathematics education. I: M. Niss (red.), *Investigations into assessment in mathematics education* (s. 1-30). Dordrecht: Kluwer Academic Publishers.
- Niss, M. & Jensen, T.H. (red.). (2002). *Kompetencer og matematiklæring*. Ideer og inspiration til udvikling af matematikundervisningen i Danmark, "KOM-projektet". Uddannelsesstyrelsens temahæfteserie nr. 18. Undervisningsministeriet.
- Rotberg, I.C. (2001). *Phi Delta Kappan, October, 83(2)*, s. 170-171.
- Ruthven, K. (1994). Better judgement: Rethinking assessment in mathematics education. *Educational Studies in Mathematics* 27, s. 433-450.
- Schoenfeld, A. (2002). Making mathematics work for all children: Issues of standards, testing, and equity. *Educational Researcher*, 31(1), s. 13-25.
- Shepard, L.A. (1991). Psychometrician's beliefs about learning. *Educational Researcher*, 20(6), s. 2-16.
- Shepard, L.A. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29(7), s. 4-14.
- Sternberg, R.J. (2004). Good intentions, bad results. A dozen reasons why the No Child Left Behind Act is failing our schools. *EdWeek.org, Wednesday, October 27, 24(9)*, s. 42-56.
- Thomson, R. (1969). *Psykologiens historie*. Det Schønbergske Forlag.
- Torpe, H. (1972). *Intelligensforskning og intelligensprøver*. J.H. Schultz Forlag.

- Valero-Dueñas, P.X. (2002). *Reform, democracy, and mathematics education*. København: Department of Curriculum Research, The Danish University of Education, Copenhagen.
- Wiliam, D., Bartholomew, H. & Reay, D. (2004). Assessment, learning and identity. I: P. Valero & R. Zevenbergen (red.), *Researching the Socio-Political Dimensions of Mathematics Education* (s. 43-61). Dordrecht: Kluwer Academic Publishers.
- Woodrow, D. (2003). Mathematics, mathematics education and economic conditions. I: A.J. Bishop, M.A. Clements, C. Keitel, J. Kilpatrick & K.S.L. Frederick (red.), *Second International Handbook of Mathematics Education* (s. 9-30). Dordrecht: Kluwer Academic Publishers.