

De nationale it-test: den første evaluering af afviklingen

af

Christen Sørensen

Discussion Papers on Business and Economics
No. 7/2008

YDERLIGERE INFORMATION
Institut for Virksomhedsledelse og Økonomi
Det Samfundsvidenskabelige Fakultet
Syddansk Universitet
Campusvej 55
DK-5230 Odense M

Tel.: +45 6550 3271
Fax: +45 6550 3237
E-mail: lho@sam.sdu.dk

<http://www.sam.sdu.dk/osbec>

ISBN 978-87-91657-22-1

De nationale it-test: den første evaluering af afviklingen

Af:

Christen Sørensen

Professor, Syddansk Universitet

1. Indledning

”Der skal en hel del justeringer og ændringer til – men det kan lade sig gøre” sagde direktør Martin Isenbecker fra Skolestyrelsen i en pressemeddelelse fra 29. oktober 2007, da den første evaluering af funktionaliteten af de nationale it-test var afsluttet med rapporten fra det såkaldte REVIEW-panel. Evalueringen vedrørte den første version af de tre test i Dansk/Læsning (8. klasse), Matematik (6. klasse) og Fysik/Kemi (8. klasse), der blev gennemført i maj-juni 2007.¹

Og direktørens udsagn var ikke en overdrivelse – for at sige det mildt! For en mere kritisk evaluering skal man lede længe efter – og givetvis forgæves, når samtidig REVIEW-panelets udpegning og sammensætning haves for øje. Konsortiet med COWI A/S i spidsen, der udviklede de nationale test, og Undervisningsministeriet udpegede nemlig i fællesskab samtlige panelets 10 medlemmer. Danmarks Lærerforening fik f.eks. ikke mulighed for at udpege et medlem. Og ikke nok hermed. Samtlige 10 medlemmer havde enten været med til at udvikle testene for konsortiet eller havde rådgivet Undervisningsministeriet. Så det var næppe de mest kritiske, der var repræsenteret i panelet. Og alligevel er panelets rapport: ”De nationale it-baserede test i folkeskolen”, der i bedste management stil er sammenskrevet af konsulentfirmaet Devoteam Consulting A/S med udgangspunkt i panelets undersøgelser og drøftelser, meget kritisk.

Det betyder ikke, at der ikke omtales positive forhold. Indledningsvist opridses f.eks. de muligheder, som it engang kan give for at udvikle og teste elevers færdigheder. Der nævnes syv vigtige læringsprocesser fra konkret viden til mere og mere komplekse processer som bl.a. evnen til at analysere og syntetisere (sammenfatte/sammendrage). Og det er da meget muligt og endda sandsynligt, at sådanne muligheder engang vil kunne implementeres. Men panelet undlader dog ikke umiddelbart efter opremsningen af de syv læringsprocesser at nævne, at de test, der blev anvendt i maj-juni 2007, helt overvejende vedrørte det første og mest enkle område: tilstedeværelse af konkret viden, dvs. paratviden! Derfor er det vanskeligt at undgå et indtryk af, at disse positive muligheder mest af alt opregnes for at muliggøre, at de efterfølgende meget kritiske bemærkninger ikke virker for voldsomme og dermed kunne risikere at sætte en stopper for dette kostbare, hovedkulde og ikke mindst meget prestigefyldte projekt.

Udover de mulige positive fremtidsperspektiver er der tre andre forhold, der vurderes positivt: 1) valget af de tre profilområder inden for de tre fag, 2) omhyggeligheden med at afprøve – via såkaldte Rasch-test – at de stillede testspørgsmål kan anvendes til at vurdere elevernes færdigheder samt 3) valget af den adaptive testform. At testformen er adaptiv betyder, at spørgsmålenes sværhedsgrad automatisk tilpasses elevens niveau, der løbende vurderes ud fra besvarelsen af de hidtil stillede spørgsmål. At dette så betyder, at det er vanskeligere og kræver langt flere spørgsmål at nå til en sikker vurdering heraf - forstærket af multiple choice dominansen i spørgsmålene - er noget andet, og en udfordring, der overhovedet ikke er løst, jf. afsnit 2.

¹ Designet – med tilhørende kritik – af de it-baserede nationale obligatoriske test blev beskrevet og analyseret i Samfundsøkonomen nr. 2, 2007, jf. Sørensen (2007).

2. Hovedproblemet: mangel på opgaver

Som de nationale test *er valgt udformet*, er der især fem forhold, der betyder, at der skal produceres virkelig mange opgaver pr. fag til den databank, som opgaverne fra elevernes trækkes fra: 1) den adaptive testform med fem sværhedsniveauer, 2) tre profilområder inden for hvert fag, 3) besvarelsesformen med primært multiple choice, der muliggør, at korrekte svar kan gives, selv om eleven ikke forstår opgaven/spørgsmålet, 4) beslutningen om, at samtlige trin- og slutmål – om muligt – skal it-testes og 5) det forhold, at eleverne frivilligt kan afprøve deres færdigheder to gange inden deltagelse i den obligatoriske test.²

Den valgte adaptive testform betyder, at sværhedsgraden øges, såfremt eleven svarer korrekt to gange i træk, og reduceres, hvis der svares forkert to gange i træk eller svares blankt. En adaptiv testform med fem sværhedsniveauer stiller dog ikke krav om hele 5 gange så mange opgaver, idet næsten alle elever vil få stillet opgaver fra mindst to sværhedsniveauer med den valgte adaptive testform. Antallet af nødvendige spørgsmål stiger også med antallet af profilområder, selv om den samlede tid til testafviklingen ikke stiger, idet der kræves et ikke ubetydeligt antal spørgsmål til med rimelighed sikkerhed at vurdere elevens færdigheder/standpunkt på de enkelte profilområder, som det gøres i det nationale it-testsystem. Dominansen af spørgsmål med på forhånd givne svarmuligheder – som f.eks. multiple choice spørgsmål – gør naturligvis ligeledes, at der må stilles flere opgaver, inden en elevs færdighedsniveau på en skala fra 1 til 5 (1: lavest, 5: højest) kan bestemmes med nogenlunde sikkerhed. At samtlige trin- og slutmål – om muligt – skal indgå øger naturligvis også kravene til antal spørgsmål. Det forhold, at eleverne frivilligt kan prøvekøre et testforløb to gange, bevirker endvidere, at der bør være så mange spørgsmål, at sandsynligheden for at få samme spørgsmål i den obligatoriske test som i de frivillige forsøg bliver beskeden, når nu systemet ikke er indrettet på at huske.

Kravene til antallet af spørgsmål, givet den valgte testform, blev da også erkendt, da it-systemet blev designet. Det blev således besluttet, at der *mindst* skulle være så mange opgaver pr. fag, at en gennemsnitlig elev (en middelhurtig elev) skulle bruge 675 min. til at løse disse opgaver. Dette skal sammenholdes med, at det *var planlagt*, at testen skulle vare max. 45 min., hvorfor der kunne siges at være en 15 gange overdækning i databanken. Men denne overdækningsfaktor reduceres jo ganske afgørende ved, at der er fem sværhedsniveauer, 3 *planlagte* profilområder, en række trin- og slutmål samt mulighed for to frivillige prøver. Overdækningen i maj-juni 2007 blev endvidere reduceret af, at grænsen på 45 min. var suspenderet, idet testen fortsatte indtil enten læreren eller eleven afbrød den, eller indtil der ikke var flere opgaver i databanken. I Dansk/Læsning, Fysik/Kemi og Matematik var der således hhv. 36, 43 og 80 pct., der brugte mere end 45 min. Den gennemsnitlige varighed for disse elever var hhv. 53, 56 og 69 minutter.

For at imødekomme det planlagte minimumskrav til antallet af opgaver, blev der udformet i alt 2000 opgaver til afprøvning omkring årsskiftet 2006/07 for de tre fag, som eleverne skulle afprøves i ved maj-juni 2007 testene.

² Trin- og slutmål blev i 2003 indført for folkeskolens fag med bekendtgørelse nr. 571 af 23. juni 2003. Trin- og slutmålene angiver de nationale mål for, hvilke kundskaber og færdigheder eleverne forventes at have tilegnet sig ved henholdsvis afslutningen af undervisningen i et fag i folkeskolen og ved afslutningen af bestemte klassetrin. Slutmålene angiver de langsigtede mål for de kundskaber og færdigheder, der skal erhverves, og som undervisningen skal sigte på i hele forløbet. Trinmålene angiver kortsigtede mål for kundskaber og færdigheder, som skal tjene til planlægning og vurdering af undervisningen, som dialogværktøj og som pejlemærker i relation til den løbende vurdering af elevens udbytte af undervisningen. Selv om alene trinmål anføres i visse tilfælde i dette papir, skal dette alligevel forstås som både trin- og slutmål i dette papir.

Imidlertid var det kun godt 38 pct. eller 760 opgaver, der kunne godkendes efter forhåndsafprøvningen ved årsskiftet 2006/07. Panelet måtte da også konkludere, at opgavebanken kun indeholdt mellem 1/5 og 1/3 af det forventede antal opgaver givet elevernes svartider. Men selv dette overvurderede i realiteten testbankens størrelse. Det var nemlig kun halvdelen af de 760 opgaver, der var selvstændige, idet den anden halvdel var varianter af en allerede stillet opgave.

Der sondres i det nationale testsystem mellem opgaver og spørgsmål, idet nogle opgaver – 376 ud af de 760 - indeholdt mere end ét spørgsmål, ja nogle helt op til 10 spørgsmål. De først afprøvede godt 2000 opgaver indeholdt således godt 4000 spørgsmål, mens de 760 godkendte opgaver indeholdt 2036 spørgsmål. Opgaverne med mere end et spørgsmål har givet anledning til en række ekstra problemer, jf. afsnit 3.2.

Det helt afgørende problem, som panelet i flere sammenhænge påpeger, er, at der er alt for få opgaver i den opgavedatabank, som der trækkes på, når eleverne deltager i en test. Hvor grelt det har været fremgår af følgende:

I Dansk/Læsning (8. klasse) kunne der ikke testes i det meget vigtige profilområde *Afkodning* grundet mangel på spørgsmål. Det var også tilfældet for profilområdet *Tal og Algebra* i Matematik (6. klasse). I Fysik/Kemi (8. klasse) kunne der godt nok testes i alle tre profilområder, men problemet var her, at spørgsmålene inden for profilområderne *Energi og energiomsætning* samt *Stoffer og materialer* faktisk testede det samme, hvorfor der i Fysik/Kemi i realiteten også alene blev testet inden for to profilområder. Opgaverne inden for hvert fag/profilområde blev stillet ud fra de trin- eller slutmål, der er fastsat for alle fag. Tabel 1 viser i et konkret eksempel – i profilområdet geometri i matematik i 6. klasse - hvordan frafaldet af spørgsmål fordelte sig på trinmål.

Tabel 1. Planlagte og faktisk antal opgaver i profilområdet geometri i matematik, 6. kl.

Trinmål	Planlagte antal opgaver	Faktisk antal Opgaver
Benytte geometriske metoder og begreber i beskrivelse af fysiske objekter fra dagligdagen, herunder figurer og mønstre	19	15
Undersøge og beskrive enkle figurer tegnet i planen	33	19
Kende til grundlæggende geometriske begreber som vinkler og parallelitet	36	14
Arbejde med fysiske modeller og enkle tegninger af disse	37	31
Kende til forskellige kulturers metode til at angive dybde i billeder	12	8
Undersøge de enkelte tegnemethoders anvendelighed til beskrivelse af form og afstand	24	13
Måle og beregne omkreds, areal og rumfang i konkrete Situationer	28	4
Tegne, undersøge og eksperimentere med geometriske figurer, bl.a. ved at benytte computer	32	23
I alt i matematik i profilområdet geometri (6. klasse)	221	127

Det fremgår af tabel 1, at der var et gennemsnitligt frafald på godt 40 pct. i profilområdet geometri i matematik i 6. klasse. Frafaldet var endvidere meget ujævnt fordelt over trinmål.

Panelet anfører endvidere flere gange, at der er så få selvstændige spørgsmål, at validiteten af testresultaterne fra maj-juni 2007 er meget begrænset, jf. bl.a. følgende uddrag fra panelets rapport:

Side 29: brugen af variantopgaver gør, ”at eleven ’lærer’ noget i løbet af besvarelsen af opgaverne, som kan benyttes” ved løsningen af efterfølgende opgaver,

Side 31: mærkværdige testforløb, som panelet formoder ”har med den begrænsede opgavebank ... at gøre”,

Side 43: ”Givet opgavebankens størrelse, og da en del af opgaverne ser ud til at blive brugt meget hyppigt, vurderer panelet, at der pt. er en betydelig risiko for teaching-to-the test”,

Side 48: manglende opgaver betyder, ”at det er tvivlsomt, om alle trinmål er repræsenteret dækkende. Desuden vurderes de enkelte trinmål at være repræsenteret ved et antal opgaver, der varierer meget. Også de enkelte læringsmål under profilområderne vurderes at være ujævnt repræsenteret”,

Side 53: i relation til Fysik/Kemi anføres: ”Og antallet af opgaver bør udvides markant.”

Side 60: i relation til Matematik anføres: ”Matematiktesten skønnes begrænset valid på grund af tre forhold. Det ene forhold er, at der mangler opgaver helt ned på trinmålsniveau. Det andet forhold er, at profilområdet *Tal og algebra* mangler. Det tredje er, at der i forhold til visse opgaver ... er skabt usikkerhed om, hvad der egentlig testes i opgaverne.”

Side 56-57: Undervisningsministeriet har valgt, at der i Dansk/Læsning skal fokuseres snævert på en særlig del af faget, nemlig læsning, selv om der normalt ikke er selvstændig fokus herpå. Dette har betydet, ”at overordentlig mange af de stillede opgaver kan karakteriseres som opgaver, der måler elevens indsigt i dansk sprog, litteratur og kultur snarere end som opgaver i læsning” Dette betyder naturligvis, at langt de fleste af de i maj-juni 2007 anvendte opgaver må kasseres.

Side 68: ”Opgavebanken bør i meget høj grad udbygges. Så vidt panelet kan vurdere, skal omfanget af opgaver som minimum tre-dobles for at leve op til de oprindelige forudsætninger.” Dette udtrykker panelets sammenfattende kritik af manglen på antal opgaver. I denne sammenhæng er det vigtigt at være opmærksom på, at opgaverne så også skal kunne anvendes, jf. afsnit 3.1.

3. Mangel på anvendelige opgaver

3.1 Paratviden

Det er ikke blot begrænsningen til test af paratviden, der sætter alvorlige spørgsmålstejn ved det nationale it-testsystem. De væsentligste færdigheder, som skolegang bør give vore børn, kan ikke testes med det system, som er besluttet. Det blev der også gjort opmærksom på lige fra starten, jf. Sørensen (2006).

Panelet undlader heller ikke at fremhæve dette:

Side 10: ”En klar begrænsning er, at it i den form, der findes i dag, ikke kan vurdere sproglige kommunikative kompetencer, kreativitet og originalitet, evne til samarbejde og mere personlige

kompetencer. Derfor kan – og skal – it ikke stå alene i vurderingen af kompetencer, men må være et supplement til andre former for vurdering.”

Side 15: ”Som grundlag for udviklingen af opgavebanken er der blandt fagets trinmål udvalgt de trinmål eller dele heraf, der egner sig til it-baserede test.”

Side 25: ”Kravet om, at de nationale test skal give en valid og reliabel indplacering af eleverne på en en-dimensional skala (inden for hvert profilområde), indebærer i sig selv den begrænsning, at kun de færdigheder, der lader sig indplacere på en og samme skala, kan indgå i testen.

Dermed udelukkes måling af en række af mulige andre færdigheder hos eleven, og dermed reduceres til en vis grad testenes værdi som diagnostisk test”

Side 40: ”de trinmål eller dele heraf, der egner sig til it-baserede test, udvælges således at **faget dækkes så bredt** som muligt”

At de væsentligste færdigheder ikke kan testes ved det nationale it-testsystem burde selvfølgelig i sig selv mane til forsigtighed med forventninger til systemet. Men dette er ikke det eneste problem. For det er i realiteten *alene paratviden*, der bliver testet inden for de felter, der kan indgå, som det også blev forudsagt, så snart designet af det kommende nationale it-testsystem var blevet offentliggjort, jf. Sørensen(2006).

Panelet kommer også flere gange ind herpå, jf.:

Side 10: ”Den første version af de nationale test bygger helt overvejende på ... tilstedeværelse af konkret viden, herunder evnen til at kunne nævne, genkende, definere, identificere.” Altså kort sagt paratviden.

Side 28: For profilområderne *Energi og energiomsætning* og *Stoffer og materialer* inden for Fysik/Kemi: ”Den faglige vurdering er, at begge profilområder, således som opgaverne er udformet, repræsenterer paratviden og ikke en faglig forståelse.”

Side 53: For Fysik/Kemi: ”Panelet finder, at testen bør suppleres med opgaver, der tester forståelse og anvendelse af viden på baggrund af generelle faglige principper. Der bør også i højere grad bruges opgaver, som giver en faglig baggrund, som eleven kan argumentere fagligt ud fra.”

Side 71: I relation til Fysik/Kemi: ”Ideelt set bør der ikke være opgaver, der tester paratviden – men gerne opgaver, der tester paratforståelse.”

3.2 Polytome opgaver

Polytome opgaver er betegnelsen for opgaver med flere spørgsmål. Som allerede anført er der flere spørgsmål i 376 af de 760 opgaver, der blev anvendt i maj-juni 2007.

Der er to hovedproblemer knyttet til opgaver med flere spørgsmål. For det første kræver det adaptive design med fem sværhedsniveauer, at alle spørgsmål i polytome opgaver har samme sværhedsgrad. For det andet kan der være en mere eller mindre afhængighed mellem spørgsmål og svar i polytome opgaver, hvilket gør disse lettere at besvare – og såfremt antallet af svarmuligheder

svarer til antallet af spørgsmål, vil der automatisk blive svaret korrekt på det sidst besvarede spørgsmål, såfremt der er svaret korrekt på de øvrige spørgsmål.

Panelet må da også konstatere, at de polytome opgaver i gennemsnit har mindre sværhedsgrader end opgaver med kun et spørgsmål. Endvidere er det i henhold til panelet et problem, at den simple Rasch-model i mange (sandsynligvis de fleste) af de polytome opgaver ikke holder over for rækken af spørgsmål. Panelet konkluderer side 66 -67 følgende om polytome opgaver:

”Anvendelsen af denne type opgaver viser sig dog at være problematisk. Dels er mange polytome opgaver blevet afvist under afprøvningen, og dels har mange af de godkendte opgaver problemer med afhængighed mellem items og/eller items af varierende sværhedsgrad. Det er problematisk, for ikke at sige umuligt, at udlede en faglig fortolkning af besvarelsen af denne type opgaver, som vil være til nytte for læreren.

Panelet vurderer samtidig, at opgaver med mange items ikke er særligt velegnede set fra elevens synspunkt, eller særligt hensigtsmæssige i forhold til ønsket om at teste så mange trinmål/emner som muligt under testen.

Psykometrisk set er det tillige et problem, hvis der er lokale afhængigheder mellem items i en opgave, ligesom det er at foretrække, hvis de enkelte items hver især er homogene i forhold til opgavebanken.

På den anden side kan polytome opgaver give et større bidrag til afklaring af elevens dygtighed end et tilsvarende antal items i dikotome opgaver, forudsat at de har en passende sværhed for eleven. Anvendes opgaverne på en sådan måde, at de enten er for lette eller for svære for eleven, kan de derimod være mindre effektive end dikotome opgaver i denne henseende. Anvendelsen af dikotome opgaver er derfor stærkt afhængige af en velfungerende adaptiv procedure.

Da panelet vurderer, at det er vanskeligt at konstruere en serie items til en bestemt, ensartet sværhedsgrad, som samtidigt ikke er lokalt afhængige, forventer panelet, at der også i kommende afprøvninger vil skulle kasseres relativt mange polytome opgaver.”³

3.3 Opgavevarianter

Som anført var det kun halvdelen af de 760 opgaver i maj-juni testen, der var selvstændige opgaver, idet den anden halvdel af opgaverne var varianter af en allerede stillet opgave. Panelet anbefaler i lyset heraf, at antallet af variantopgaver reduceres så meget, at en elev aldrig møder to eller flere varianter af samme grundopgave i et testforløb. Dette problem var især stort i Fysik/Kemi.

3.4 Andre opgaveproblemer

Her skal kun nævnes en række andre mindre væsentlige problemer, som der relativt let kan rådes bod på. Disse problemer er mest interessante derved, at deres tilstedeværelse viser, at der har været en udbredt mangel på projektstyring fra COWI's side.

Opgaverne er ikke fordelt optimalt over sværhedsgrad, trin- og slutmål m.v., hvorfor udnyttelsesgraden ikke afspejler elevfordelingen efter færdigheder (side 33). Panelet finder endvidere brugen

³ Items er det samme som spørgsmål. Dikotome opgaver er opgaver med kun et spørgsmål.

af distraktorer (forkerte svarmuligheder) problematisk (side 49). Overskrifter til opgaverne er ikke konsekvent udformet (side 50). Uklare illustrationer, sprog og formuleringer bør fjernes (side 50). Scrolling bør undgås (side 50). Metadata anvendes usystematisk og inkonsekvent (side 51). For få trinmål indgår i testforløbene (side 51).

4. Den adaptive mekanisme

Selv om panelet som anført anbefaler en adaptiv mekanisme, er der store problemer ved den implementering heraf, der blev anvendt i maj-juni 2007

Panelet har eksempelvis fundet testforløb, der tyder på, at den adaptive algoritme ikke i alle tilfælde har udvalgt opgaver på en optimal måde (side 42). Den adaptive algoritme har således ikke sikret en tilstrækkelig spredt anvendelse af opgaverne (side 35).

Der er også en alt for høj hyppighed af svarekvenser, der har en *lille sandsynlighed*. I Dansk/Læsning og Matematik er forekomsten af svarekvenser med lille sandsynlighed over 25 pct. og i Fysik/Kemi helt op til 40 pct. afhængigt af, hvor lange svarekvenser, der betragtes (side 29). Undertiden kan der ligefrem konstateres ganske bemærkelsesværdige svarekvenser, der strider mod den procedure, der skulle være implementeret til skift af sværhedsgrad (side 31). Den alt for hyppige forekomst af sekvenser med lille sandsynlighed og bemærkelsesværdige svarekvenser er givetvis en følge af for få opgaver i opgavebanken kombineret med fravær af en stopmekanisme, når en elev er ved at tømme opgavebanken, samt i tilknytning hertil tilsidesættelse af proceduren for ændring af sværhedsgrad.

Den adaptive algoritme tager endvidere ikke højde for, om en elev tidligere har fået stillet en variant af samme opgave (side 43). Derudover sikrer den adaptive procedure heller ikke, at de forskellige trinmål er tilstrækkeligt repræsenteret i testene (side 51-53).

5. Øvrige problemer

5.1 Den tidsmæssige placering af de nationale test

Det blev i designet af det nationale it-testsystem besluttet, at de skulle gennemføres i slutningen af skoleåret. Og testen i skoleåret 2006/07 blev da også gennemført i maj-juni 2007. Det er imidlertid samtidigt tilkendegivet af såvel Undervisningsministeriet og -ministeren, at hovedformålet er at sikre et bedre fremadrettet forløb af elevens læringsproces, jf. bl.a. følgende formulering fra lovforslaget om indførelsen af de nationale it-test:

”Testene skal udgøre et værktøj, der kan medvirke til et godt overblik over den enkelte elevs faglige niveau og derigennem bidrage til grundlaget for faglige fremskridt for den enkelte elev ved at målrette undervisningen til elevens behov og forudsætninger. Testene anvendes sammen med andre elementer som led i den løbende evaluering fremadrettet til brug i vejledningen af den enkelte elev, den videre planlægning af undervisningen og underretning af forældrene med henblik på at tilrettelægge en undervisning og et forældresamarbejde, der understøtter eleven bedst muligt.”

Men hvis dette er det vigtigste, er det vanskeligt eller rettere umuligt at forstå, at testene ikke gennemføres i starten af skoleåret, så læreren hurtigt kunne få et indblik i sine elevers standpunkt. Det er jo endvidere ikke sikkert, at det er samme lærer, der fortsætter undervisningen i det testede

fag på næste trin. Og da testresultaterne kun må kendes af læreren, der har undervisningen på det pågældende klassetrin, og skolelederen, kan indsigten fra testresultaterne ikke videreføres til en ny lærer, hvis reglerne skal overholdes. Derfor kan det med rette spørges, hvor meget af det mulige fremadrettede, der i realiteten kan udnyttes ved en placering af testene i slutningen af skoleåret.

Den tidsmæssige placering af testene har naturligvis også betydning for de opgaver, der kan stilles, idet det relevante pensum jo skal være gennemgået, før det har mening af teste opgaver heri. Derfor vil det give betydelige problemer, såfremt testene ikke afholdes i enten starten eller slutningen af skoleåret, da rækkefølgegangen kan variere fra skole til skole. Trods dette, er det besluttet, at de obligatoriske test i de tre fag, der blev testet i maj-juni 2007, i skoleåret 2007/08 skal afholdes i marts-april 2008. Dette vil helt givet give ekstra problemer, som det også indirekte er gjort opmærksom på af panelet (side 64).

5.2 Sammenlignelighed over tid

Selv når opgavebanken har fået en tilstrækkelig størrelse, er det forventningen, at der hvert år skal tilvejebringes mellem 15-20 pct. nye opgaver. Dette medfører, som også panelet fremfører (side 63), at der vil være betydelige problemer med at sammenligne elevernes præstationer fra år til år. Derfor opfordrer panelet også til, at der tages stilling til, om der som ønsket skal sammenlignes over tid, idet dette yderligere vil øge kravene til opgaver i it-systemets opgavedatabank.

5.3 Skævvridning af undervisningen

Test giver ikke blot muligheder, selv når de er udformet på bedst mulig måde, men også problemer. Et gennemgående problem er, at undervisningen kan blive rettet imod, at eleverne opnår de højeste resultater i testene. Dette er naturligvis kun godt, hvis testene nu også afprøvede det, der var det væsentligste i læringsprocessen. Det er i afsnit 3.1 kort redegjort for, at det *ikke er tilfældet* for det nationale it-testsystem.

Risikoen for en u hensigtsmæssig skævvridning af undervisningen forstærkes naturligvis, hvis der rangordnes mellem klasser/skoler, og hvis lærerne ved eller kan regne ud, hvad der testes i. Begge disse forhold kendetegner det valgte design.

Skoleledere (skolebestyrelsen), kommuner (skoleudvalgsmedlemmer) og Undervisningsministeriet vil modtage rapporter, der rangordner skolerne i forhold til landsgennemsnittet. Rangordningen bliver præciseret ved, at elevernes præstationer bliver korrigeret for social baggrund (forældres uddannelsesbaggrund og indkomstforhold, etnicitet, køn m.v.). Om disse rapporter, de såkaldte nationale præstationsprofiler, er det dog fastsat i den grundlæggende aftale af 23. sep. 2005 om de nationale test:

”Oplysninger om testopgaver og testresultater er fortrolige og må ikke offentliggøres. Uberettiget videregivelse eller udnyttelse af oplysninger om testresultater kan straffes i medfør af straffelovens regler om tavshedspligt.”

Givet den relativ store modtagerkreds kan det betvivles, at dette kan overholdes.

Men måske er der – i hvert fald på kort sigt – ingen grund til at beskæftige sig med evt. følger af rangordning. Panelet påpeger således (side 36-37), at de nationale it-test fra maj-juni 2007 kun i

begrænset omfang kunne rangordne skolerne, selv om dette ellers som i f.eks. PISA-test er normen. Dette er endnu et eksempel på de problemer, som det nationale testsystem står over for, hvis det ønskede system altså skal realiseres.

Det kan derimod ikke betvivles, at læreren får kendskab til opgaver i it-systemets databank. Panelet vurderer – forstærket af den alt for lille opgavedatabank, hvor der endvidere er mange varianter – at lærere ”forholdsvis let vil kunne huske en række af problemstillingerne og følgende år gøre ekstra meget ud af disse i undervisningen.” (side 43) Lærerens kendskab vil dels hidrøre fra tilstedeværelse under testafviklingen og fra de testrapporter med spørgsmål og svar, som de modtager for samtlige de af deres elever, som deltager i en test.

5.4 Ingen sikker vurdering af elevernes færdigheder

Panelet anfører (side 63), at der (endnu) er så mange problemer ved det nationale it-testsystem, at der er stor risiko for fejlklassifikationer af elevernes færdigheder på færdighedsskalaen med fem trin (1-2-3-4-5-skalaen). I værste fald kan det være tilfældigt, om en elev indplaceres på trin to eller trin 4 på denne skala.

6. Konklusion

REVIEW-panelets opgave var begrænset til to forhold. For det første skulle panelet ud fra især erfaringerne fra maj-juni 2007 testen komme med forslag til ændringer, der kunne sikre en bedre opfyldelse af målet med de it-baserede test. I relation hertil har panelet fokuseret på kvaliteten af opgavebanken og funktionaliteten af den adaptive opgaveudvælgelse. For det andet skulle panelet bidrage til at afdække, i hvilket omfang konsortiets leverance har levet op til de i kontrakten med Ministeriet stillede krav. Panelet skulle således ikke vurdere den it-mæssige eller den vejledningsmæssige side af testsystemet, ligesom omkostninger ved og effekter af systemet ikke skulle vurderes.

Panelets bidrag til det første punkt er forsøgt sammenfattet i det ovenstående. Det fremgår heraf, at der er rigtig meget at tage fat på, såfremt blot det tilsigtede paratvidenssystem skal leve op til de stillede krav. For som direktør fra Skolestyrelsen Martin Isenbecker sagde i den i indledningen anførte pressemeddelelse: ”Forskernes evaluering har sat en tyk streg under, at vi ikke er i mål endnu.” Og det er ganske vist! Noget andet er, om det overhovedet er fremmende for læreprocessen i dette land, at elever i grundskolealderen nu skal igennem sådanne paratvidensforløb. Er der virkelig fremtid i dette i et google-samfund?

Panelets bidrag til det andet punkt er ikke offentlig tilgængelig, idet modtageren her er styregruppen for projektet. Uden at kende til indholdet er det nok ikke særlig voveligt at antage, at dette bidrag må kunne bruges til en ordentlig rabat!

Litteratur:

REVIEW-panelet 2007. *De nationale it-baserede test i folkeskolen*. 29. oktober 2007.

Sørensen, C. 2006. De nationale it-baserede obligatoriske test – frelser eller uhyre? *Discussion Papers on Business and Economics No. 11/2006*. Institut for Virksomhedsledelse og Økonomi. Syddansk Universitet.

Sørensen, C. 2007. De nationale it-baserede obligatoriske test - endnu et flop?. *Samfundsøkonomen* 2007:2:15-19.

Undervisningsministeriet 2005. *Forslag til Lov om ændring af lov om folkeskolen*. København.

Undervisningsministeriet 2006. *Bilag 2. Kravspecifikation og testafvikling*. København.

Efterskrift

Denne artikel blev skrevet, før undervisningsminister Bertel Haarder udskød de obligatoriske nationale test et år, jf. Politiken 26. jan. 2008. Det er positivt, at Bertel Haarder dog har haft modet til at udsætte dette for især ham meget prestigefyldte projekt. Hermed undgås i det mindste, at vores skolebørn endnu engang bliver brugt som prøveklude. Men Bertel Haarder er dog endnu ikke kommet så langt, at han kan se, at der er gjort noget forkert. ”Man kan altid være bagklog, men en gang skal jo være den første, og hvis vi ikke havde prøvet sidste år, havde vi ikke fået alle de gode råd til forbedringer, vi arbejder videre med nu”, jf. Bertel Haarder til Politiken 26. jan. 2008., idet han tilsyneladende også har glemt, at der på forhånd blev advaret mod en så hovedkulds iværksættelse af de nationale it-test.

Efter dette ”fantastiske” udsagn forstår jeg i øvrigt meget bedre, at så mange it-projekter i relation til politi, kommunesammenlægning mv. er så fejlbehæftede, at de mest af alt forhindrer, at nødvendigt arbejde udføres. Det er jo åbenbart ikke meningen, at de skal fungere fra starten. Der skal først indhentes ideer til forbedringer! Og så kan jeg også meget bedre forstå de øvrige fejlslagne projekter fra undervisningsministeriet så som gymnasireformen mv.