

Är det svårare att dela med fyra än med två när man läser matte C?

Ewa Bergqvist är filosofie doktor i matematikdidaktik och ingår för tillfället i ett forskningsprojekt som handlar om språk i prov.

Anna Lind Pantzare är doktorand i beteendevetenskapliga mätningar på Umeå universitet och arbetar dessutom med de nationella proven i matematik för gymnasieskolan.

Inledning

Våren 2004 infördes på försök två olika versioner av de nationella kursproven i matematik, på kurserna B, C och D. Året innan var ett av proven ute till försäljning dagarna innan provdagen, och målet med dubbla provversioner var att minska riskerna för och följderna av eventuellt fusk. De båda versionerna distribuerades till olika skolor vilket gjorde att eleverna inte visste vilken version de skulle möta på provdagen. Skolverkets tanke är att konstruktionen av två likvärdiga provversioner ska minska sårbarheten i det nationella provsystemet.

De nationella kursproven i matematik har flera syften. De ska bland annat ge läraren stöd i bedömning av om eleverna har uppnått uppställda mål i kursplanerna. De ska även stödja en likvärdig bedömning och en rättvis betygssättning. Det är därför viktigt att provversionerna är parallella, dvs. att det provbetyg en elev uppnår inte beror på vilken provversion han eller hon har blivit tilldelad, utan på de faktiska kunskaper eleven har. De båda provversionerna har i dessa fall identiska inre strukturer och poängsättningar, och dessutom behandlar uppgifterna på samma position samma typ av matematisk kunskap.

Syftet med denna studie var att undersöka om, hur och varför olika typer av förändringar i innehållet i uppgifter i form av kontext och tal påverkar uppgifternas svårighetsgrad.

Bakgrund

Kursproven i matematik B, C och D består samtliga av två delar. Den första delen innehåller uppgifter som ska lösas utan räknare, medan räknare är nödvändiga och tillåtna som hjälpmedel på den andra delen. Den sista uppgiften i provet är en större uppgift som läraren ska aspektbedöma. Det innebär att läraren bedömer kvalitetsnivån på varje elevs lösning utifrån tre olika kvalitetsaspekter; *metodval och genomförande*, *matematiskt resonemang* samt *redovisning och matematiskt språk*. Beroende på vilken typ av kompetens som eleven i huvudsak har möjlighet att visa prov på i lösningen av en viss uppgift tilldelas uppgiften g- och/eller vg-poäng. Denna bedömning görs utifrån kursmål och betygskriterier i styrdokumentet. Vissa uppgifter som ingår i proven är $\square$ -markerade (sol-markerade). Denna markering indikerar att eleven i sin lösning av denna uppgift har större möjlighet, än i lösningar till övriga uppgifter, att visa prov på kunskaper motsvarande betygskriterierna för Mycket väl godkänd.

Vid konstruktionen av de båda provversionerna utarbetades först en version av provet på vanligt sätt. Det innebär bland annat att provet granskas i flera omgångar och ur olika perspektiv av rutinerade lärare som anlitas av Institutionen för beteendevetenskapliga mätningar. Därefter fastställdes de kravgränser som skulle gälla för provet. För mer information om provkonstruktionsprocessen se Lindström & Nyström (1996). Den andra versionen konstruerades som en modifiering av den första genom att en del av de ingående uppgifterna förändrades. De förändringar som gjordes var av olika karaktär men gjordes för

alla tre kursproven på ett sådant sätt att de båda versionerna har exakt samma fördelning poängmässigt. Dessutom användes samma kravgränser för båda versionerna av respektive kursprov. För en utförligare beskrivning av arbetet med Nationella prov se Lindström (2003).

Det finns ett antal uppgifter som är identiska i de båda provversionerna och som dessutom har samma position i respektive provversion. Dessa uppgifter kallas *ankaruppgifter*. Antalet ankaruppgifter och poängen på dessa skiljer sig mellan de olika kursproven. Ankaruppgifterna används i den här studien för att kunna jämföra de olika provversionerna trots att elevgrupperna har olika sammansättning och olika resultat på proven som helhet.

Uppgifter som skiljer sig åt men är tänkta att vara parallella kallas för *parvis motsvarande uppgifter*. De har samma position i proven, samma poängsättning, samma matematiska innehåll och ofta samma formulering.

De skillnader som finns mellan två parvis motsvarande uppgifter går att dela upp i två olika typer. Den första typen består av *skillnad mellan de ingående talen* i uppgifterna. Parvis motsvarande uppgifter som skiljer sig åt på detta sätt har identiska formuleringar, så när som på de tal som ingår och i ett enda fall ett namn. Uppgifterna kan antingen vara inommatematiska eller ha kontext. Det finns ett flertal uppgiftspar i varje prov som skiljer sig åt på detta sätt. Den andra typen består av *skillnad i kontext*. Det finns ett par parvis motsvarande uppgifter i vardera B- och C-provet som skiljer sig åt på detta sätt. Det matematiska innehållet är detsamma i de båda versionerna av uppgiften, men situationen och beskrivningen av situationen är olika.

Data

De data vi använt oss av i denna studie är den insamling som SCB genomförde varje termin för de nationella proven. I denna datamängd ingår ungefär en sjättedel av Sveriges skolor och alla elevers resultat på dessa skolor rapporteras in.

Metod

För att kunna påstå att två motsvarande uppgifter faktiskt fyller samma funktion i ett prov är det många olika faktorer som bör vara uppfyllda. Uppgifterna bör ha samma egenskaper ur ett statistiskt perspektiv, men naturligtvis också likna varandra innehållsmässigt och pröva samma kompetens se till exempel (Crocker & Algina, 1986; Henriksson, Sundström, & Wiberg, 2004). Som tidigare nämnts är de parvis motsvarande uppgifterna i de prov som undersöks i denna rapport mycket lika. Till exempel har de samma position i provet och samma poängfördelning. De har också samma matematiska innehåll och för det mesta samma formulering. Vi definierar i denna studie att *uppgifterna är parallella om de har samma statistiska egenskaper*.

För att kunna jämföra två uppgifters svårighetsgrad använder vi uppgiftskaraktäristiska kurvor (Lindström, Nyström, & Palm, 1996). Dessa tas fram genom att man beräknar sannolikheten för att elever på olika poängnivå lyckas lösa en uppgift. Till dessa värden anpassar man sedan, med hjälp av regression, logaritmiska funktioner av en viss typ, nämligen

$$P_k(x) = p_k \cdot \frac{e^{a(x-b)}}{1 + e^{a(x-b)}}$$

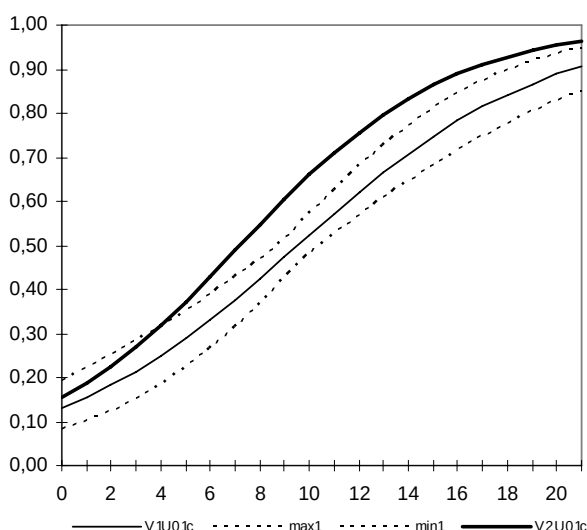
där funktionsvärdet $P_k(x)$ betecknar sannolikheten för att en elev med poängnivå x ska lyckas lösa uppgift k . Parametern p_k betecknar uppgiftens totalpoäng medan parametrarna a och b skattas vid regressionen. Parametern a är lutningen på kurvan där den är som brantast.

Parametern b är ett mått på uppgiftens svårighetsgrad och anger den provpoäng som krävs för att få en förväntad lösningsproportion på 0,5. Grafen till denna funktion kallas uppgiftens *uppgiftskaraktäristiska kurva*. Dessa kurvor används för att undersöka om de parvis motsvarande uppgifterna har olika svårighetsgrad.

Resultat

Vid en sammanställning av de resultat vi fått så kan man konstatera att en förändring av de ingående talen i en uppgift i de flesta fall inte påverkar uppgiftens svårighetsgrad alls. De gånger när en skillnad faktiskt visar sig gäller det oftast lätta g-uppgifter, det vill säga uppgifter som de svagare eleverna ofta gör ett ordentligt försök att lösa, på den del av provet där eleverna inte får använda räknare. Något som också är märkbart i dessa sammanhang är elevernas problem med att hantera bråk när de inte har tillgång till räknare.

En intressant uppgift är det första av de parvis motsvarade uppgifterna på C-kursprovet de har följande uppgiftskaraktäristiska kurvor.



Figur 9. Grafer till uppgiftsparet på position 1c, C-kursprovet, vt04.

De parvis motsvarande uppgifterna på denna position gav eleverna i uppgift att derivera ett bråkuttryck. Uppgiften återfinns som en del av första uppgiften på den miniräknarfria delen. Uttrycket i de två versionerna av uppgiften hade samma täljare men olika nämnare. Nämnaren var i båda fallen en konstant som i version 1 var fyra (4) och version 2 var två (2).

Enligt de uppgiftskaraktäristiska kurvorna ovan var version 2 lättare för de flesta eleverna. Det är bara för eleverna längst ner på poängskalan som det inte är någon skillnad. Troligtvis är det så att de svagaste eleverna inte löser uppgiften i någon större utsträckning oavsett vilken nämnare bråket har. Det kändes vid en första anblick underligt att en skillnad i konstantterm i nämnaren skulle göra en sådan tydlig skillnad mellan svårighetsgraden för de olika versionerna. Det visade sig dock vid en genomgång av tillgängliga elevlösningar att när nämnaren är 2 så har eleverna i flera fall skrivit om bråket som ett decimaltal. Denna metod har vi inte kunnat se eleverna välja när nämnaren varit 4. Det är uppenbarligen lättare för eleverna att hantera ett bråkuttryck när man delar med 2 än när man delar med 4. Omskrivningen kan underlätta den efterföljande deriveringen betydligt och vi tror att denna skillnad är orsaken till skillnaden i svårighetsgrad mellan dessa två versioner av uppgiften. Troligen skulle denna skillnad inte ha påverkat svårighetsgraden om uppgiften legat på den del av provet där räknare är tillåtet som hjälpmedel.

När det gäller olika kontext i olika uppgiftsversioner så verkar det, av de få exempel vi analyserat, inte vara omöjligt att få parallella uppgifter på detta sätt. Om upplägget i uppgiften, det vill säga hur problemet presenteras, är lika så är det möjligt att själva kontexten inte spelar så stor roll.

Vår bedömning är att skillnaden i provpoäng mellan två lika duktiga elever som tagit olika versioner av provet är mycket liten. Detta innebär att det inte borde spela någon roll vilken version av provet som eleven fått, resultatet blir detsamma vad gäller uppnått provbetyg.

För en komplett redovisning av resultatet se Bergqvist och Lind (2005).

Litteratur

- Bergqvist, E., & Lind, A. (2005). *Är det lättare att dela med fyra än med två när man läser matte C? En jämförelse av svårighetsgrad mellan olika versioner av matematikuppgifter i Nationella kursprov* (BVM Nr 12). Umeå universitet: Institutionen för beteendevetenskapliga mätningar.
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Belmont: Wadsworth Group.
- Henriksson, W., Sundström, A., & Wiberg, M. (2004). *The Swedish Driving-Licence Test, A summary of studies from the Department of educational measurement, Umeå university* (EM no 45). Umeå: Department of Educational Measurement.
- Lindström, J.-O. (2003). *The Swedish National Course Tests in Mathematics* (EM no 43). Umeå: Department of Educational Measurement.
- Lindström, J.-O., & Nyström, P. (1996). *Nationellt kursprov i matematik Kurs A, Ht -95. Resultat och kommentarer*. (PM Nr 112). Umeå Universitet: Enheten för pedagogiska mätningar.
- Lindström, J.-O., Nyström, P., & Palm, T. (1996). *Nationellt kursprov i matematik. Kurs A, C och E, VT -96. Resultat och kommentarer* (PM Nr 117). Umeå universitet: Enheten för pedagogiska mätningar.