

De digitale afgangsprøver har høj kvalitet

Keld Nørgaard, formand for opgavekommissionen for naturfag

Lise Steinmüller & Michael Lund-Larsen, @ventures

Kommentar til anmeldelsen "Test og kompetenceudvikling" i MONA, 2006(2).

For første gang i mange år skulle eleverne i 9. klasse i skoleåret 05/06 til afgangsprøve i faget biologi. Det var politisk besluttet at prøven både skulle bestå af en praktisk-mundtlig prøve og en digital prøve. Denne kommentar beskæftiger sig kun med den digitale prøve i december og maj – og det sker med udgangspunkt i Eigil Larsens anmeldelse af decemberprøven i juniudgaven af MONA¹.

I december 2005 blev den digitale prøve gennemført med 500 deltagere. I marts 2006 testede 30.000 elever prøveformen, som baserede sig på december 2005-prøven. I maj 2006 gennemførte 33.500 deltagere prøven i elektronisk form, og ca. 25.000 elever på papir. Desuden gennemførte et lille antal elever sygeprøven i juni 2006.

I prøverne blev der anvendt testformer med forskellige varianter af multiple-choice og rigtigt/forkert. Det vil sige at eleverne primært har skullet vælge deres svar mellem en række på forhånd udarbejdede udsagn. Eleverne blev stillet over for ca. 20 opgaver med forskellige sværhedsgrader og et varieret antal svarmuligheder.

I 2005/06 er eleverne blevet bedømt med en standpunktskarakter, en karakter for den praktisk-mundtlige prøve og en for den digitale prøve. Bedømmelsen ved den digitale prøve afviger fra de to førstnævnte:

For det første er den digitale prøve objektiv. Det vil sige at der ikke tages andre individuelle hensyn til eleverne end dem der er indtænkt i opgaveformuleringen, ligesom der ikke er mulighed for fortolkning af elevernes svar. Dette har både positive og negative effekter, som det dog ligger uden for denne kommentar at behandle.

For det andet er prøven skriftlig forstået på den måde at eleven stilles over for skriftlige udsagn som skal behandles med afkrydsning eller indtastning. Dette kan være en hæmsko for nogle elever idet de ikke kan give uddybende svar, men det kan samtidig være en styrke for elever som ikke trives ved mundtlige prøveformer, eller

1 Prøven kan findes på <http://us.uvm.dk/grundskole/proeverogevaluering/naturfag.htm/?menuid=1020>.

som har svært ved at udtrykke sig skriftligt. Man kan altså med en vis ret sige at der med indførelse af digitale prøver åbnes mulighed for at nogle elevgrupper vil få bedre mulighed for at vise hvad de kan, idet de hverken stilles over for et krav om skriftlig formuleringsevne eller mundtlig fremlæggelse.

Samtidig er det vigtigt at være opmærksom på at den digitale prøve skal suppleres af den praktisk-mundtlige prøve (eller omvendt).

Udviklingsprocessen for prøverne

Formålet med den digitale prøveform er at prøve eleven i forståelse og anvendelse af faglige begreber inden for fagets slutmål.

Udgangspunktet for formuleringen af opgaverne har derfor været fagets slutmål. Men da slutmålene er meget bredt formuleret, har det været nødvendigt også at inddrage de relevante trinmål i forsøget på at præcisere det faglige indhold. Trinmålene er imidlertid også meget upræcise. Det har derfor været nødvendigt ud fra slut- og trinmål at forsøge at udlede relevante og grundlæggende biologiske begreber, forståelser og sammenhænge – uden at skele til lærebøger.

Det er nødvendigt både for mange lærere, for den faglige progression i undervisningen, for det tværnaturfaglige samarbejde samt selvfølgelig for evaluering og afgangsprøvesituationen at målene bliver præciserede.

Det er desuden ikke realistisk at forvente at én digital afgangsprøve skal komme "hele vejen rundt" og teste eleverne i alle slutmål – hvilket vel heller ikke forventes ved den praktisk-mundtlige prøve. I decemberprøven og især i majprøven lykkedes det alligevel at udarbejde opgaver som stort set berører alle slutmål i de tre første CKF-områder. At det i decemberprøven ikke umiddelbart opleves at der er opgaver inden for 'Biologiens anvendelse', skyldes dette områdes begrænsede eksempler. Opgave 13 omhandler netop *naturpleje* som eksempel på anvendelse af biologisk viden.

Sammenhæng, kontekst og paratforståelse

Prøvens indledende tekst skal hjælpe eleven til at se spørgsmål og svar ind i en sammenhæng. Det er centralt i forhold til Signalementet i Læreplanen at eleverne får mulighed for at bruge deres viden og forståelse til at besvare opgaven, dvs. at eleverne ud fra deres biologiske viden og indsigt i et rimeligt omfang får mulighed for at ræsonnere sig frem til en svarmulighed. Derfor er alle opgaver sat ind i et emne, en kontekst, hvor eleverne skal anvende deres tilegnede biologiske viden og indsigt – deres paratforståelse – til i sammenhæng med opgavetekstens informationer at tage stilling til svarmulighederne. Opgaver som ensidigt tester elevens paratviden, søges undgået.

Det vil stille eleven i en urimelig situation hvis begreberne i opgaverne skulle testes i et tomrum, klinisk rensat for sammenhæng – og dermed fri af enhver form for kontekst.

Opgaverne er udformet så det principielt er uden betydning om eleven har arbejdet med landbruget, søen, skoven, havet eller noget helt femte. Det centrale er om eleven har forstået betydningen af det udvalgte grundlæggende biologiske begreb, f.eks. *sympiose*, og kan anvende denne viden, uanset kontekst.

Et andet eksempel kan være *ukønnet formering*. Har eleven lært om *ukønnet formering* og forstået indholdet af dette begreb, bør det være uden betydning om en opgave vedrørende *ukønnet formering* er eksemplificeret ved engplanter eller landbrugsafgrøder.

I decemberprøven var der et eksempel på en opgave med fejl. Det var opgaven om en fødekæde med gedde, aborre og skalle. Der kan naturligvis ikke stilles krav om at eleverne har kendskab til disse specifikke dyr. Denne opgave var ikke hensigtsmæssig og omfatter en fejltipe som vil blive forsøgt undgået fremover.

Som nævnt skal opgaverne være lærebogsuafhængige. Der må ikke opstå den situation at elever der har arbejdet med et bestemt lærebogssystem, kan drage fordel heraf i prøvesituationen. Illustrationen i decemberprøvens opgave 10 er ganske principiel og adskiller sig fra den omtalte bogillustration ved at have færre kurver fordelt ikke på 3, men 4 opdelinger. Det er en væsentlig faglig kompetence at eleven kan aflæse sådanne illustrationer.

Udviklingsprocedure

Når testspørgsmålene udvikles, er de gennem en række faser med en række udviklere som bidrager med hver deres faglige ekspertise. Først og fremmest er der opgavekommissionen som udarbejder opgavernes indhold. Dernæst er der faglig, pædagogisk, sproglig og almen forståelseskorrektur. Hver enkelt opgave med tilhørende svarmuligheder er igennem mindst 15 gennembearbejdsfaser og ca. lige så mange personer inden den endeligt godkendes.

Illustrationer indsættes hvor det skønnes at det kan være en hjælp for elever med manglende førfaglig forståelse eller med manglende sprogkompetence. Illustrationerne er unikke for den enkelte opgave, og deres relevans og indhold er gennem samme kvalitetskontrol som opgaverne.

Kvalitetskontrol af test

Anvendelse af test til at bedømme elevens præstation inden for et fag eller et fagområde bygger på den forudsætning at det er muligt at konstruere testopgaver der reliabelt og validt kan danne grundlag for en sådan bedømmelse.

En test er reliabel hvis det kan sandsynliggøres at testopgaverne besvares af eleverne på (stort set) samme måde ved gentagende besvarelser. En test er valid hvis den faktisk måler det den forventes at måle. Dette udtrykkes bl.a. gennem målingens sikkerhed.

En Rasch-analyse er en anerkendt empirisk baseret statistisk metode til at kvalitetskontrollere test. Undervisningsministeriet har fået foretaget en sådan Rasch-analyse af testene.

Prøvens kvalitet

En tommelfingerregel siger² at reliabiliteten på en test bør ligge i intervallet 0,7-0,9. Reliabiliteten for decemberprøven er på ca. 0,85 og for majprøven på ca. 0,77.

Usikkerheden på testpersonernes præstationsmål udtrykkes ved standardfejlen på målingen³. Denne skal helst ligge så tæt på det teoretiske minimum som muligt. I Den Digitale Naturfagsprøve ligger det teoretiske minimum for SEM omkring 0,3 logit-enheder, og SEM for både decemberprøven og majprøven ligger på ca. 0,3.

Eleverne gætter ikke

Da Rasch-analysen omfatter en analyse af elevernes svarmønstre, er det muligt at vurdere gætnings/sløseris betydning for elevens præstationsbedømmelse. Analysen afviser at der i nævneværdig grad har været tale om gætning eller sløseri i forbindelse med elevernes besvarelse af prøvesættene. Således vurderes andelen af elever der kan have gættet eller sløset, generelt til at ligge mellem 0,01 % og 0,09 %⁴.

Sandsynligheden for gæt ved fastformulerede svarmuligheder kan ikke fjernes helt, og det kan den næppe ved nogen prøveformer. Man kan heller ikke ved en mundtlig prøve gardere sig mod at en elev kommer op i sit yndlingsemne, eller at en opgave i et skriftligt prøvesæt svarer til det læreren lige har gennemgået i spørgetimen. Sådan er eksamen. Men med de skriftlige naturfagsprøver kan man sige at man har mindsket konsekvensen af gæt, idet den skriftlige karakter bør ses sammen med den praktisk-mundtlige prøve. Disse to prøver vil med al sandsynlighed sammen give et ret godt billede af hvad eleven rent faktisk kan.

Testens sværhed

De testopgaver der indgår i Den Digitale Naturfagsprøve, kan alle omsættes til et antal dikotome items – dvs. spørgsmål med 2 svarmuligheder: rigtigt eller forkert. Dette gør det muligt at anvende en Rasch-analysemodel.

Hver testopgave består af et eller flere spørgsmål der hver for sig i Rasch-analysen bliver opfattet som selvstændige items. I visse tilfælde indeholder et spørgsmål 2 eller flere rigtige svar (f.eks. en multiple-choice med 6 svarmuligheder, hvoraf 2 eller flere

2 McColl, Christiansen, König-Zahn (1997): Making the right choice of outcome measure.

3 SEM: Standard Error of the Measure.

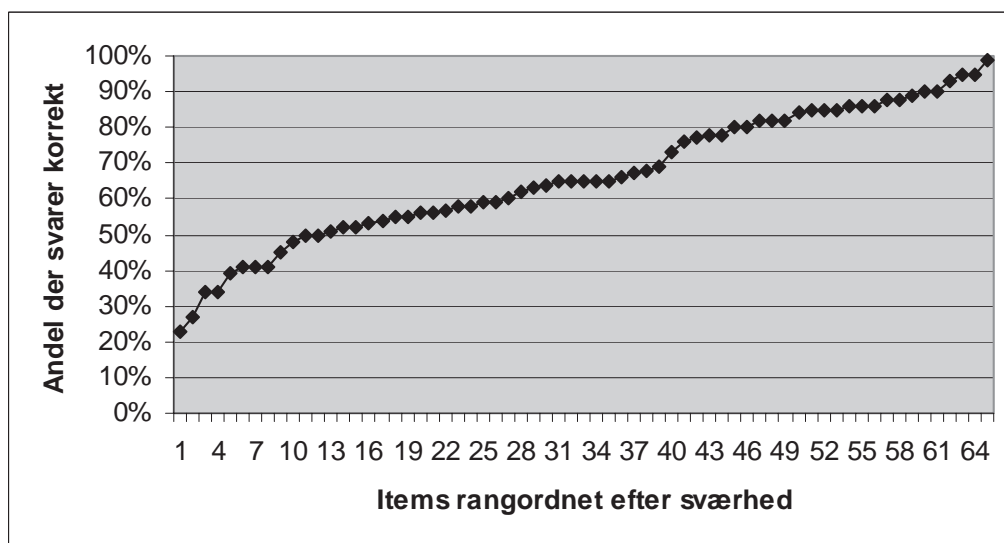
4 Kriterier defineret af Linacre, J.M. & Wright, B.D. (1994). Chi-Square Fit Statistics. Rasch Measurement Transactions 8:2 s. 350. (<http://rasch.org/rmt/rmt82a.htm>), dvs. OutfitMS > 3,0 og Infit < 0,9 eller Infit > 1,1. Denne undersøgelse er ikke foretaget på decemberprøven (under 500 besvarelser) men alene på majprøven med ca. 33.500 besvarelser.

skal afkrydses). I disse tilfælde bliver hver svarmulighed opfattet i Rasch-analysen som selvstændige items.

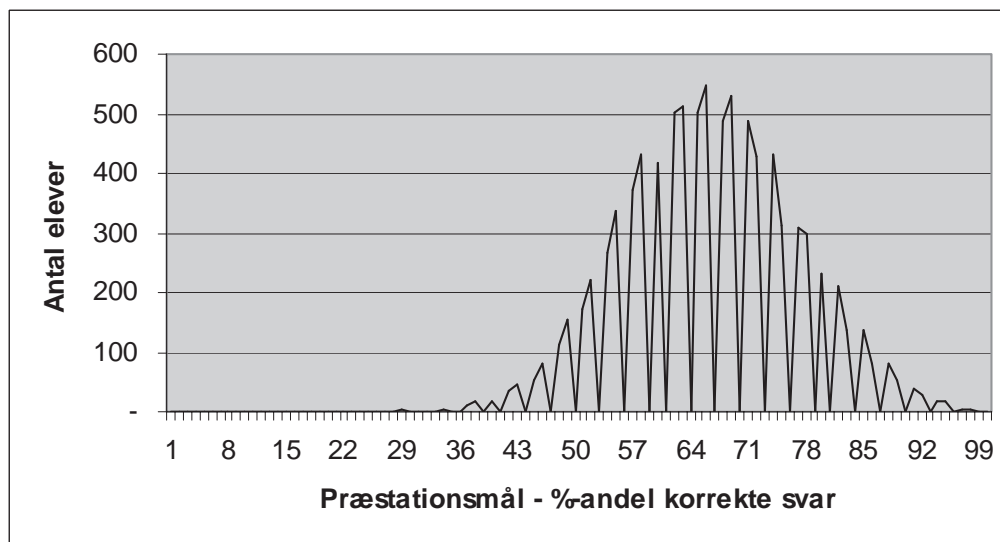
Derved består et prøvesæt med 20 testopgaver af et større antal items, der alle er dikotome. I majprøvesættene var der 65 items i hvert prøvesæt. I decemberprøven var der 66 items.

Der dannes ved analysen en fælles skala for henholdsvis items sværhed og testpersonernes præstationsmål (dygtighedsniveau) i forhold til det område som testen skal bedømme. Items rangordnes i forhold til deres sværhed (sandsynligheden for at en vilkårlig testperson svarer korrekt på det givne item) og giver det bedste målegrundlag hvis de fordeler sig jævnt på hele skalaen. Testpersonernes præstationsmål (sandsynligheden for at en testperson kan svare rigtigt på et vilkårligt item) fordeler sig tilnærmelsesvis efter en normalfordeling på skalaen.

Graferne i figur 1 og 2 viser Biologi prøvesæt 2 som eksempel på items og præstationsmåls fordeling (de øvrige prøvesæt har næsten samme fordelinger). Items fordeler sig jævnt over sværhedsskalaen uden items i intervallet 0-20 % (de sværeste items). Tilsvarende fordeler testpersonerne sig efter en normalfordeling på præstationsskalaen med toppunkt hvor eleverne har ca. 65 % korrekte svar.



Figur 1.



Figur 2.

Testopgavernes kvalitet

Rasch-analysen kan som nævnt bl.a. afdække hvilke items der ikke besvares på en måde så de kan bidrage til at fastlægge den bedømmelsesskala hvorpå respondenterne indplaceres. Eller med andre ord: Hvilke items besvares ikke som forventet i forhold til en såkaldt endimensionel skala.

I både majprøvens og decemberprøvens biologisæt, hvor der er 65-66 items i hvert af sættene, passer ca. 6 % af itemsene ikke til modellen, hvilket erfaringsmæssigt er en meget lille andel⁵. Årsagen til denne høje kvalitet skal ses i testkonstruktionsmetoden der er baseret på en række afgørende kvalitetsfaktorer vedrørende testmål og spørgsmålsudformning. Det vurderes at de 6 % items ikke har påvirket den konkrete bedømmelse af eleverne nævneværdigt, men det vil en nærmere analyse kunne afdække.

Negative point ændrer intet

Eigil Larsen skriver: "Forkerte besvarelser trækker ikke ned og giver f.eks. ikke et negativt antal point. Eleven belønnes altså for at gætte". Der er her tale om en fejlslutning. Negative point ændrer ikke på bedømmelsen af eleven. Om et fejlsvaret bedømmes med nul point eller et negativt pointtal, har alene betydning for pointskalaens længde –

⁵ Ved opbygning af testopgave-banker forkaster Rasch-analyser normalt op mod halvdelen af de konstruerede items.

ikke for elevernes rangordning og dermed heller ikke elevernes karakterer, idet karakterskalaens sammenhæng med pointskalaens intervaller ændres tilsvarende.

Afslutning

Egentlig er det vel ikke helt usandsynligt at den digitale prøveform kan medvirke til at fremme en god naturfagsundervisning ved netop at sætte fokus på forståelse og anvendelse af *grundlæggende biologiske begreber*. Det er jo fint at undervise i biologiske emner, men man skal som lærer også gøre sig klart hvad det er eleverne skal tilegne sig ved arbejdet med emnet/biotopen. Læreren skal sætte de faglige mål for undervisningen, dvs. fastlægge hvilke grundlæggende faglige begreber, forståelser og sammenhænge eleverne skal tilegne sig. Så kan man sammen med eleverne vælge emne, biotop osv. som kan tilgodese målene. I forbindelse med planlægningen af forløbene i klassen er det en central forudsætning at læreren målfastsætter forløbet og dermed kan evaluere sin undervisning og elevernes målopfyldelse.