

De nationale it-baserede obligatoriske test – frelser eller uhyre?

af

Christen Sørensen*

Discussion Papers on Business and Economics
No. 11/2006

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

Tel.: +45 6550 3271

Fax: +45 6615 8790

E-mail: lho@sam.sdu.dk

<http://www.sam.sdu.dk/depts/virkl/about.shtml>

ISBN 87-91657-10-5

* Christen Sørensen er professor i økonomi ved Institut for Virksomhedsledelse og Økonomi

De nationale it-baserede obligatoriske test – frelser eller uhyre?

Af: Christen Sørensen

Abstract

Following PISA 2003, which showed lower than expected/acceptable results for the achievements of Danish students, a majority in the Danish Parliament decided in September 2005 to introduce ten national compulsory tests of student performance at various class levels in, respectively, Danish (4), English (1), mathematics (2), geography (1), biology (1) and chemistry/physics (1). This test system, which will be introduced in mathematics, Danish and chemistry/physics in the current school year 2006/07, is designed as an adaptive system on an IT platform. The adaptive feature implies that a student will meet increasingly difficult questions when a number of correct answers are given. In contrast, the difficulty of questions will be lowered, if a student cannot answer a number of questions. Both the mentioned features were emphasized as being essential.

Even though OECD in PISA 2003 found negative, yet insignificant, effects of standardised testing, an implementation of the aforementioned system was decided on without any serious further investigations. However, the purpose of this article is not to explore that question, but rather to display the design of the coming national system.

Three essential features imply that the system will centre on ready knowledge (lexical knowledge) and not on students' ability to reflect on or to solve problems: 1) an average student must manage to answer 50-60 questions (items) in 45 minutes, 2) the questions are framed as multiple choice questions and 3) the IT medium will encourage a quick response. The adaptive feature and the necessary number of questions in evaluating a student's possible ranking requires around 700-900 questions in each of the 10 tests. Furthermore, the multiple choice design etc. will necessitate that many new questions will have to be prepared every year, although secrecy by law is demanded in relation to the questions.

This article poses the following questions: 1) is it conducive to learning in a knowledge society to test student performance by lexical knowledge? 2) was the majority in Parliament aware that an essential feature of the coming national system was to test lexical knowledge? 3) why did Parliament not ask for an in-depth investigation of the Finnish school system? For at least two reasons: the Finnish school system has essential features in common with the Danish system (e.g. no grades in the lower class levels, no standardised test) and at the same time Finnish students have topped both in the PISA 2000 and the PISA 2003 assessments.

The author of this article – who has a rather negative opinion of the coming test system and especially of the analytical foundation for decision-making in relation to school policy at the central level in Denmark – suggests that the ability in the Danish Ministry of Education to supply the Government and the general public with reliable results is far from an acceptable standard. Several examples testify to this, as explained in the article.

JEL classification: I20 and I28.

Key words: Student performance, national tests, adaptive tests, multiple choice tests.

1. Indledning

Med vedtagelsen af lovforslag 101, hvorved folkeskoleloven blev ændret, besluttede et flertal i Folketinget (90 for, 23 imod, 66 fraværende)¹ 30. marts 2006 i forlængelse af forliget af 23. sep. 2005 mellem regeringen (Venstre og Det Konservative Folkeparti), Socialdemokraterne og Dansk Folkeparti, at der bl.a. skulle indføres obligatoriske, it-baserede nationale test i Folkeskolen.²

I figur 1 er det vist, hvordan disse test, der påbegyndes i skoleåret 2006/07, er fordelt på fag og klassetrin.³ Et tal i tabelkroppen angiver, at der er fastsat trin- eller slutmål på det med tallet angivne klassetrin. Et t i tabellen angiver, at der på det pågældende klassetrin er en it-baseret test. De it-baserede test skal teste en delmængde af læringsmålene i relation til de relevante trin- eller slutmål, jf. nærmere herom i afsnit 4.1. Figur 1 viser endvidere, at der for folkeskolens 23 fag er fastsat i alt 70 trin- eller slutmål.

Testdesignet for de it-baserede nationale test blev i hovedtræk fastlagt i februar 2006, jf. Undervisningsministeriet (2006). Selv om den konkrete udformning af designet ikke dermed blev endeligt fastsat i detaljen, er designet herved blevet så specificeret, at det er ret så klart, hvad der kommer til at karakterisere de it-baserede nationale test.

2. Afgrænsning og problemstilling

Formålet med det følgende er *ikke* at analysere, om obligatoriske nationale test medvirker til at forøge elevernes præstationer i internationale test som PISA-test eller i andre test. Her skal det blot anføres, at dette endnu er et uafklaret spørgsmål. I den originale PISA-2003 test, hvor hovedemnet var matematik, var konklusionen, at det ikke kunne vises, at nationale test forbedrede præstationerne i matematik, jf. OECD (2004). Bl.a. anvender topscorerlandet i PISA-undersøgelserne, Finland, ikke nationale obligatoriske test. Der er dog andre undersøgelser, der kunne tyde på, at nationale test kan forbedre elevernes præstationer ved test. Velfærdskommissionen (2006, side 257) anfører således med udgangspunkt i Fuchs og Woessmann (2004), at der er belæg for, at nationale test kan hæve det faglige niveau i grundskolen, hvis der er klarhed om formålet med undervisningen.⁴

¹ Jastemmerne hidrørte alle fra forligspartierne Venstre, Det Konservative Folkeparti, Socialdemokraterne og Dansk Folkeparti, mens nejstemmerne hidrørte fra Det radikale Venstre, SF og Enhedslisten samt fra 6 socialdemokrater.

² Friskolerne og øvrige grundskoler på folkeskoleniveau kan også benytte disse test, hvis de ønsker det.

³ Testene er blevet forsinket næsten et år, idet første udbudsrunde måtte opgives. Så vidt jeg kan vurdere, var årsagen hertil, at kravene til testsystemet ikke var tilstrækkeligt specificeret i første runde. Dette er bl.a. formuleret således: "På baggrund af den viden, Undervisningsministeriet har erhvervet i løbet af udbudsprocessen, og på baggrund af ministeriets seneste risikoanalyser, er det besluttet ikke at indgå kontrakt på nuværende tidspunkt. Udbuddet er derfor den 28. oktober 2005 blev aflyst med henblik på, at Undervisningsministeriet snarest muligt vil indlede et nyt EU-udbud med henblik på implementering af testene i 2007.", Undervisningsministeriet (2005) og videre "Baggrunden herfor var, at Undervisningsministeriets risikoanalyser, der er blevet gennemført sideløbende med tilbudsevalueringen, gav ministeriet anledning til at revurdere de tidsmæssige rammer for gennemførelsen, navnlig for implementeringen af bl.a. de it-baserede test i skolerne. Endvidere gav analyserne ministeriet anledning til nøje at overveje den overordnede økonomiske ramme for projektet.", Undervisningsministeriet (2006, side 5).

⁴ Og ikke nok hermed. Velfærdskommissionen (2006, side 249) anførte endvidere: "at nationale tests i udvalgte fag på en række klassetrin med målrettet opfølgning er en god idé." Dette blev anført, selv om Velfærdskommissionen, der afsluttede sit arbejde i 2005, endnu ikke vidste, hvordan det kommende nationale danske test-system ville blive udformet! Derfor står det også hen i det uvisse, om Velfærdskommissionen går ind for et paratvidens testsystem, som er blevet resultatet, jf. afsnit 2.

Figur 1. Oversigt over placeringen af de nationale it-baserede test på fag og klassetrin

Fag	Klassetrin										Antal målsatte trin
	1	2	3	4	5	6	7	8	9	10	
Dansk		2t		4t		6t		t	9	10	5
Engelsk				4			7t		9	10	4
Kristendomskundskab			3			6			9	10	4
Historie				4		6		8	9		4
Samfundsfag									9	10	2
Idræt		2			5		7		9	10	6
Musik		2		4		6					3
Billedkunst		2			5						2
Håndarbejde				(4	5	6	7) ^a				1
Sløjd				(4	5	6	7) ^a				1
Hjemkundskab				(4	5	6	7) ^a				1
Matematik			3t			6t			9	10	4
Natur/teknik		2		4		6					3
Geografi								8t	9		2
Biologi								8t	9		2
Fysik/kemi								8t	9	10	3
Tysk							7		9	10	3
Fransk							7		9	10	3
Dansk som andetsprog ^b		2			5t		7t		9	10	5
Færdselslære			3			6			9		3
Sundheds- og seksual- undervisning og familiekundskab			3			6			9		3
Uddannelses-, erhvervs- og arbejdsmarkeds- orientering			3			6			9		3
Modersmålsundervisning ^c			3			6			9		3
											70

Anm.: Et tal i tabellen angiver, at der er fastsat trinmål eller slutmål på det med tallet angivne klassetrin. I matematik er der således trin eller slutmål på 3, 6, 9 og 10 klassetrin. Et t i tabellen angiver, at der på det pågældende klassetrin er en it-baseret test. For matematik er der således en it-baseret test på 3 og 6 klassetrin.

- Der er kun fastsat slutmål, der dækker 4., 5., 6. og 7. klassetrin under et.
- Trinmål er ikke kun relateret til klassetrin. Test for dansk som andetsprog er ikke obligatoriske.
- Trinmål er ikke kun relateret til klassetrin.

Det afgørende og lidet flatterende for den hjemlige beslutningsproces er, at beslutningen om at iværksætte de obligatoriske nationale test blev truffet, uden at dette helt afgørende spørgsmål var blevet forsøgt besvaret af Undervisningsministeriet.⁵ Og der er – så vidt jeg er orienteret – heller ikke gjort forsøg på at sandsynliggøre, at det konkrete besluttede testdesign, som vil blive anvendt ved de it-baserede test i Danmark fra skoleåret 2006/07, vil føre til forbedrede præstationer uden at føre til forringelser på andre områder.⁶

⁵ Her og i det følgende angives Undervisningsministeriet som hovedansvarlig for, at der ikke tilvejebringes et tilstrækkeligt beslutningsgrundlag i relation til folkeskolen m.v. Det kan naturligvis diskuteres, om det ikke rettere var undervisningsministeren, regeringen eller det relevante flertal i Folketinget, der i en eller anden kombination sammen med Undervisningsministeriet bærer ansvaret herfor.

⁶ I skoleåret 2006/07 er det kun i flg. fag og klassetrin, at der gennemføres obligatoriske nationale it-test: matematik i 6. klasse, dansk/læsning i 8. klasse samt fysik/kemi i 8. klasse. Disse obligatoriske test skal gennemføres i perioden fra 1. maj til 29. juni 2007.

Formålet med det følgende er *derimod* at præsentere essensen i det it-baserede testdesign på en måde, der gør det muligt for bl.a. ikke-it-eksperter at diskutere dette. Det må jo være afgørende at analysere, hvordan den konkrete udformning af disse test vil påvirke læringssituationen i folkeskolen. Det grundlag, der lægges her, er jo helt afgørende for, hvordan forløbet vil forme sig i den videre skolegang. Dette vil derfor også være afgørende for, om velfærdsaftalens målsætning om, at mindst 95 pct. af en ungdomsårgang kommer igennem en ungdomsuddannelse senest med virkning fra 2015, kan realiseres.⁷ Og dette er endvidere helt afgørende for den enkeltes muligheder for at fastholde en tilknytning til arbejdsmarkedet i det sædvanligt erhvervsaktive aldersinterval, hvilket igen er en nødvendig forudsætning for at kunne finansiere det danske velfærdssamfund uden et alt for højt skattetryk. Derfor er det selvfølgelig helt afgørende, hvordan vor folkeskole fungerer og derfor også afgørende, at de rammer, der sættes herfor, fastlægges på et oplyst grundlag og er med til at bidrage til et optimalt læringsforløb.

3. Paratviden vil blive testgrundlaget

Det ligger som anført allerede klart, hvad der kommer til at karakterisere det nationale it-baserede testsystem. Det helt essentielle er – efter min vurdering – at det bliver et system til test af paratviden. Dette resultat følger overordnet af to grundlæggende krav til systemet. For det første skal gennemsnitseleven kunne besvare mindst 50-60 spørgsmål på 45. min., jf. Undervisningsministeriet (2006, side 62). Og systemet skal for det andet operere som et multipel choice system med kun to testvurderinger: korrekt og ikke-korrekt.⁸ Hertil kommer, at selve it-mediet ansøger til hurtig respons, som personer, der arbejder med it, ved alt for godt. Det siger sig selv, at disse to krav indbygget i et it-medie i realiteten kun gør det muligt at stille paratviden-spørgsmål, idet tiden til at fundere over et spørgsmål er yderst begrænset.⁹ I denne sammenhæng er det vigtigt at bemærke, at de ellers meget kritiserede PISA-spørgsmål var af en helt anden kontemplativ karakter (lagde op til fundering).¹⁰

Selv om det bliver paratviden, der bliver grundlaget for testresultaterne, er det vigtigt at bemærke, at Undervisningsministeriet ikke ønsker, at de it-baserede test skal føre til, at læringsfokus i folkeskolen trækkes i retning af, at eleverne i højere grad bliver omvarende mini-leksikoner. Gang på gang understreges – nogen vil sikkert sige besværges – at hovedformålet er at sikre et bedre videre forløb, jf. bl.a. følgende formulering:

⁷ Dette blev besluttet i Velfærdsforliget af 20. juni 2006 mellem regeringen (Venstre og Det Konservative Folkeparti), Dansk Folkeparti, Det radikale Venstre og Socialdemokraterne.

⁸ Derudover kan der naturligvis også undlades at besvare et spørgsmål. Ikke-besvarede spørgsmål og forkert besvarede spørgsmål får begge samme scoringsværdi, nemlig nul. Multipel choice grundlaget fremgår foruden af opdelingen i tre svarkategorier (korrekt, ikke-korrekt, ikke-besvaret) af bl.a. følgende: "Fx bør der være opgaveskabeloner, der skal besvares med 'peg og klik', farvelæg område, skriv tal, vælg rigtig knap, osv.", Undervisningsministeriet (2006, side 58).

⁹ Et spørgsmål kan give anledning til flere delspørgsmål (også benævnt items). Dette kan naturligvis muliggøre, at der bliver lidt mere tid til fundering, men vil – efter min vurdering – ikke forhindre, at det primært bliver paratviden, der bliver testgrundlaget. I det følgende sondres ikke mellem spørgsmål med og uden delspørgsmål. Et spørgsmål med eksempelvis fem delspørgsmål (items) tæller derfor i det følgende som fem spørgsmål.

¹⁰ Som det fremgår af Sørensen (2005), har jeg ikke deltaget i den principielle kritik af disse PISA-2003 spørgsmål, om end der også efter min opfattelse har været anvendt spørgsmål, der burde være formuleret klarere. Hovedproblemet ved PISA-2003 var ikke OECD-spørgsmålene eller OECD-rapporten, men den hovedløse modtagelse af PISA-2003 i Undervisningsministeriet.

”Testene skal være et værktøj til at sikre faglige fremskridt for den enkelte elev ved at målrette undervisningen til den enkelte elevs særlige evner.”, Undervisningsministeriet (2006, side 13).

og videre:

”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”, Undervisningsministeriet (2005).

Det er efter min opfattelse centralt at analysere og drøfte, om det overhovedet er muligt eller sandsynligt, at dette ønske fra Undervisningsministeriets side kan realiseres givet det besluttede testdesign. Men denne væsentlige udredning er *ikke* målet for dette papir, idet jeg ikke mener at have kompetence til en sådan udredning. Til gengæld mener jeg, at jeg kan blotlægge grundstrukturen i det it-design, der må være forudsætningen for at kunne gennemføre en sådan analyse på et oplyst grundlag. Derfor er analysen af grundstrukturen i det kommende it-test-system hovedemnet for dette papir. Formen sigter på, at også ikke-it-eksperter får forståelsen for dette grundsystem uden at blive overvældet eller fortabt i it-tekniske termer el.lign.

4. Designet af it-testsystemet

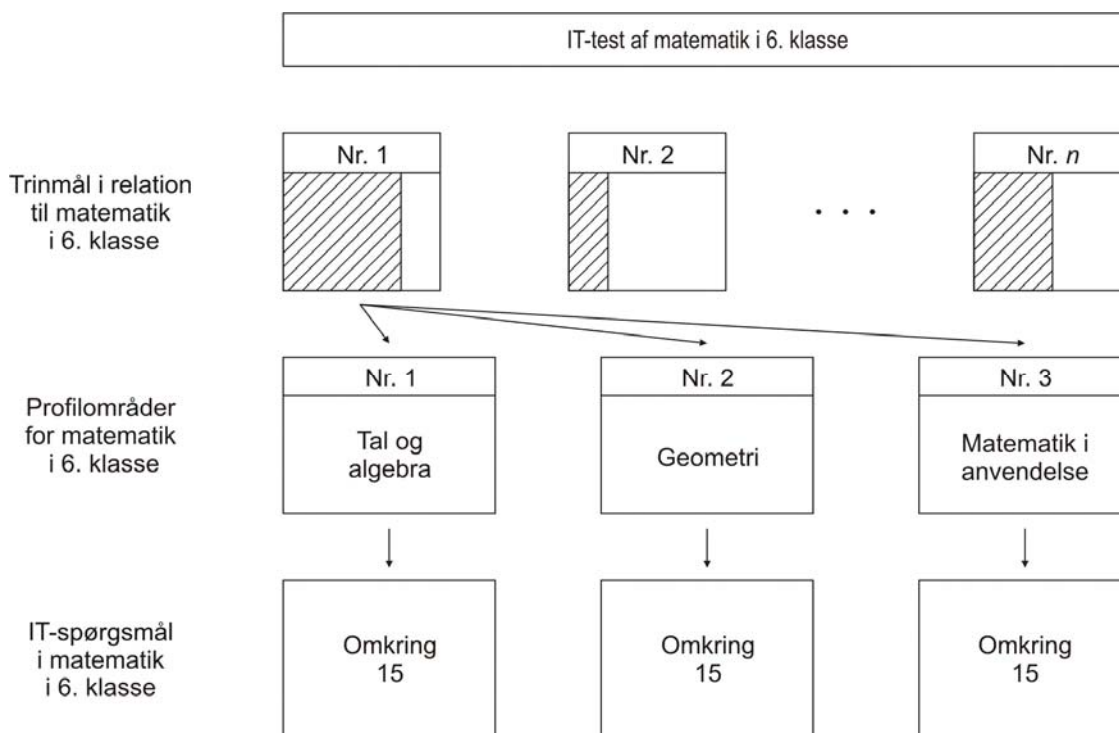
Den følgende analyse og beskrivelse af designet af it-testsystemet i det kommende nationale it-testsystem analyserer først i afsnit 4.1 sammenhængen mellem på den ene side de trin- og slutmål, der blev indført for folkeskolens fag i 2003 og på den anden side de profilområder og it-spørgsmål, der bliver kendetegnende for it-testsystemet. Herefter ses i afsnit 4.2 på kravene til det antal spørgsmål pr. test, der er nødvendige i et adaptivt testsystem, hvis der som ønsket skal gives blot nogenlunde valide og reliable standpunktvurderinger af den enkelte elevs kundskaber og færdigheder på de områder, der indgår i it-testene. Det såkaldte sværhedstal, der knyttes til og karakteriserer de enkelte spørgsmål, spiller en afgørende rolle i opbygningen af it-testsystemet. Hvordan sværhedstallet fastsættes og proceduren for afprøvning af nye it-spørgsmål er hovedemnet i afsnit 4.3. I afsnit 4.4 beskrives i hovedtræk, hvordan tilbagemeldingen til elever, skoleledere, kommuner og Undervisningsministeriet tænkes udformet. Afsnit 4 afsluttes med en kort redegørelse i afsnit 4.5 for de forventede udviklings- og driftsomkostninger i indkøringsperioden frem til og med 2009 for de samlede initiativer til at fremme evalueringskulturen i folkeskolen, herunder til it-systemet.

4.1 Trin- og slutmål, profilområder og testspørgsmål

Som det fremgår af figur 1, indføres der i folkeskolen 10 obligatoriske it-test (4 i dansk, 2 i matematik og 1 i hhv. engelsk, geografi, biologi og fysik/kemi) og 2 frivillige it-test i dansk som andetsprog.

Læringsspørgsmålene, som eleverne skal besvare i it-testene, er relateret til de for fagene fastsatte trin- og slutmål via profilområder. Denne sammenhæng, der gælder for alle 12-test, er skematisk skitseret i figur 2 med matematik i 6. klasse som eksempel.

Figur 2. Trin- og slutmål, profilområder og it-spørgsmål



Trin- og slutmål blev i 2003 indført for folkeskolens fag med bekendtgørelse nr. 571 af 23. juni 2003. Slutmålene og trinmålene 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 formål 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 evaluering af undervisningen, som dialogværktøj og som pejlemærker i relation til den løbende vurdering af elevens udbytte af undervisningen.

Det overordnede formål med faget matematik i folkeskolen er formuleret således:

”Formålet med undervisningen i matematik er, at eleverne bliver i stand til at forstå og anvende matematik i sammenhænge, der vedrører dagligliv, samfundsliv og naturforhold. Analyse og argumentation skal indgå i arbejdet med emner og problemstillinger.

Stk. 2. Undervisningen tilrettelægges, så eleverne opbygger matematisk viden og kunnen ud fra egne forudsætninger. Selvstændigt og i fællesskab skal eleverne erfare, at matematik både er et redskab til problemløsning og et kreativt fag. Undervisningen skal give eleverne mulighed for indlevelse og fremme deres fantasi og nysgerrighed.

Stk. 3. Undervisningen skal medvirke til, at eleverne oplever og erkender matematikkens rolle i en kulturel og samfundsmæssig sammenhæng. Med henblik på at kunne tage ansvar og øve indflydelse i et demokratisk fællesskab, skal eleverne kunne forholde sig vurderende til matematikkens anvendelse.”, se www.faellesmaal.uvm.dk.

Trinmålene for matematik i 6. klasse er formuleret således:¹¹

”Arbejde med tal og algebra

Undervisningen skal lede frem mod, at eleverne har tilegnet sig kundskaber og færdigheder, der sætter dem i stand til at

- kende til de hele tal, decimaltal og brøker
- benytte erfaringer fra hverdagen sammen med arbejdet i skolen ved opbygningen af talforståelse
- kende tallenes ordning, tallinjen, positionssystemet og de fire regningsarter
- benytte hovedregning, overslagsregning og skriftlige udregninger
- anvende lommeregner og computer ved gennemførelse af beregninger
- arbejde med optællinger og eksempler på sammenhænge og regler inden for de fire regningsarter
- kende til eksempler på brug af variable, herunder som de indgår i formler, enkle ligninger og funktioner
- kende til procentbegrebet og forbinde begrebet med hverdagserfaringer
- regne med decimaltal og benytte brøker knyttet til procent og konkrete sammenhænge
- arbejde med "forandringer" og strukturer, som de indgår i bl.a. talfølger, figurrækker og mønstre
- kende til koordinatsystemet og herunder sammenhængen mellem tal og tegning.

Arbejde med geometri

Undervisningen skal lede frem mod, at eleverne har tilegnet sig kundskaber og færdigheder, der sætter dem i stand til at

- benytte geometriske metoder og begreber i beskrivelse af fysiske objekter fra dagligdagen, herunder figurer og mønstre
- undersøge og beskrive enkle figurer tegnet i planen
- kende til grundlæggende geometriske begreber som vinkler og parallelitet
- arbejde med fysiske modeller og enkle tegninger af disse
- kende til forskellige kulturers metoder til at angive dybde i billeder
- undersøge de enkelte tegnemethoders anvendelighed til beskrivelse af form og afstand
- måle og beregne omkreds, areal og rumfang i konkrete situationer
- tegne, undersøge og eksperimentere med geometriske figurer, bl.a. ved at benytte computer.

Matematik i anvendelse

Undervisningen skal lede frem mod, at eleverne har tilegnet sig kundskaber og færdigheder, der sætter dem i stand til at

- vælge og benytte regningsarter i forskellige sammenhænge
- anvende og forstå enkle informationer, som indeholder matematikfaglige udtryk

¹¹På netstedet www.faellesmaal.uvm.dk er der bl.a. en oversigt over samtlige trin- og slutmål for folkeskolens fag.

- anvende faglige redskaber, herunder tal, grafisk afbildning og statistik, til løsningen af matematiske problemstillinger fra dagligliv, familieliv og det nære samfundsliv
- arbejde med enkle procentberegninger, herunder ved rabat Køb
- beskrive og tolke data og informationer i tabeller og diagrammer
- indsamle og behandle data samt udføre simuleringer, bl.a. ved hjælp af en computer
- foretage eksperimenter, hvori tilfældighed og chance indgår.

Kommunikation og problemløsning

Undervisningen skal lede frem mod, at eleverne har tilegnet sig kundskaber og færdigheder, der sætter dem i stand til at

- kende til eksperimenterende og undersøgende arbejdsformer
- beskrive løsningsmetoder gennem samtaler og skriftlige notater
- opstille hypoteser, og efterfølgende ved at "gætte og prøve efter" medvirke til at opbygge faglige begreber og indledende generaliseringer
- formulere, løse og beskrive problemer og i forbindelse hermed anvende forskellige metoder, arbejdsformer og redskaber
- samarbejde med andre om at anvende matematik ved problemløsning
- undersøge, systematisere og begrunde matematisk ud fra arbejde med konkrete materialer.”, se www.faellesmaal.uvm.dk.

Det er disse trinmål, der er opdelt i et antal – n – i figur 2.

For hvert fag, der indgår i en it-baseret national test, skal der formuleres to til fire profilområder, som led i opbygningen af it-designet. De profilområder, der kommer til at indgå for matematik i 6 klasse, er – som det også fremgår af figur 2 – fastsat til: tal og algebra, geometri samt matematik i anvendelse. Det er ikke blot for matematik i 6. klasse, der vil blive anvendt 3 profilområder. Det gælder også for dansk/læsning i 8. klasse samt for fysik/kemi i 8. klasse. Generelt forventes det, at 3 profilområder bliver det typiske resultat.

Trin- og slutmålenes kundskabs- og færdighedsstandarder i relation til læring skal fordeles på profilområder. Dette er en følge af, at it-testene ikke blot skal lede frem til en vurdering af elevens standpunkt i faget, men også i fagets profilområder.¹² I figur 2 er denne fordeling kun vist for et mål – nr. 1 – for ikke at overbrodere figuren med forbindelseslinier fra målene til profilområder. Det er i figur 2 antaget, at mål 1 har indvirkning på alle tre profilområder. Som det fremgår af figur 2, er det kun en del – og endvidere en varierende del – af trin- og slutmålene, der påvirker de profilområder, der indgår i relation til udformningen af it-spørgsmålene. Dette skyldes, at den valgte evalueringsform i it-designet – multipel choice formen – ikke gør det muligt at vurdere læringsmålene i relation til en række aspekter af trin- og slutmålene. Dette er bl.a. formuleret således:

”De it-baserede test kan kun omfatte en del af fagets trinmål. For en dækkende vurdering af elevens kundskaber og færdigheder inden for alle fagets trinmål må yderligere evalueringsmetoder også tages i brug. Testene kan således ikke stå alene – heller ikke for de trinmål, som testen omfatter.”, Undervisningsministeriet (2006, side 48).

¹² Hvis en elev besvarer for få spørgsmål inden for et profilområde, vil it-systemet ikke levere en standpunktvurdering for denne elev i det pågældende profilområde, men besvarelsen af disse spørgsmål vil indgå i standpunktvurderingen for hele faget.

Og videre:

”Der vil blive gennemført supplerende undersøgelser, der har til formål at belyse elevernes kundskaber og færdigheder i udvalgte fag og klasetrin på områder (trinmål), der ikke er omfattet af testene. Det gælder fx trinmål, som retter sig mod, at eleven demonstrerer evne til selvstændig formulering, analyse, problemorientering og -løsning mv.”, Undervisningsministeriet (2006, side 18).

Hvordan disse supplerende undersøgelser skal gennemføres er der endnu ikke taget beslutning om!

Selvom en række (væsentlige) aspekter således på forhånd er holdt uden for den evaluering, der gennemføres med it-test, forudsættes det alligevel, at it-testene giver en *valid og reliabel* vurdering af de enkelte elevers standpunkt, som skal fremgå af den tilbagemelding, som den enkelte elev skal have efter en obligatorisk test i et fag via it-systemet, se nærmere herom i afsnit 4.4.

Dette er bl.a. formuleret således:

”Tjenesteyder skal desuden komme med sit begrundede forslag til, hvordan de it-baserede test kan tilrettelægges, således at den enkelte tests resultat udgør en valid og reliabel måling af elevens standpunkt samt kan danne grundlag for en tilbagemelding i forhold til profilområder..., således at læreren i videst muligt omfang direkte kan drage nytte af tilbagemeldingen ved tilrettelæggelse af den videre undervisning.”, Undervisningsministeriet (2006, side 53).¹³

Det er vel en selvfølge, at et mekanisk it-system kun kan producere tilbagemeldinger, der kan udledes fra det input, der indgår i systemet, hvorfor det valide og reliable i hvert fald kun kan omfatte den delmængde af trin- og slutmålene, der er inddraget i systemet, jf. bl.a. figur 2. Den it-baserede tilbagemelding vil derfor i bedste fald kun kunne give en reduktionistisk evaluering af den enkelte elevs kundskaber og færdigheder.

It-spørgsmålene inden for hvert profilområde vil blive opdelt i 5 sværhedsniveauer, idet de bl.a. vil blive tilordnet et såkaldt sværhedstal, der anvendes til denne opdeling i 5 niveauer.

Et spørgsmåls sværhedstal vil blive normeret efter en numerisk skala fra 0 til 100. Et sværhedstal på nul vil blive givet til et spørgsmål, som alle elever på det pågældende klasetrin kan besvare korrekt. Et sværhedstal på 100 karakteriserer en opgave, som ingen elever på klasetrinet kan løse korrekt. Et spørgsmåls sværhedstal skal derfor fastsættes ud fra den andel af eleverne på det pågældende klasetrin, der korrekt kan besvare spørgsmålet. Dette rejser et interessant problem, idet designet med spørgsmål efter multipel choice konceptet betyder, at der også vil blive givet korrekte svar på spørgsmål med sværhedsgrad 100, selv om ingen af disse elever kan redegøre for den korrekte løsningsmåde for spørgsmålet. Antag, at der til et spørgsmål er knyttet fem svarmuligheder. Multipel choice designet betyder som udgangspunkt, at omkring 1/5 af besvarelserne af et spørgsmål med fem svarmuligheder, som ingen af eleverne kan redegøre for den korrekte løsningsmåde af, alligevel rent tilfældigt vil blive besvaret korrekt i den aflestning af spørgsmålet, der fører frem til fastsættelsen af sværhedstallet, se også afsnit 4.3. Det er endnu ikke offentliggjort, hvilken algoritme eller regneregul, der vil blive anvendt til at gå fra den relative andel af korrekte

¹³ Tjenesteyder er betegnelsen for det firma/konsortium, der skal udvikle it-testsystemet. Tjenesteyder er COWI A/S, der har dannet et konsortium med en række erfarne underleverandører, jf. pressemeddelelse af 26. juni 2006 fra Undervisningsministeriet.

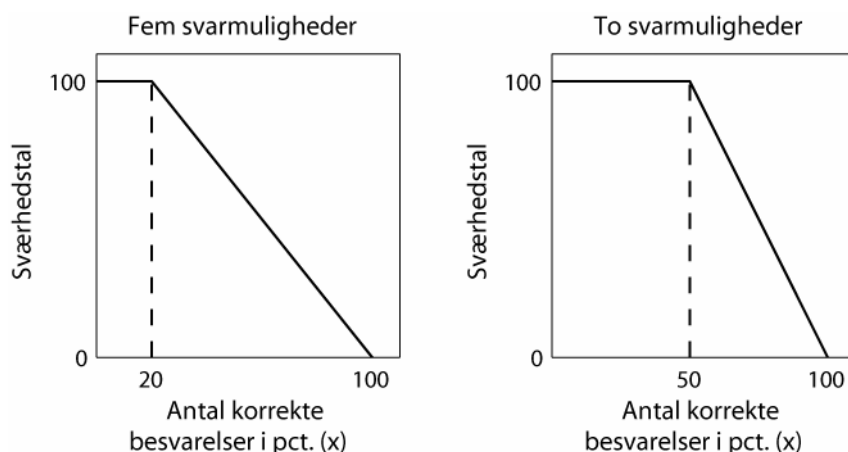
sværhedstallet. Den i det følgende skitserede algoritme eller regneregul er derfor kun en blandt flere mulige.

Antag, at den andel, der besvarer et spørgsmål korrekt, er x pct. x varierer altså mellem 0 og 100. Og antag, at der er y svarmuligheder for det pågældende spørgsmål. Følgende regneregul vil derfor kunne oversætte den korrekte besvarelsesprocent x fra et spørgsmål med y svarmuligheder til sværhedstallet, givet at korrekte svar kan skyldes rene tilfældigheder:

$$\text{sværhedstal} = \begin{cases} 100, & \text{hvis } x \leq 100/y \\ (100-x) \cdot \frac{100}{100-100/y}, & \text{hvis } x > 100/y \end{cases}$$

Den sammenhæng, der er mellem sværhedstallet og den korrekte andel af besvarelsen målt i procent som følge af denne regneregul, er også skitseret i figur 3 i et tilfælde med hhv. 5 og 2 svarmuligheder, hvorfor $100/y$ hhv. er 20 og 50.

Figur 3. Mulig sammenhæng mellem andel af korrekte besvarelser af et spørgsmål med hhv. fem og to svarmuligheder og sværhedstallet



Anm.: Betegner y antallet af svarmuligheder, er knækket på kurverne givet ved x -værdien: $100/y$. Når y er 5, er knækket altså givet ved en x -værdi på 20. Når y er lig 2, er knækket givet ved en x -værdi på 50.

Sværhedstallet benyttes foruden til scoring af elevens standpunkt, jf. nærmere herom i afsnit 4.3, også til opdeling af spørgsmålene i fem sværhedsniveauer, jf. følgende formulering:

”Sværhedsniveauet af et item er en klassifikation på baggrund af sværhedstallet. Opgaverne, der skal indgå i testene, skal fordele sig omtrentlig symmetrisk med middelsværhedstal på ca. 50. Klassificeringen sker på en 5-trinsskala, hvor opgaverne efter sværhedstal fordeles i 5 grupper ..., således at opgaverne med sværhedsniveauet 1 udgør de ca. 20% letteste (målt ved sværhedstallet), og opgaverne med sværhedsniveauet 5 udgør de ca. 20% sværeste. Sværhedsniveauet (samt evt.

sværhedstallet) bruges ved udvælgelse af item/opgaver i testforløbet, mens sværhedstallet anvendes ved scoring af besvarelsen.”, Undervisningsministeriet (2006, side 65).¹⁴

4.2 Kravene til antal it-spørgsmål i et adaptivt testsystem baseret på multipel choice konceptet

I afsnit 4.1 blev der redegjort for, hvordan sammenhængen er fra trin- og slutmål og profilområder til it-spørgsmål med tilhørende sværhedstal og -niveau. I dette afsnit redegøres for, hvor mange spørgsmål, der (mindst) skal indgå i den databank til it-systemet, hvorfra spørgsmålene udtrækkes i en it-test, når denne endvidere skal danne grundlag for en valid og reliabel vurdering af elevens standpunkt i den delmængde af trin- og slutmålene, der er omfattet af it-spørgsmålene efter multipel choice modellen. I relation hertil er det afgørende, at test-designet skal udformes adaptivt, at en elev i hver it-test kan afprøve denne frivilligt to gange foruden den obligatoriske deltagelse, og at spørgsmålene udtrækkes tilfældigt fra databanken.

Et adaptivt testsystem er kendetegnet ved, at spørgsmålenes sværhedsgrad tilpasses på grundlag af forudgående besvarelser. Det rejser umiddelbart to problemer: 1) hvordan skal beslutningsreglen være, når efterfølgende spørgsmåls sværhedsgrad mekanisk skal besluttes i et it-system og 2) hvor stort skal springet være i sværhedsgrad?

I kravspecifikationen er det fastsat, at en *højere* sværhedsgrad skal vælges, når der er afgivet to rigtige svar i træk. Multipel choice konceptet gør, at det kan være tilfældigt, at der er afgivet to rigtige svar i træk. Hvor stor, sandsynligheden herfor er, afhænger naturligvis af, hvor mange svarmuligheder, der er i spørgsmålene. I henhold til kravspecifikationen skal en *lavere* sværhedsgrad vælges, når der er svaret blankt eller afgivet to urigtige svar i træk. Det kan ikke antages, at ikke-besvarelse eller urigtige svar er tilfældige. Heraf kan dog ikke sluttes, at elever, der er på samme kundskabs- og færdighedsniveau, vil få helt samme standpunkt i it-systemet, selv om der startes med ikke-besvarelse eller to urigtige svar, idet der efterfølgende sandsynligvis vil blive afgivet to korrekte svar i træk givet det adaptive design.

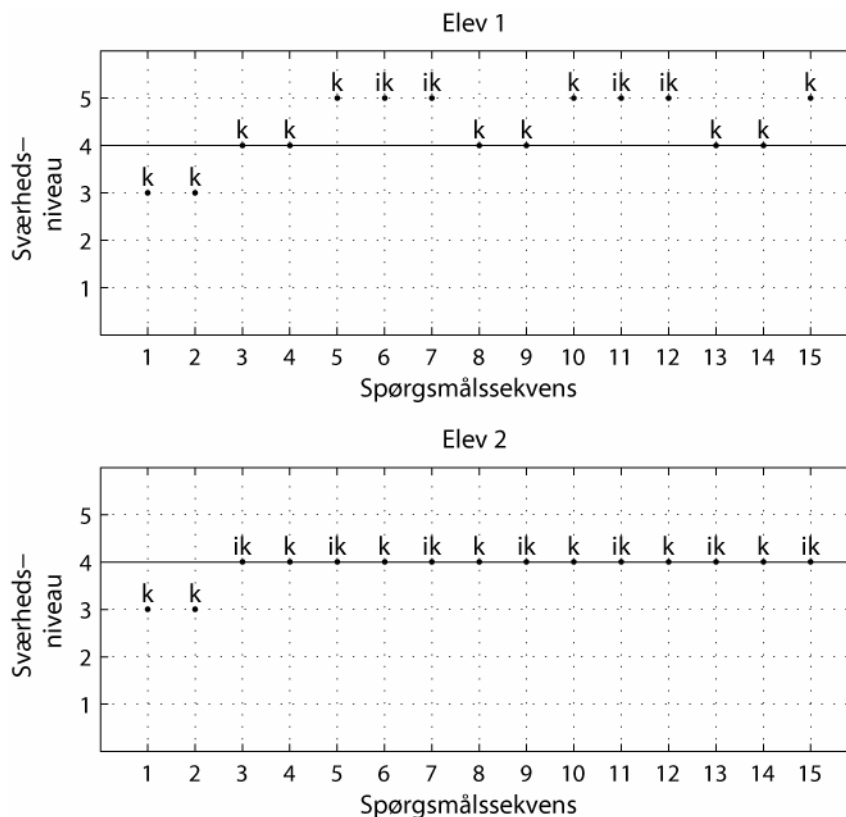
I ovenstående beskrivelse af justering i op- eller nedgående retning af spørgsmålenes sværhed er der som i kravspecifikationen brugt betegnelsen *sværhedsgrad* og ikke de ellers definerede sværhedsmål: *sværhedstallet* og *sværhedsniveau*. Dette skyldes, at det endnu er uklart, hvad der præcist vil blive tilfældet.¹⁵ Inden for hvert sværhedsniveau varierer sværhedstallet. Derfor kan sværhedsgraden naturligvis ændres, uden at sværhedsniveauet ændres. Og på dette område har Undervisningsministeriet ikke været præcise i kravspecifikationen, jf. Undervisningsministeriet (2006).

Da den enkelte elevs standpunkt ikke kendes af it-systemet, når en test påbegyndes, er det nærliggende at forestille sig, at der startes med spørgsmål, der svarer til gennemsnitsstandpunktet, altså med spørgsmål med et sværhedstal omkring 50 i sværhedsniveau 3. I det følgende lægges til grund, at der skiftes sværhedsniveau, når sværhedsgraden efter den specificerede beslutningsregel skal ændres. I figur 4 er der skitseret et tænkt testforløb med 15 spørgsmål (f.eks. inden for et profilområde) for to elever, der antages at have et kundskabs- og færdighedsniveau svarende til sværhedsniveau 4.

¹⁴ Item anvendes også som betegnelse for et spørgsmål/delspørgsmål og en opgave/delopgave.

¹⁵ Sværhedsgrad kan således opfattes om et overordnet begreb, der både kan fortolkes som sværhedstal og sværhedsniveau.

Figur 4. Eksempel på et it-testforløb inden for et profilområde med 15 spørgsmål i et adaptivt it-testsystem, der bygger på multipel choice. k betegner et korrekt besvaret spørgsmål, ik betegner et ikke-korrekt/forkert besvaret spørgsmål



Anm.: Årsagen til, at elev 1 svarer korrekt på spørgsmål 5, 10 og 15, kan antages at skyldes rene tilfældigheder som følge af multipel choice grundlaget for it-spørgsmålene.

Det fremgår af figur 4, at elev 1 har elleve korrekt besvarede spørgsmål med 2, 6 og 3 fra henholdsvis sværhedsniveau 3, 4 og 5, hvorfor det gennemsnitlige sværhedsniveau baseret på alle 15 spørgsmål er 3. Inddrages kun de korrekt besvarede spørgsmål fås et gennemsnitsniveau på 4,1, der efter sædvanlige afrundingsregler vil give et gennemsnitsniveau på 4.¹⁶ Elev 2 har svaret korrekt på 8 spørgsmål med 2 og 6 fra hhv. sværhedsniveau 3 og 4, hvorfor det gennemsnitlige sværhedsniveau baseret på de 15 spørgsmål er 2. Inddrages kun de korrekt besvarede spørgsmål fås et gennemsnit på 3,8, der også giver et sværhedsniveau på 4 efter afrunding. Det korrekte sværhedsniveau kan også findes ved at basere sig på det sværhedsniveau, der får flest korrekte besvarelser. Både elev 1 og 2 har i eksemplet i figur 4 flest korrekt besvarede spørgsmål på sværhedsniveau 4, nemlig 6. Der er derfor mindst to enkle metoder til at fastlægge sværhedsniveauet, og derfor også det hertil svarende sværhedstal til karakterisering af elevens kundskaber og færdigheder: gennemsnittet af de korrekte svar og det typisk korrekte sværhedsniveau.

¹⁶ Bemærk, at der anvendes en afvigende opdeling i tilbagemeldingssystemet, hvad der øger vanskelighederne med at afgive en korrekt standpunkt-vurdering i tilbagemeldingssystemet, der må baseres på sværhedstallet, jf. nærmere herom i afsnit 4.4.

Det væsentlige i denne sammenhæng er imidlertid, at typetallet for sværhedsniveau/sværhedstal først med rimelig sikkerhed er fastsat efter mindst 8 spørgsmål for såvel elev 1 som 2 i eksemplet i figur 4. Anvendes gennemsnitsmetoden er sværhedsniveauet tidligst fastsat efter 5 spørgsmål for elev 1 og efter 8 spørgsmål for elev 2.

Dette eksempel viser endvidere, se figur 4 for elev 2, at der mindst skal stilles 6 spørgsmål på hvert sværhedsniveau på hvert profilområde, dvs. 30 spørgsmål pr. profilområde, hvis et spørgsmål ikke skal gentages selv under forudsætning af, at et profilområde kan evalueres med kun en streng af spørgsmål. Og det forudsætter endvidere, at der ved skift i sværhedsgrad springes fra et sværhedsniveau til et andet. Og et spring i et sværhedsniveau til et andet tilgrænsende sværhedsniveau indebærer, at sværhedstallet for spørgsmålet før og efter springet i *gennemsnit* ændres med 20 (idet forskellen kan variere fra 1 til 40), idet sværhedstallet i sværhedsniveau 1-5 ligger i intervallerne 0-20, 21-40, 41-60, 61-80 og 81-100. Reduceres springet i sværhedsgrad fra i gennemsnit 20 til i gennemsnit 10 på sværhedsskalaen for at reducere usikkerheden i vurderingen af elevens kundskabs- og færdighedsstandpunkt, må der altså stilles det dobbelte antal spørgsmål.

I kravspecifikationen er det anført, at den enkelte test skal kunne benyttes op til 3 gange af den enkelte elev, uden at eleven bliver ”præsenteret for de samme opgaver i overvejende grad – uanset hvordan eleven besvarer opgaverne.”, Undervisningsministeriet (2006, side 61). Heraf følger, at der mindst skal stilles 18 spørgsmål pr. sværhedsgrad inden for et profilområde. Antages det, at der skal stilles det dobbelte antal opgaver for at undgå gentagelser, jf. kravspecifikationen, kan antallet af opgaver pr. test herefter opgøres til:

Antal opgaver på hvert sværhedsniveau på hvert profilområde	36
Antal opgaver på 5 sværhedsniveauer på hvert profilområde	180
Antal opgaver på 5 sværhedsniveauer i test af et fag med tre profilområder	540

I dette regnestykke er det antaget, at et profilområde kan dækkes ved kun en type af sammenhængende spørgsmål. Hvorvidt dette muliggør en valid og reliabel standpunktevaluering, må indtil videre være et åbent spørgsmål.

I kravspecifikationen er der også skønnet over det nødvendige antal opgaver i databanken pr. test. Den anvendte metode er dog en anden. I kravspecifikationen tages udgangspunkt i, at der er afsat 45 min. til effektiv test. Dette skal ses i sammenhæng med, at der til hver opgave også skal tilordnes en anslået besvarelsestid. De vanskeligste opgaver skal således kunne besvares på maks. 1 min. af den middelhurtige elev. Den middelhurtige elev antages i fortsættelse heraf at skulle besvare 50 spørgsmål pr. 45 min., jf. også afsnit 2.¹⁷

Det antages videre, at de hurtigste elever kan besvare 50 pct. flere opgaver end de middelhurtige elever, og at der skal være det dobbelte antal af besvarelser i databanken for at undgå for mange gentagelser af de samme spørgsmål i de tre mulige forsøg.¹⁸ I databanken skal der for hver test følgelig være 750 spørgsmål: 50 (antal spørgsmål til de middelhurtige) * 1.5 (grundet hensyntagen

¹⁷ I afsnit 2 er det anført, at der mindst skal svares på 50-60 spørgsmål. I opgørelsen af det nødvendige antal opgaver er det her forudsat, at der skal svares på 50. Det er naturligvis ligetil at gennemføre regnestykket med i stedet 60 spørgsmål.

¹⁸ Elever, der medvirker i aftestningen af it-spørgsmålene, vil i denne fase også kunne få stillet opgaver, som de senere kan blive stillet over for i bl.a. den obligatoriske test.

til de hurtigste elever) * 5 (antal sværhedsniveauer) * 2 (for at undgå genbrug af spørgsmål i overvejende grad).

Uanset om den ene eller den anden opgørelsesmetode lægges til grund bliver resultatet, at der skal stilles virkelig mange opgaver grundet valget af den adaptive testform kombineret med tilfældig udtrækning. Dette skyldes ikke blot kravene til antallet af opgaver pr. test samt det forhold, at der skal laves spørgsmål til 12 test. Der regnes nemlig med, at halvdelen af de foreslåede spørgsmål kasseres i testfasen, jf. følgende formulering:

”Erfaringer viser, at ca. 50% af de afprøvede items må kasseres. Det medfører, at der bør udarbejdes og afprøves mindst dobbelt så mange opgaver, som opgavebanken skal indeholde på hvert sværhedsniveau i hver test.”, Undervisningsministeriet (2006, side 67).

Hertil kommer, at selv om det er besluttet, at it-spørgsmålene er fortrolige, fordi de dels skal kunne genbruges i flere år og dels fordi it-systemet skal kunne anvendes hele året, er der flere grunde til, at det ikke er en engangsupgave at udforme opgaver: 1) den faglige og pædagogiske udvikling slutter ikke, 2) selv om der er tavshedspligt, er det usandsynligt, at nogle opgaver ikke bliver kendt af større kredse af elever og evt. også af offentligheden. Multipel choice designet gør det f.eks. i mange tilfælde lettere at redegøre for et spørgsmål og det rette svar. Denne problemstilling er bl.a. beskrevet således:

”Fører problemer med en opgave eller et item til, at der skal ske en kassering, forventes det, at opgavebanken suppleres med mindst et tilsvarende antal opgaver/items, der dækker de berørte trinmål. Det samme gælder for forældede, eksponerede eller for hyppigt anvendte opgaver. Der skal som minimum udskiftes 5% af opgaverne pr. år.”, Undervisningsministeriet (2006, side 63), 3) antallet af it-spørgsmål skal desuden hvert år vokse med 5%.

Efter igangsættelsen af systemet skal der altså hvert år mindst stilles nye opgaver svarende til 10% af det antal opgaver, der indgik i databanken for forrige år. Og da antallet af opgaver hvert år skal stige med mindst 5%, skal der endvidere absolut set stilles flere og flere nye opgaver fra år til år.

4.3 Fastsættelse af sværhedstallet for it-spørgsmål og afprøvning af nye spørgsmål

Det er vigtigt at være opmærksom på, hvordan sværhedstallet for et it-spørgsmål vil blive bestemt. Sværhedstallet er nemlig grundlaget for såvel fastsættelsen af standpunktet (scoringen) for den enkelte elev som for opdelingen af spørgsmålene i fem sværhedsniveauer, der igen med stor sandsynlighed vil blive et centralt element i den adaptive styring af sværhedsgraden i testforløbet, jf. afsnit 4.2. Hertil kommer, at tilbagemeldingen til den enkelte elev – se afsnit 4.4 – er baseret på sværhedstallet, hvilket derfor også vil gælde for dele af de aggregerede tilbagemeldinger til lærere, skoleledere, kommuner og Undervisningsministeriet.

Fastsættelsen af sværhedstal for it-spørgsmål og afprøvningen af disse skal ske ved, at et repræsentativt udsnit af elever (ca. 500) på det relevante klassetrin afprøver disse spørgsmål. I henhold til lov nr. 313 af 19. april 2006 har undervisningsministeren, jf. § 13, stk. 4, bemyndigelse til at udpege repræsentative grupper af elever til afprøvning af testenes relevans og pålidelighed mv.

Fastsættelse af sværhedstal for og afprøvning af it-spørgsmål skal ske i det udviklede it-system med to afgørende modifikationer:

1. Det adaptive princip og den mekaniske udtrækning af spørgsmålene er sat ud af kraft, idet alle testeleverne (ca. 500) skal besvare alle spørgsmål, og
2. Det er ikke muligt at svare blankt, men der skal være *mulighed* for (ved afkrydsning), at eleven tilkendegiver, at han/hun ikke har nogen ide om, hvad der er det rigtige svar på et spørgsmål, og at en besvarelse derfor ville være udtryk for et rent gæt.

Disse to modifikationer er en naturlig følge af, at andelen af korrekte svar er udgangspunktet for fastsættelsen af sværhedstallet, jf. den nærmere beskrivelse heraf i afsnit 4.2.¹⁹

4.4 Tilbage melding til elever, lærere, skoleledere, kommuner og Undervisningsministeriet

I afsnit 4.1 blev der redegjort for, at sværhedstallet var udgangspunkt for inddelingen i fem sværhedsniveauer med lige lange intervaller som målt ved sværhedsstal: 0-20, 21-40, 41-60, 61-80 og 81-100.

I de standpunktvurderinger, der skal udskrives til eleverne, skal der også anvendes fem niveauer, idet hhv. 10, 25, 30, 25 og 10 pct. af eleverne skal indgå i disse niveauer efter stigende præstationer. Derfor vil det blive sværhedstallet og ikke sværhedsniveauet, der bliver grundlaget for placeringen i de fem tilbagemeldingsniveauer.²⁰

Læreren, som har undervist i det fag, som eleverne er it-testet i, får fire rapporter:

Profilbeskrivelse for hver elev:

Profilbeskrivelsen er den vurdering, som it-testsystemet producerer for det niveau af kundskaber og færdigheder, som eleven har opnået som målt i it-testen. Profilbeskrivelsen, der udformes verbalt, er baseret på de fem tilbagemeldingsniveauer. Profilbeskrivelsen modtages af læreren og danner grundlaget for den diskussion, som læreren hurtigst muligt skal tage med den enkelte elev om dennes udbytte af undervisningen med særlig henblik på det fremadrettede undervisningsforløb, som også forældrene skal inddrages i som led i forældresamarbejdet. Eleven får profilbeskrivelsen udleveret af læreren efter denne diskussion.

Testrapport for hver elev:

Denne giver en detaljeret oversigt over elevens fulde testforløb med udgangspunkt i en gengivelse af spørgsmålene i den rækkefølge, eleven har fået disse under testen.²¹ For hvert spørgsmål er angivet: sværhedsniveau, normeret og anvendt tid til besvarelse, forbindelsen til trin-/slutmål og profilområde samt om eleven har svaret korrekt, ikke-korrekt eller undladt at besvare.

¹⁹ Tilkendegiver en elev, at han/hun ikke har nogen ide om, hvad der er det rigtige svar på et spørgsmål, og at en besvarelse derfor ville være udtryk for et rent gæt, tælles en sådan besvarelse naturligvis ikke med i andelen af korrekt besvarede spørgsmål.

²⁰ Dette øger naturligvis den usikkerhed, der er knyttet til vurderingen af en elevs standpunkt, hvis springet i sværhedsgrad alligevel styres af sværhedsniveauet. I det hele taget er det svært at se begrundelsen for at anvende sværhedsniveau-begrebet. Det komplicerer blot systemet – og i hvert fald forståelsen af det.

²¹ Dvs. lærerne kommer til at kende de test, som deres elever har skullet besvare.

Testrapport for klassen:

Bearbejdet oversigt over klassens præstationer, samt om omfanget af deltagelse o.lign.

Profiloversigt for klassen:

Denne giver for hver elev en oversigt over de enkelte elevers profiler.

Tilbagemeldingen til elever og lærere er derfor et relativt evalueringssystem bestemt af, at hhv. 10, 25, 30, 25 og 10 pct. af eleverne skal placeres i de fem tilbagemeldingsniveauer. Et sådant relativt system gør det ikke muligt at vurdere den absolutte udvikling i elevernes præstationer fra år til år, hvilket er i overensstemmelse med, at testene i relation til lærer og elev ”skal være et værktøj til at sikre faglige fremskridt for den enkelte elev ved at målrette undervisningen til den enkelte elevs særlige evner.”, Undervisningsministeriet (2005, side 13).

I de rapporter, der udskrives til skoleledere, kommuner og Undervisningsministeriet vil der imidlertid også indgå opgørelser, hvoraf den absolutte udvikling fremgår. I disse rapporter vil elevernes præstationer endvidere blive korrigeret for elevernes sociale baggrund (forældres uddannelsesbaggrund og indkomstforhold, etnicitet, køn m.v.). Rapporteringen vil altså blive opbygget, så de enkelte skoler og kommuner *kan* rangordnes i forhold til landsresultatet, der hvert år vil fremgå af den såkaldte nationale præstationsprofil fra Undervisningsministeriet. Og skolelederen og skolebestyrelsen for den enkelte skole *vil få oplyst*, hvordan skolens præstation er i forhold til landsgennemsnittet på basis af de tal, der er korrigeret for elevernes sociale baggrund. Kommunerne, der har tilsynet med skolerne, *vil ligeledes få oplyst*, hvordan de enkelte skolers placering er i forhold til landsgennemsnittet.

I denne sammenhæng er det dog centralt at inddrage den grundlæggende aftale af 23. sep. 2005 om de nationale test. Heri er det aftalt:

”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.”

Det følger heraf, at både skolelederen, skolebestyrelsen og en række personer og udvalg m.m. i kommunerne vil få kendskab til, hvordan en skole/skolerne i kommuner er placeret i forhold til landsgennemsnittet på et grundlag, der er forsøgt gjort sammenligneligt ved at korrigere for elevernes sociale baggrund.

Det bliver interessant at følge, om det i praksis er muligt at opretholde fortroligheden i et principielt åbent samfund som det danske, hvor bl.a. flere personer, der har børn eller børnebørn i kommunens skoler, får kendskab til disse skolers placering i forhold til landsgennemsnittet/den nationale standard.

Denne problemstilling har da også helt åbenbart voldt problemer, som det kan læses af notatet af 7. april om ”Testresultaters anvendelse og fortrolighed”. Her står bl.a. i relation til kommunalbestyrelsens drøftelse i åbne møder af den årlige kommunale kvalitetsrapport om kommunes folkeskoler:

”På de åbne møder må man ikke offentliggøre de enkelte skolers testresultater eller oplyse, hvordan skolerne ligger resultatmæssigt i forhold til hinanden. Når konklusioner om det faglige niveau er truffet ved afvejning af de forskellige indikatorer, må offentliggørelse heraf ske i en form, så fortroligheden om testresultaterne opretholdes. Offentligheden vil f.eks. kunne gøres bekendt med, at visse navngivne skoler ud fra en helhedsvurdering af de forskellige indikatorer har dårlig kvalitet, men uden at de ud fra kvalitetsrapporten eller udtalelser på åbne møder får kendskab til oplysninger om de pågældende skolers testresultater.

De samme forhold gør sig gældende, når det samlede skolevæsen skal beskrives. Man vil således f.eks. kunne udtale, at det faglige niveau for kommunens skoler er tilfredsstillende set ud fra en samlet afvejning af de forskellige indikatorer. Kommunalbestyrelsen må imidlertid ikke udtale sig om, hvordan testresultaterne er for de enkelte skoler eller alle skoler i kommunen eller sammenlignet med landsresultaterne.

Det afgørende for, hvilke udtalelser der kan gives offentligt, er som nævnt, at det sker således, at testresultaterne ikke offentliggøres, herunder at der gives oplysninger om, hvordan skolerne resultatmæssigt ligger i forhold til hinanden. Testresultaterne må således ikke indgå i udtalelser, der kan skabe grundlag for rangordning.”, se www.uvm.dk.

Det er vist ikke nødvendigt yderligere at kommentere dette uddrag.

4.5 Udviklings- og driftsomkostninger

I relation til den ændring af folkeskoleloven, der inkorporerede forliget af 23. sep. 2005 mellem regeringen (Venstre og Det Konservative Folkeparti), Socialdemokraterne og Dansk Folkeparti blev der skønnet over udviklings- og driftsomkostningerne i introduktionsperioden 2005-09. Det blev anslået, at den samlede udgiftssum til indsatsen for at fremme evalueringskulturen i folkeskolen ville blive 90 mio. kr. i årene 2005-09, heraf mindst 50 mio. kr. til it-relaterede udgifter.²²

Personligt vil jeg blive overrasket, hvis denne ramme ikke væsentligt overskrides. Undervisningsministeriet har tidligere helt undervurderet udviklingsomkostninger til it-systemer. STADS (studieadministrativt system) og ØSS (økonomistyringssystem) til de højere læreanstalter udløste således udgifter på ca. 220 mio. kr. alene til Undervisningsministeriet, hvortil kommer brugerbetaling fra institutionerne. Men værre: systemet var så problemfyldt og fejlbehæftet, at Undervisningsministeriet i realiteten måtte kaste håndklædet i ringen og overlade det afsluttende lappearbejde til de højere læreanstalter, der ville benytte systemet. Bl.a. Københavns Universitet og Handelshøjskolen i København ønskede overhovedet ikke at anvende systemet, hvorved et vigtigt formål med systemet – opbygningen af en fælles databank med bl.a. gennemførelsesprocenter og studietider – endvidere måtte opgives.

Det kan naturligvis ikke udelukkes, at ministeriet denne gang har skønnet nogenlunde korrekt, idet projektet jo også er udliciteret. Men fortiden og det forhold, at Undervisningsministeriet har måttet lade EU-udbudet om it-testsystemet gå om, styrker ikke tiltroen hertil, jf. også flg. begrundelse for aflysningen af det første udbud:

²² I Undervisningsministeriets pressemeddelelse af 26. juni 2006 er den samlede udgiftsramme allerede blevet forhøjet til 110,8 mio. kr. for årene 2005-09, heraf medgår 55 mio. kr. til COWI A/S for treårsperioden: juli 2006-juni 2009.

”På baggrund af den viden, Undervisningsministeriet har erhvervet i løbet af udbudsprocessen, og på baggrund af ministeriets seneste risikoanalyser, er det besluttet ikke at indgå kontrakt på nuværende tidspunkt. Udbudet er derfor den 28. oktober 2005 blevet aflyst med henblik på, at Undervisningsministeriet snarest muligt vil indlede et nyt EU-udbud med henblik på implementering af testene i 2007.”, Undervisningsministeriet (2005).

Her anføres jo, at Undervisningsministeriet ikke havde forberedt den første udbudsrunde tilstrækkeligt. Spørgsmålet er, om dette nu er sket. Hvis ikke er overskridelse af budgettet sandsynligvis ikke den eneste konsekvens – systemet kan også blive mere eller mindre amputeret selv henset til den indsnævring, der på forhånd er sket ved, at det kun er en delmængde af trin- og slutmålene, der indgår i it-testsystemet.

5. Konklusion

Aftalen af 23. sep. 2005 mellem regeringen (Venstre og Det Konservative Folkeparti), Socialdemokraterne og Dansk Folkeparti indeholder ingen spor af, at aftalepartnerne var klar over, at it-testsystemet vil blive opbygget på basis af multipel choice konceptet, hvor gennemsnitseleven endvidere skal besvare mindst 50-60 spørgsmål på 45 min. At resultatet med andre ord for alle praktiske formål bliver et system til test af paratviden. Offentlige tilkendegivelser i tilknytning til aftalen af 23. sep. 2005 tyder mest på, at aftalepartnerne mest var ”imponeret” over moderniteten – det it-mæssige – i aftalen!

Når noget vælges, vælges andet fra. Derfor burde det også være overvejet, om der ikke var andre ændringer, der hellere skulle være gennemført i relation til folkeskolen i lyset af PISA-evalueringerne. Dette er så meget desto mere påkrævet i lyset af de ganske betydelige omkostninger, som det obligatoriske nationale testsystem vil føre med sig, jf. afsnit 4.5. I den sammenhæng må det vække forundring, at der ikke er gennemført en dybtgående analyse af årsagerne til, at finske elever indtager førstepladsen i PISA-undersøgelserne, når det finske skolesystem har afgørende fælles kendetegn med det danske, se Teknologirådet (2005). Bl.a. ingen karaktergivning i de mindre klasser og ingen nationale test!

Hvordan kunne det være, at den daværende Undervisningsminister kunne reagere på PISA-2003 ved primært at kræve nationale test i folkeskolen, når PISA-2003 viste, at nationale test ikke havde virkning, jf. uddraget i boks 1 af pressemeddelelse af 6. december 2004?

Og hvordan kunne det være, at det blev centralt for den nuværende undervisningsminister at få flere timer i dansk og historie, når Teknologisk Institut i en undersøgelse for EU-kommissionen, bl.a. på basis af PISA-2003, kunne vise, at flere timer ikke i sig selv hjalp på elevernes præstationer, jf. følgende uddrag:

”The weak statistical relation between student time devoting to learning and performance reminds policy makers that there is no clear-cut and simple relation between instruction time and academic results. More instruction hours do not necessarily lead to better academic performance among students. The contents of teaching and learning may be more important than the amount of time allocated to teaching and learning.”, Danish Technological Institute (Nov. 2005, side 8-9).

Boks 1. Testforslagene i undervisningsministerens pressemeddelelse af 6. december 2004

I pressemeddelelsen foreslår undervisningsministeren primært at øge testningen af folkeskoleeleverne ved:

- 1) en obligatorisk *sprogscreening* af alle børn ved skolestart,
- 2) inddragelse af gældende trin- og slutmål for folkeskolens fag i *vurderingen* af den enkelte elevs udbytte af undervisningen,
- 3) *obligatoriske test* (profiltest i slutningen af 2., 4. og 6. klasse, læsetest i begyndelsen af 8. klasse, profiltest af matematik i 6. klasse, af engelsk i 7. klasse og af naturfag i 8. klasse) samt
- 4) skriftlig tilbagemelding til forældrene af deres børns udbytte af undervisningen, herunder af *testresultaterne*.

I tilknytning hertil skal der stilles en *testværktøjskasse* til rådighed for Folkeskolen.

Endelig skal lærernes uddannelse og efteruddannelse forbedres.

På trods af dette blev der efterfølgende med lovforslag L170 af 1. marts 2006 alligevel fremsat forslag om at øge timetallet i dansk og historie med henblik på at øge elevernes præstationer. Men som ovenstående uddrag viser, og som der ikke blev taget stilling til i relation til lovforslaget, er der ingen sikkerhed for at flere timer øger præstationerne.²³

En mulighed er naturligvis, at de politisk ansvarlige blæser på faglige resultater, hvis disse strider imod deres forudfattede mening. En anden og efter min opfattelse mere nærliggende mulighed er, at det faglige niveau i topledelsen i Undervisningsministeriet er så ringe, at det politisk opportune får alt for meget spillerum.

Hvis det er korrekt, at det faglige niveau i topledelsen af Undervisningsministeriet er meget ringe, kan det også bedre forstås, hvorfor ønsket om at fremme en evalueringskultur i folkeskolen nu vil blive kompromitteret af, at det lovpriste it-testsystem blot vil blive et system til test af paratviden. Dette vil i øvrigt blive sat yderligere i relief af de efter min mening meget betydelige meromkostninger, der bliver resultatet af, at det straks blev besluttet at iværksætte et ambitiøst it-testsystem, selv om der ikke forelå seriøse undersøgelser af, hvad et sådant system skulle teste, hvordan det skulle gøres, og om det ikke havde været mere hensigtsmæssigt at starte langt simplere med alm. testskemaer, idet resultater herfra så kunne være benyttet til langt mere præcist at designe et it-system, hvis dette skulle vise sig at være den rette løsning.

Det kan med rette spørges, hvorfor jeg allerede nu kommer med denne negative vurdering af det kommende it-testsystem. Hvorfor ikke vente til systemet er færdigudviklet, afprøvet og testet ved de kommende obligatoriske testrunder? Svaret herpå er enkelt og fremgår allerede af den vestlige verdens to første grundbøger: Homers Iliaden og Odysseen. Der er ikke meget anseelse forbundet med ikke at vove noget. Derfor.

²³ Lovforslaget blev vedtaget ved 3. behandling 2. juni 2006 med 86 stemmer for (Venstre, Det Konservative Folkeparti, Socialdemokraterne og Dansk Folkeparti samt en fra Socialistisk Folkeparti, der stemte forkert) og 23 imod (Det radikale Venstre, Socialistisk Folkeparti og Enhedslisten samt fire socialdemokrater).

Litteraturliste

Danish Technological Institute 2005. *Explaining Student Performance. Evidence from the international PISA, TIMSS and PIRLS surveys*. Copenhagen.

Finansministeriet 2006. *Aftale om fremtidens velstand og velfærd og investeringer i fremtiden. Aftale om fremtidig indvandring*. 20. juni 2006. København.

Fuchs, T. and L. Woessmann 2004. *What Accounts for International Differences in Student Performance? A Re-examination using PISA Data*. CESifo Working Paper No. 1235.

OECD 2004. *Learning for Tomorrow's World. First Results from PISA 2003*. Paris.

Sørensen 2005. PISA-2003. Hvad kan der konkluderes? *Working Paper No. 02/2005*. Institut for Regnskab og Finansiering, Syddansk Universitet.

Teknologirådet 2005. *PISA-undersøgelsen og det danske uddannelsessystem*. Folketingshøring om PISA-undersøgelsen. Mandag den 12. september 2005. København.

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

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

Velfærdskommissionen 2006. *Fremtidens velfærd – vores valg*. Analyserapport I. Januar 2006. København.