**Vurdering vedr. tekstforståelsesopgaver med høj sværhedsgrad**

12. februar 2011

**Indledning**

Jeg er blevet bedt om at vurdere, om det inden for de rammer, der er beskrevet i et tilsendt notat om de nationale test på området tekstforståelse, kan lade sig gøre at udfærdige opgaver, som opfylder følgende to krav:

1. Maksimalt 10 % af eleverne på det klassetrin, der testes, må kunne svare rigtigt.

2. Opgaverne må ikke blive forkastet i en Raschanalyse af opgavehomogenitet.

Min vurdering bygger dels på det omtalte notat, "Tekstforståelse – rammerne og processen bag" (udarbejdet af Skolestyrelsen og COWI-konsortiet og dateret september 2010 – herefter "rammenotatet"), samt på rekvirerede (og naturligvis fortrolige) eksempler på opgaver med høj sværhedsgrad (enten formodet eller empirisk), opdelt efter om de har klaret kravet om opgavehomogenitet.

Indledningsvis kan det bemærkes, at formålet med at inkludere opgaver med høj sværhedsgrad i en læsetest er at give de dygtigste testdeltagere mulighed for at vise, hvor dygtige de faktisk er. Det bemærkes i rammenotatet s. 3, at de nationale test trods relativt få svære tekstforståelsesopgaver opfylder et krav om, at der indenfor 45 minutter skal kunne opnås statistisk sikre bedømmelser af mindst 90 % af eleverne. Om dette vil jeg bemærke, at 45 minutter er en forholdsvis lang prøvetid, når man tager i betragtning at der er tale om adaptive prøver. En attraktion ved adaptive prøver er netop muligheden for at minimere prøvetiden ved i udvælgelsen af opgaver at zoome ind på den enkelte elevs dygtighedsniveau. Muligheden for at stille meget svære opgaver kan have væsentlig betydning for, hvor hurtigt man opnår en sikker bedømmelse, og dermed også for hvordan brugerne oplever testen.

**Vurdering**

Det kan konstateres, at der i det forelagte ER (enkelte) eksempler på opgaver, der har opfyldt de nævnte krav. Det korte svar på spørgsmålet kunne derfor være, at det KAN lade sig gøre at indfri kravene. Visse forhold ved de nævnte kravopfyldende opgaver får mig imidlertid til at fraråde at de gøres til forbillede for udarbejdelsen af yderligere vanskelige tekstforståelsesopgaver.

Disse forhold har dels at gøre med den tekniske udformning af opgaverne (primært antallet af svarmuligheder), dels med den rent indholdsmæssige udformning (primært formuleringen af svarmuligheder). Vanskelighederne ift. tekniske og indholdsmæssige rammer uddybes nedenfor.

Samlet set er min vurdering, at det ER muligt at udforme svære opgaver inden for de beskrevne rammer – men det må medgives, at udformningen af svære tekstforståelsesopgaver er forbundet med mange vanskeligheder, og at en lav succesrate derfor er forventelig. Endvidere vil jeg gøre opmærksom på at face-validiteten (den oplevede validitet) kan udgøre et særligt problem ved meget svære opgaver.

**Vedr. tekniske vanskeligheder**

Det påpeges i rammenotatet s. 3, at polytome opgaver (opgaver, hvor der indgår flere delspørgsmål) kan gøres sværere end dikotome opgaver (med kun en enkelt spørgsmål), og at sikre bedømmelser opnås hurtigst med polytome opgaver. Årsagen til dette er indlysende nok den, at det er nemmere at svar korrekt på et enkelt spørgsmål end på en række spørgsmål (to eller flere), når det kræves, at alle besvares korrekt.

Ikke desto mindre er der meget få polytome opgaver blandt de eksempler på empirisk svære opgaver jeg har fået forelagt. Den typiske opgave har fem valgmuligheder – hvoriblandt en enkelt mulighed skal markeres som korrekt. Gættesandsynligheden i den slags opgaver er 20 %: Hvis samtlige deltagere markerer en helt tilfældig svarmulighed, er det forventeligt, at cirka hver femte (20 %) vil ramme den rigtige mulighed og blive noteret for et korrekt svar. Hvis man forestiller sig, at de 10 % dygtigste læsere faktisk finder frem til det rigtige svar på egen hånd, og at de øvrige 90 % giver op og gætter, vil ikke færre end 28 % forventeligt ramme det rigtige svar.

På de vilkår er det ikke overraskende, at opgaver, hvor højst 10 % af deltagerne svarer rigtigt, bliver sjældne. For at nå under 10 %-grænsen, skal en eller flere af de forkerte svarmuligheder (distraktorer) virke så fristende, at mange deltagere TROR, de har fundet det rigtige svar (sådan at de ikke bare gætter). Og den slags distraktorer er vanskelige at udforme når de stadig skal være entydigt forkerte (se indholdsafsnittet).

Problemet havde for mig at se kunne løses, eller i hvert fald gøres mindre, hvis man i stedet havde satset på polytome opgaver, hvor gættechancen bliver langt mindre. Hvis man kombinerer to MC-spørgsmål (eller cloze-spørgsmål med valgmuligheder), der hver har en gættesandsynlighed på 20 %, reduceres den samlede gættesandsynlighed til 4 % (20 % af 20 %).

En anden mulighed er at bruge opgaver, hvor der skal markeres en kombination af valgmuligheder. Hvis der blandt fem svarmuligheder skal markeres to, er der i alt ti forskellige kombinationsmuligheder – sådan at gættesandsynligheden bliver 10 %.

Begge de nævnte muligheder har, så vidt det fremgår, kunnet vælges inden for de tekniske rammer. Opgaver med to eller flere indgående spørgsmål i cloze-format skal dog holde sig under de ca. 1.000 enheder, som ifølge rammenotatet er det maksimale, når deltagerne skal kunne se hele opgaven uden at scrolle. Dette kan nok være en vanskelighed – men jeg vurderer ikke, at det er direkte umuligt.

Det ser således ud til, at opgaveudformerne ikke har set (eller ikke i tilstrækkelig grad er blevet gjort opmærksomme på), at meget svære opgaver forudsætter en lav gættesandsynlighed.

Når der blandt de forelagte opgaver med høj empirisk sværhedsgrad næsten kun er dikotome opgaver, kan det muligvis have sin forklaring i, at polytome opgaver if. rammenotatet s. 4 indgår i opgavebanken med (som jeg forstår det) de indgående spørgsmåls GENNEMSNITLIGE sværhedsgrad – hvor det relevante mål for mig at se havde været de polytome opgavers SAMLEDEsværhedsgrad.

En teknisk løsning, der ligger klart uden for de rammer, der har gjort sig gældende her, men alligevel bør nævnes mhp. videreudviklinger af formaterne, er at inddrage svartiderne. Dermed ville man ud fra afprøvninger af de enkelte opgaver kunne fastsætte en svartid, der afgrænsede de 10 % hurtigste svartider for korrekte svar. Dermed ville opgavernes gættesandsynlighed blive mindre afgørende for indfrielsen af kravene.

**Vedr. indholdsmæssige vanskeligheder**

De fleste af de forelagte opgaver, der faktisk opfyldte kravene, forekommer IKKE egnede som forbilleder for udarbejdelsen af yderligere vanskelige tekstforståelsesopgaver. Dette har især at gøre med det indholdsmæssige krav, at der ikke må være tvivl om, hvilket svar der er det rigtige (rammenotatet s. 20).

En lav andel af korrekte svar, skyldes i sagens natur, at de afgivne svar har været koncentreret om en eller flere af distraktorerne. I mange opgaver forekommer det mig imidlertid at valget af en eller flere distraktorer godt kan forsvares. Distraktorerne er mao. ikke altid entydigt forkerte. Eller: Distraktorerne kan være forkerte på en måde, der ikke har så meget med tekstforståelse at gøre – som når to distraktorer i opgave 010203050613617-2 har udskiftet det korrekte ord "frosten" med ortografiske nærliggende (men i sammenhængen meningsløse) ord, hhv. "frossen" og "fronten".

I forlængelse af denne generelle kritik (jeg vil undlade at gå nærmere ind i en analyse af konkrete opgaver) må det dog medgives at det ER en vanskelig sag at formulere distraktorer, der BÅDE er entydigt forkert OG fristende for en stor del af eleverne (sådan som det valgte format kræver, jf. ovenfor).

Man kan gøre gældende, at der er tale om opgaver, der har bestået kravene mht. opgavehomogenitet, sådan at det formentlig især er mindre gode læsere, der falder for de tvivlsomme distraktorer. Kravet om entydighed mht. rigtigt/forkert er imidlertid stadig meget vigtigt for at sikre at opgaverne OPLEVES som valide. Hvis 90 % af eleverne i en typisk 8. klasse ikke kan finde det rigtige svar i en opgave, vil der givetvis også være mange andre (inkl. dansklærere), der ikke uden videre finder det rigtige svar. Hvis der i sådanne situationer kan findes eksempler på at svar, der regnes som forkerte, faktisk IKKE er entydigt forkerte, så vil det påvirke testens oplevede validitet (face-validitet).

Dette er ikke så meget en faglig, som en rent pragmatisk grund til ikke at stå meget stejlt fast på et krav om svære opgaver, hvor ønskelige de end kan være psykometrisk set. Man kan nemt få en dårlig sag i dagspressen.

Selv hvor kravet om entydighed mht. rigtigt og forkert faktisk KAN indfris, kan de svære opgaver rejse diskussioner om validiteten, fordi den korrekte løsning af opgaver ofte vil kræve kognitive færdigheder, som ikke alle betragter som noget, der hører under begrebet læsefærdigheder, men snarere ville forbinde med intelligenstestning (fx at sammenholde udsagn fra en tekst og på den baggrund foretage følgeslutninger).

Min faglige vurdering er dog stadig, at det godt kan lade sig gøre at udarbejde meget svære opgaver. Det er ikke nemt, men med et hensigtsmæssigt valg af teknisk opgaveformat er det ikke umuligt.

Som et eksempel på en for mig at se vellykket opgave (selv i det mindre hensigtsmæssige MC-format) vil jeg fremhæve opgave 0104030515018. Her er der, så vidt jeg kan vurdere, ingen af distraktorerne der kan anklages for ikke at være forkerte (selv om flere af dem umiddelbart kan virke fristende), mens udvælgelsen af det rigtige svar vitterligt tekstforståelse: Det er i denne opgave en fordel, hvis man kan udlede (eller allerede ved), at dysser og jættestuer består af store sten; hvis man er med på, at svarets "de store sten blev solgt" modsvares af tekstens "der var gode penge i at fjerne de store sten"; og hvis man kender/udleder forbindelse mellem begrebet skærver og anlæggelsen af veje.

Summa summarum: Det KAN lade sig gøre - men det er ikke nemt, og med det valgte format har man gjort det endnu sværere.

**Med venlig hilsen**

Holger Juul

lektor, Institut for Nordiske Studier og Sprogvidenskab, Københavns Universitet

juul@hum.ku.dk