

313a

Hur lika bedömer matematiklärare?

Anna Lind Pantzare, Institutionen för beteendevetenskapliga mätningar, Umeå universitet. Provtutvecklare av de nationella proven i matematik för gymnasieskolan och doktorand i beteendevetenskapliga mätningar.

Likvärdigheten i bedömningen av elevers kunskaper och den efterföljande betygssättningen har på senare år diskuterats livligt. Förutsättningarna för att göra en likvärdig bedömning av elevers kunskaper varierar dock mellan skolans olika ämnen. Dels ser kursplanerna olika ut och dels är tillgången på stödjande material varierande. Hur enkelt är det då att bedöma lika i matematik? De finns de som hävdar att matematiken är så enkel att bedöma eftersom uppgifterna genererar ett svar som antingen är rätt eller fel. Jag vill hävda att så enkelt är det inte. Även matematiken innehåller en del svårbedömda uppgifter.

I några ämnen däribland matematik finns nationella prov med tillhörande bedömningsanvisningar. Uppgifterna som ingår i dessa prov är definitivt inte bara av typen som ger rätt eller fel svar. Många av uppgifterna kräver att eleverna ger mer resonerande svar vilket innebär att både elevens svar men även bedömningsanvisningarna kräver en tolkning av läraren. Bedömningsanvisningen till dessa öppna uppgifter tenderar att konstrueras så att de kräver en tolkning av läraren. De öppna uppgifterna har oftast inte endast en korrekt metod med endast ett rätt svar utan det finns olika sätt att angripa problemen. Då kan inte heller bedömningsanvisningen endast innehålla ett konkret exempel utan de måste skrivas på ett sådant sätt att alla acceptabla lösningar och svar ska godkännas. Bedömningsanvisningarna ska utgöra ett stöd för den bedömning som läraren ska göra av elevernas svar.

Hur lika bedömer då matematiklärare utifrån de givna bedömningsanvisningarna i de nationella proven?

Under den senaste tioårsperioden har fyra svenska studier av bedömaröverensstämmelse i de nationella matematikproven presenterats (Boesen, 2004; Lindström, 1998; Olofsson, 2006; Palm, 2008). Den första studien (Lindström, 1998) byggde på omdömning av 58 elevarbeten från det nationella kursprovet för matematik E våren 1997. Några år senare gjordes en liknande studie (Boesen, 2004) men denna gång handlade det om ett nationellt prov i matematik B. Syftet med studien var att undersöka bedömarreliabiliteten vid de nationella proven i matematik, med särskilt fokus på den så kallade aspektbedömda uppgiften i de nationella proven i matematik. Olofssons (2006) studie av bedömaröverensstämmelse utgick från 100 elevarbeten, slumpvis valda bland arbeten som skickades in från landets skolor i samband med resultatrapporteringen för kursprovet i matematik A våren 2001. Den fjärde svenska studie av bedömarreliabilitet som presenteras här gäller interbedömarreliabiliteten i ett muntligt prov som utvecklats för gymnasieskolans matematik C (Palm, 2008). Palm konstaterar att muntliga prov av ämneskunskap eller kommunikationsförmåga i allmänhet är behäftade med reliabilitetsproblem, vilket orsakar svårigheter med dess användning. Reliabilitetsstudier om muntliga prov i matematik är dock relativt ovanliga.

Dessa studier av nationella prov för olika kurser i gymnasieskolans matematik visat på en relativt god bedömaröverensstämmelse. Föga överraskande är det dock så att ju mer av tolkningar och förklaringar som eleverna förväntas göra vid lösandet av provuppgifterna och ju mer bedömningen gäller sådant som redovisningens kvalitet, desto större blir variationerna

mellan bedömarna. Att detta i ännu högre grad gäller ett muntligt prov som har en större komplexitet är inte heller anmärkningsvärt, men det bör påpekas att den studie som presenterats gäller en modell för bedömning av muntlighet i matematik som sedan utvecklats vidare.

Den allra vanligaste utgångspunkten i studier av bedömaröverensstämmelse är att söka svaret på frågan ”hur olika är bedömningsresultaten från olika bedömare?”. Det är hot av detta slag mot tolkningen av provresultat som i allmänhet avses med termen interbedömarreliabilitet. Bedömares överensstämmelse i absoluta termer kan mätas med en enkel parvis procentuell överensstämmelse, dvs. en beräkning av hur stor andel av bedömningarna som är lika mellan två bedömare. Hur konsekventa lärare är i sin bedömning kan studeras med korrelationer. Observera att en låg korrelation betyder att bedömarna i hög grad är oense om bedömningarna. En hög korrelation behöver inte betyda att två bedömare är överens om enskilda bedömningar, utan visar primärt bara att de två bedömarna har ungefär samma rangordning av de bedömda elevsvaren.

För att på nytt undersöka bedömaröverensstämmelsen genomförde Skolverket hösten 2008 ett antal studier på nationella prov där ibland ett nationellt prov i matematik C från våren 2007 (Nyström & Lind Pantzare, 2009). Det analyserade provet består av 16 uppgifter. Vissa av dessa uppgifter består av en eller flera deluppgifter vilket innebär att totalt 26 deluppgifter har analyserats. Deluppgifterna kan ge mellan en till tre poäng. I matematikproven finns g- och vg-poäng, vilka kan kopplas till betygskriterierna. Dessutom finns ett antal uppgifter som är markerade med \boxtimes , vilka ger en större möjlighet än övriga uppgifter att visa på kvaliteter knutna till betygskriterierna för MVG.

Tre bedömare har genomfört ombedömningen i denna studie. Bedömare 1 är en erfaren lärare som sedan flera år arbetar som provutvecklare vid Institutionen för beteendevetenskapliga mätningar. Bedömare 2 är en erfaren lärare som dessutom har mångårig erfarenhet som granskare i referensgrupper, dvs. de grupper sammansatta av lärare som granskar de nationella proven. Bedömare tre är en erfaren lärare som aldrig varit inblandad i något arbete med granskning av de nationella proven.

I denna studie av bedömaröverensstämmelse används såväl procentuell överensstämmelse som korrelationer för att söka svaret på frågan om hur olika lärare är i sina bedömningar. För detta syfte är det nödvändigt att låta olika bedömare bedöma ett relativt stort antal elevsvar. I denna studie är det i första hand de tre externa bedömarna som kan användas. Varje elevsvar har dessutom en bedömning som är gjord av någon lärare på elevens skola. Dessa lärare har ju i de allra flesta fall endast bedömt ett enda elevsvar och kan därför inte användas för interbedömarreliabilitetsstudier, eftersom fokus är på hur olika enskilda bedömare bedömer. Det kan dock vara intressant att betrakta alla dessa bedömningar som om de var gjorda av en lärare, för att undersöka hur denna elevnära bedömning förhåller sig till de övriga (externa) bedömarna. Jämförelsen bör dock göras med försiktighet.

I vår studie har vi i första hand analyserat provbetygen, men vi har även studerat olikheter mellan bedömarna när det gäller provpoäng (totalpoäng) och även poäng på olika uppgifter. Det sistnämnda syftar i första hand till att identifiera uppgifter som tycks särskilt svårbedömda, och eventuellt komma fram till någon tänkbar förklaring till detta utifrån bland annat bedömningsanvisningen.

Granskningen av bedömaröverensstämmelsen visar acceptabla eller till och med bra nivåer för det prov i matematik C som ingår i studien. I provet ingår en hel del kortsvarsuppgifter där bedömningen är entydig och där bedömarna i princip är helt eniga. Dessutom är bedömningen relativt överensstämmande för de uppgifter som kräver en längre lösning eller ett mer resonerande svar, de flesta uppgifterna har en procentuell överensstämmelse på över 90 %.

Generellt sett är matematikens bedömningsanvisningar relativt utförliga. Det finns oftast en beskrivning för varje poäng även om den kan vara ganska generell. Det beror dock oftast på att det finns flera olika lösningsmetoder som är acceptabla. Vissa skillnader är naturliga på grund av att det finns en frihet i bedömningen. I och med att lärarna förväntas bedöma utifrån den undervisning som skett så kommer det att uppstå vissa skillnader. Det innebär att om överensstämmelsen skulle vara för hög så finns inget sådant friutrymme. Denna brist i överensstämmelse kan inte anses vara ett interbedömarreliabilitetsproblem eftersom vi har det system vi har. En mer styrande bedömningsanvisning kan troligen öka bedömaröverensstämmelsen, men kommer samtidigt att minska lärarens möjligheter att göra lokala tolkningar av målen.

Skillnader i bedömningen vad gäller elevernas användning av räknare är inget som har varit möjligt att granska i denna studie eftersom det inte funnits någon elevlösning där eleven tydligt använt sin räknare. Det är däremot något som diskuterats ingående i samband med provkonstruktionen. Till vissa uppgifter är det naturligtvis helt i sin ordning att eleverna använder sitt hjälpmedel i form av räknare. Det finns däremot inga principer för hur eleverna ska redovisa detta. Det här är något som kommer att bli mer intressant när andelen elever med de mer avancerade symbolhanterande räknarna kommer att öka. Det blir naturligtvis ingen orättvisa inom varje enskild klass men mellan klasser kan det uppstå skillnader.

Det som överlag verkar orsaka störst problem att bedöma är uppgifter där eleven ska göra någon form av resonemang och där läraren ska bedöma elevernas språk. Det finns bland lärarna ganska olika synsätt på vad eleven måste göra för att få poäng och anses uppfylla MVG-kriterier för sitt resonemang och språk. Ofta finns det till denna typ av uppgifter bedömda elevsvar med kommentarer men det är inte alltid helt lätt att hitta lösningar som "ligger på gränsen". Dessutom finns det till vissa uppgifter nästan lika många formuleringar av svar som det är elever, vilket gör att det i slutändan är upp till läraren att göra en tolkning av bedömningsanvisningen och relatera den till den undervisning som föregått provet. Har man inte lärt eleverna att göra ett korrekt bevis så är det naturligtvis svårt för eleven att prestera ett sådant på ett prov. Å andra sidan har man höga krav på eleverna så kommer det att avspeglas även i bedömningen av de nationella proven. Detta kommer att generera en skillnad mellan bedömarna. Kursplanerna ger inte heller här något stöd för vad som är det korrekta utan det är upp till de lokala tolkningar som gjorts. För att uppnå högre överensstämmelse vad gäller resonemang samt redovisning så är det säkert möjligt att försöka förtydliga bedömningsanvisningarna vilket vi arbetar med hela tiden. Dessutom kan säkert informationen i de bedömda elevlösningarna förbättras ytterligare. Om dessa åtgärder är tillräckliga är däremot osäkert. Problemet med uppgifter som genererar ett resonerande svar är att beroende på den lokala tolkning läraren gjort av kursplanen så kommer olikheter att uppstå.

Vad ska man då göra om man vill öka överensstämmelsen mellan olika bedömare i matematik? För det första kan man som angett ovan specificera bedömningsanvisningarna ytterligare men då uppstår en risk att det sker en begränsning i de lösningsmetoder som undervisas. För det andra skulle sannolikt skillnader i bedömningen vad gäller elevernas

metoder vid lösandet av uppgifterna kunna minskas om bedömningsanvisningarna i vissa fall inte bara beskrev vilka kvaliteter som ska ge poäng, utan också specificerade vad eleven inte får göra. För det tredje skulle en träning av bedömarna kunna genomföras. Stemler (2004) nämner vid ett flertal tillfällen i sin artikel vikten av att träna bedömarna. Detta är någonting som är en viktig del i t.ex. bedömningen av de stora internationella komparativa studierna så som TIMSS och PISA. Där får bedömarna träna på att bedöma ett antal elevhäften innan den egentliga bedömningen startar, allt för att öka överensstämmelsen i de bedömningar som görs. Det skulle vara möjligt att göra något motsvarande till de nationella proven. Till bedömningsanvisningen skulle man kunna bifoga ett antal bedömda elevlösningar som lärarna förväntas träna på och diskutera med övriga lärare inom ämnet innan den egentliga rättningen startar. Det skulle sannolikt öka överensstämmelsen i bedömningen. Det kräver dock betydligt mer utförligt bedömda elevlösningar än vad som är fallet i dagens bedömningsanvisningar.

Syftet med detta arbetsseminarium är att undersöka om det blir en bättre överensstämmelse om man tränar de som ska bedöma elevernas lösningar. Tanken är att deltagarna till att börja med ska få rätta ett antal elevlösningar till en given uppgift. Därefter sker en diskussion om bedömningen för att därefter efterföljas av en ny omgång av elevlösningar som ska bedömas.

Referenser

- Boesen, J. (2004). *Bedömarreliabilitet. Med fokus på aspektbedömningen i det nationella B-kursprovet i matematik våren 2002* (No. 195). Umeå: Enheten för pedagogiska mätningar.
- Lindström, J.-O. (1998). *Rättvis rättning i nationella prov* (Pm No. 144). Umeå: Umeå universitet, Enheten för pedagogiska mätningar.
- Nyström, P., & Lind Pantzare, A. (2009). *Bedömaröverensstämmelse för ett nationellt prov i matematik C*. Hämtad 2009-11-26 från <http://www.skolverket.se/publikationer?id=2200>
- Olofsson, G. (2004). *Likvärdig bedömning? En studie av lärares bedömning av elevarbeten på ett nationellt prov i matematik kurs A*. Stockholm: Stockholms universitet: PRIM-gruppen.
- Palm, T. (2008). *Interrater reliability in a national assessment of oral mathematical communication*. *Nordic Studies in Mathematics Education*, 13(2).