

# SJÄLVSTÄNDIGT ARBETE

*15 Högskolepoäng, avancerad nivå, magister  
Masterprogrammet i utbildningsvetenskap 120 hp*

***Våren 2010***

*Sektionen för Lärarutbildning  
Pedagogiskt arbete*

# **Ett acceptabelt koordinatsystem**

**Om bedömning av nationella prov med stöd av  
bedömningsmatriser**

**Författare**

Pia Thornberg

**Handledare**

Lars-Erik Nilsson

## Förord

Detta är ett arbete utfört inom ramen för Masterprogrammet i utbildningsvetenskap 120 hp. Kursen heter Pedagogiskt arbete - självständigt arbete, 15 hp och ligger på avancerad nivå. Kursen ges som avslutande del inom programmets första 60 hp som motsvarar magisternivån. Enligt kursplanen ska studenten för sin examination välja en etablerad vetenskaplig genre. Uppsats betraktas som en genre som endast används inom utbildningskontexter och kan inte väljas. Studenten ska enligt kursanvisningar till sitt arbete medsända ett missiv där vald genre och villkoren för publiceringen presenteras. Missivet för varje arbete återfinns som en bilaga, sist i arbetet.

I missivet presenteras t.ex. sådant som:

- vilken tidskrift den är tänkt för,
- vilka områden tidskriften riktar sig mot och
- vilka anvisningar tidskriften ger till sina skribenter.

## Ur aktuell kursplan MUM60L, Dnr: 850/333-08

### **”Förväntade läranderesultat**

#### *Kunskap och förståelse*

Efter genomgången kurs ska studenten

- visa fördjupade kunskaper inom den valda problemställningen utifrån dess teoretiska och empiriska kontext
- visa fördjupade kunskaper om vetenskapliga metoder och konsekvenser vid tillämpning av dessa.

#### *Färdighet och förmåga*

Efter genomgången kurs ska studenten

- ha förmåga att identifiera, formulera, bearbeta och analysera problemställningar inom ett specifikt problemområde inom pedagogiskt arbete
- kunna värdera, välja och använda relevanta metoder och arbetssätt
- kunna använda informationskompetens på sätt som kännetecknas av relevans och kvalitet i relation till specifikt kunskapsområde
- kunna rapportera vetenskapliga arbeten skriftligt och muntligt enligt aktuella akademiska traditioner samt muntligt i lämplig form för en bredare målgrupp.

#### *Värderingsförmåga och förhållningssätt*

Efter genomgången kurs ska studenten

- utifrån vetenskaplig grund och med en tydlig självständighet kunna kritiskt granska det egna och andras arbeten
- visa fördjupad förmåga att utifrån ett kritiskt förhållningssätt analytiskt och konstruktivt diskutera konkreta forskningsarbeten
- visa medvetenhet om enskilda vetenskapliga arbetens bidrag och begränsningar till kunskapsområdet
- kunna göra överväganden kring vad det ökade personliga kunnandet kan föra med sig i relation till omgivning och personlig utveckling, t.ex. i termer av ansvar och frågor som rör identitet.

### **Genomförande**

Inom kursen ska studenten genomföra ett vetenskapligt arbete i pedagogiskt arbete med kopplingar till den forskningsmiljö varinom studenten har sin handledning eller vid annat nationellt eller utländskt lärosäte som forskningsmiljön har samarbete med.

Arbetet kan utföras ensamt eller som en del av ett samarbete med andra

studenter/forskare. Arbetet innehåller en rad självständiga moment och situationer där studenten kan utveckla sin förmåga att inta ståndpunkter och argumentera för dessa.

Detta sker bland annat i handledning, när studenten diskuterar med andra studenter och forskare om sin egen och deras pågående forskning. Studenten ska under handledning genomföra en empirisk studie och rapportera den i en etablerad vetenskaplig genre.

Studenten erbjuds under kursen att delta i seminarier och workshops dels i den forskningsmiljö studenten är knuten till samt vid andra forskningsmiljöer. Studenten ges ett tillfälle att presentera det pågående självständiga i den egna forskningsmiljön. Som komplement till handledningen förekommer även föreläsningar och workshops samt tillgång till bibliotekets sökverkstad.”

# Ett acceptabelt koordinatsystem: om bedömning av nationella prov med stöd av bedömningsmatriser

Pia Thornberg

*Mastersstudent, Högskolan Kristianstad, Sverige*

I dagens kursplaner betonas att elever ska utveckla komplexa förmågor och inte enbart fakta och färdigheter. Vid bedömning av sådana komplexa förmågor ställs krav på de bedömningsverktyg som används att de ska kunna bedöma dessa på ett tillförlitligt sätt. Ett verktyg för ändamålet som ofta lyfts fram är bedömningsmatriser. I denna studie analyseras filmad interaktion mellan lärare för att synliggöra hur lärare skapar mening kring sin uppgift att bedöma mer omfattande elevarbeten med stöd av bedömningsmatriser, vilka överväganden de gör och vilka dilemman som uppstår. Resultaten visar att lärare vid bedömningen förutom de instruktioner som tillhandahålls även lyfter in andra faktorer som blir avgörande vid bedömningen. Utifrån resultatet förs även en diskussion kring utveckling av bedömarkompetens hos lärare.

## Introduktion

Användandet av olika typer av bedömningsmatriser har de senaste åren fått stort genomslag i det svenska skolsystemet på olika nivåer. Exempel på detta är att de används som en del av bedömningsunderlaget vid nationella prov i både matematik och naturvetenskap, på lärarutbildningar i samband med examinationer och i många svenska kommuner görs för närvarande stora satsningar på att fortbilda pedagoger i bedömningsfrågor där utarbetandet av bedömningsmatriser fått en central roll. Bakgrunden till detta ligger i en tilltro till att bedömningsmatriser ska lösa problematiken med att bedöma komplexa förmågor på ett trovärdigt sätt.

Denna studie syftar till att ge ytterligare ett perspektiv på bedömning med bedömningsmatriser och tar sin utgångspunkt i teorier om bedömning och användning av matriser som en situerad aktivitet. Filmad interaktion mellan lärare används och det görs en analys av hur lärare skapar mening kring sin uppgift att bedöma mer omfattande elevarbeten utifrån bedömningsmatriser, vilka överväganden de gör och vilka dilemman som uppstår. Ytterligare utgångspunkter för studien hämtas från några grundläggande antaganden om utvecklingen av bedömningsformer. Det är idag ett stort, både nationellt och internationellt, fokus på frågor som rör bedömning och kvalitetskontroll i utbildning. Det svenska utbildningssystemet står mitt i en omfattande förändringsprocess som bl.a. innebär ny skollag, ny läroplan med tillhörande kursplaner för samtliga skolformer och ny betygsskala. Sverige deltar även i samarbeten med andra länder, t.ex. TIMSS och PISA, för att kunna göra internationella jämförelser av elevers kunskaper. På nationell nivå görs motsvarande jämförelser mellan klasser, skolor och kommuner utifrån resultat på nationella ämnesprov. I landets skolor pågår det även omfattande arbete med införandet av IUP (individuella utvecklingsplaner) med skriftliga omdömen i alla ämnen redan från tidiga skolår.

Samtidigt som nya kontrollsystem införs finns ett intresse att utvärdera dessa. Skolinspektionen (2010) har på uppdrag av regeringen genomfört en central rättning av ett urval av de nationella ämnesproven i syfte att stödja en likvärdig bedömning och betygssättning av proven samt att kvalitetssäkra provsystemet för att motverka

felbedömningar som kan äventyra dess tillförlitlighet. Resultatet av denna omrättning visar att det finns betydande avvikelser mellan den bedömning som gjorts av ursprungs rättarna och den som gjorts av kontrollrättarna och att avvikelserna är som störst i den typ av uppgifter där eleverna fritt får formulera sina svar. I ämnesprovet för åk 9 i matematik ingår en omfattande uppgift som prövar elevens förmåga att lösa problem, reflektera över och tolka sina resultat samt bedöma resultatens rimlighet. Vid omrättningen visade det sig att endast i drygt hälften av fallen hade ursprungs rättarna och kontrollrättarna gjort samma bedömning.

Det nya ”provparadigm” (Korp, 2003) som fått genomslag i bedömningsforskningen sedan 1980-talet har bl.a. inneburit en förändrad syn på kunskap och lärande som bygger på social-konstruktivistiska eller sociokulturella teorier. Denna syn på kunskap, vilken kommer till uttryck i skolans styrdokument (Utbildningsdepartementet 1998; Skolverket 2000), beskriver kunskap som sociala konstruktioner som hjälp att förstå den värld vi befinner oss i och skapa mening och sammanhang i denna (SOU 1992). Det nya paradigmet har även medfört ett ökat intresse för provens pedagogiska kvaliteter snarare än dess mättekniska kvaliteter (Broadfoot, 1996; Shepard 2000). Den läroplan med tillhörande kursplaner som kom 1994 innebar även ett vidgat kunskapsbegrepp och var ett tydligt steg bort ifrån en tidigare syn på kunskap som ensidig fokusering på faktakunskaper och färdigheter. Den nya läroplanen kom även att inkludera förståelse och förtrogenhet. I kursplanerna kan man hitta spår av denna förändring i de så kallade strävansmålen. Strävansmålen i matematik innebär bl.a. att eleverna ska utveckla problemlösningsförmåga, dra slutsatser, jämföra, tolka, argumentera samt kunna uttrycka sig i tal och skrift. Detta har medfört att prov som endast är riktade mot återgivning av fakta inte längre är relevanta i förhållande till de uppsatta undervisningsmålen (Korp 2003).

Att bedöma elevers prestationer i syfte att synliggöra flera olika typer av förmågor ställer helt andra krav på uppgiftsformat och bedömningsverktyg. 1998 användes i de nationella ämnesproven i matematik för första gången skriftliga uppgifter av mer omfattande karaktär i syfte att bedöma elevernas förmåga att använda sin matematiska kompetens för att upptäcka samband, föreslå lösningar, reflektera över resultat, dra slutsatser och visa klarhet och noggrannhet vid redovisningen (Skolverket 1998). Vid bedömning av alla typer av uppgifter ställs krav på att bedömningsverktygen de ska kunna synliggöra den typ av kunskaper och förmågor som uppgiften avser att pröva (Gipps 1994). Ett exempel på ett bedömningsverktyg som på senare tid fått stort genomslag vid bedömning av komplexa förmågor är bedömningsmatriser. 1999 användes vid de nationella ämnesproven i matematik i åk 9, för första gången en bedömningsmatris som underlag för bedömning av en av provets uppgifter (Kjellström 2000). Bedömningsmatrisen som användes är exempel på en analytisk bedömningsmodell, dvs. den fokuserar olika kunskapsaspekter/kriterier samtidigt som olika kvalitativa nivåer beskrivs inom varje kriterie. Det finns studier (Jönsson och Svingby 2007) som lyfter fram att användandet av bedömningsmatriser kan öka tillförlitligheten vid bedömning om matrisen är analytisk, uppgiftsspecifik och kompletteras med exempel på bedömda elevarbeten samt om lärare får bedömningsträning. Andra studier (Wyatt-Smith et al. 2010) visar att lärare i sitt arbete med att bedöma utifrån kriterier, förutom att använda sig av olika skriftliga artefakter så som bedömningsanvisningar och exempel på elevarbeten även använder s.k. ”tyst kunskap” av olika slag t ex kunskaper om enskilda elever och vad ”en genomsnittlig elev” i en given årskurs bör kunna.

Då bedömningsmatriser används i stor omfattning och i flera sammanhang där det har stor betydelse för elevers framtida möjligheter och då anspråk görs på att matriser ökar tillförlitligheten vid bedömning av omfattande uppgifter finns ett behov av mer kunskap om

hur lärare går till väga när de bedömer sådana omfattande uppgifter med stöd av bedömningsmatriser.

Vad blir föremål för lärares överväganden vid bedömning med bedömningsmatris?

Vilka resurser tar lärare in vid bedömning med bedömningsmatris?

## Bedömning

Traditionellt sett har kunskapsbedömning haft som syfte att sortera och rangordna elever för att göra ett urval till högre utbildningar och anställningar (Korp 2003). Formerna för bedömning har haft sin grund i den psykometriska traditionen som utgått ifrån antaganden om att intelligensen i en population är normalfördelad med några få individer i skalans topp och botten och med flertalet fördelade där emellan. I Sverige fick de sitt genomslag i mitten av 1900-talet (Korp 2003) i samband med införandet av ett rättvist system för urval till realskolan som skulle grundas på betyg och inte som tidigare, på antagningsprov. I många länder använder man examensprov som är direkt avgörande för vilket betyg en elev ska få i ett ämne, sk. "high-stake prov". I Sverige har vi nationella ämnesprov som bl.a. ges i årskurs 9 och vilkas resultat ska vara en del av lärarens samlade bedömning vid betygsättningen (Skolverket, 2009). De har därmed inte som enskilt prov samma avgörande betydelse för slutbetyget som det har i flera andra länder. Dock kan det lokalt avgöras hur stor del ett nationellt ämnesprov ska utgöra av det underlag som ligger till grund för det slutgiltiga betyget. En synlig tendens i rapporter från Skolverket (2007) och i medias rapportering<sup>1</sup> är sambandet mellan elevers resultat på de nationella ämnesproven och deras slutbetyg. Detta har satts i fokus på senare tid, vilket kan tolkas som att de nationella ämnesproven i årskurs nio är att betrakta som "high-stake" för elever, då t.ex. intagningen till gymnasieskolan sker utifrån detta slutbetyg. I samband med att den svenska grundskolan 1994 fick nya styrdokument skedde inte endast en förändring av kunskapssynen, utan vi fick även ett helt nytt betygssystem. Ett normrelaterat betygssystem som byggde på samma principer som intelligenstesterna övergavs. Det system som ersatte, och som fortfarande lever kvar, är ett s.k. mål- och kriterierelaterat betygssystem, vilket innebär att eleverna bedöms utifrån på förhand uppsatta kriterier i ämnet vilka anger vad eleven ska kunna inom ett aktuellt kunskapsområde. Motsvarande övergång till ett mål- och kriterierelaterat betygssystem har skett på flera håll i de västerländska skolsystemen.

Prov kan ha olika utformning och kan mäta olika typer av kunskaper. Prov med fasta svarsalternativ, s.k. flervalsfrågor, eller prov där svaren ges i form av ett ord eller ett numeriskt resultat, s.k. "kortsvarsfrågor", har fördelen att de jämfört med andra prov ger resultat som är lätta att jämföra och de går snabbt att rätta (Cunningham 1998). Nackdelen är att de inte ger någon egentlig kunskap om vad eleven kan utan bara om han eller hon svarat rätt eller fel. Detta är inget problem så länge man är intresserad av fakta- och procedurkunskaper men om man vill utvärdera mer komplexa kunskaper behövs andra uppgiftsformat (Cunningham 1998). Det finns studier (Gipps 1994) som för fram att flervalsfrågor är det uppgiftsformat som ger högst bedömaröverensstämmelse medan

---

<sup>1</sup> <http://www.sydsvenskan.se/lund/article250634/Lund-satter-hoga-betyg-i-nian---igen.html> 100812

bedömning av komplexa uppgifter ger sämre överensstämmelse. De styrdokument vars uppdrag lärare är satta att förvalta ställer krav på att elever ska utveckla komplexa kunskaper och förmågor som kan vara svåra att bedöma på den typ av prov där svaren kan bedömas som antingen rätt eller fel. Både i Sverige och i andra länder (Broadfoot 1995) har gjorts satsningar på att utveckla nationella provsystem som är mer inriktade på komplexa färdigheter och som har ett mer öppet svarsformat. Utmaningen består då i att utforma bedömningsanvisningar som gör att tillförlitligheten inte riskeras.

### ***Bedömningsmatriser***

För att bedöma komplexa uppgifter och kvaliteter i elevers kunskaper behövs tydliga bedömningsanvisningar. Ett exempel på en sådan bedömningsanvisning är en bedömningsmatris. Även om bedömningsmatriser kan vara utformade på olika sätt har de gemensamt att de är verktyg för kvalitativ bedömning av komplexa elevarbeten (Arter and McTighe 2001). Det som kännetecknar en bedömningsmatris är att de både består av ett antal kriterier/aspekter samt beskrivningar av dessa kriterier på olika kvalitativa nivåer. Genom att göra kriterier och nivåbeskrivningar explicita kan både elever och bedömare bli uppmärksamma på vad det är som ska bedömas. Om matrisen delas upp i olika aspekter som bedöms var för sig, t ex metodval, genomförande och redovisning i en matematikuppgift, får man en analytisk bedömningsmatris. Om en sådan uppdelning inte sker utan man bedömer kvaliteten i sin helhet blir matrisen istället holistisk (Arter och McTighe 2001).

Det finns studier som lyfter fram att reliabilitet och validitet vid bedömning av komplexa uppgifter med stöd av bedömningsmatriser ökar (Jönsson och Svingby 2007(review)). Även om de värde på överensstämmelse mellan olika bedömare generellt sett är lägre än vad många av forskarna anser är godtagbart (korrelation  $< 0,7$ ) drar de flesta ändå slutsatsen att de erhållna värdena är tillräckligt höga, åtminstone om bedömningen är "low stake". Reliabiliteten kan ökas om bedömningsmatrisen är analytisk, uppgiftsspecifik och om bedömarna får träning. Reliabiliteten ökar även om bedömarna har tillgång till bedömda elevarbeten (Jönsson och Svingby 2007). Om bedömningsmatriserna har för många kriterier och görs för detaljerade kan detta få negativ konsekvenser i form av minskad användbarhet. Eisner (1991) har även framfört kritik mot att använda för detaljerade bedömningsanvisningar eftersom han menar att dessa kan hämma kreativiteten hos elever och lärare.

### **Andra utgångspunkter**

I Australien har en forskargrupp studerat lärare (åk 4-9) som betygsatt elevarbeten utifrån beskrivningar av olika kunskapsnivåer (A-E), sk. "standars" (Wyatt-Smith et al. 2010). Efter lärarnas individuella bedömning diskuterade och enades en grupp lärare om ett slutgiltigt betyg. Studien visar att lärarna vid bedömning utifrån kriterier använder sig både av skriftliga artefakter, som kriteriebeskrivningar och bedömda elevlösningsexempel, men även "tyst kunskap" av olika slag, så som ämneskunskaper, kunskaper om kursplanen, tidigare erfarenheter, kunskap om enskilda elever och åsikter om vad "en genomsnittlig elev" i en given årskurs förväntas kunna och sociala processer, som dialog och förhandling. De drar slutsatsen att kriterier spelar en roll vid bedömningsprocessen men att de inte räcker till för att förstå hur lärare värderar och betygsätter elevarbetena. Även Sadler (1998) beskriver ett antal faktorer som utgör de källor lärare förlitar sig på när de bedömer. Dessa är goda ämneskunskaper, djupkunskap om kriterier och nivåer i förhållande till det som ska bedömas, bedömningsskicklighet eller erfarenhet av att ha bedömt liknande uppgifter tidigare samt en uppsättning attityder och förhållningssätt till undervisning och elever. Utgångspunkt har även tagits i Sadlers (1989) antagande kring kvalitativ bedömning och då särskilt i att kriterier vid

kvalitativ bedömning är *fuzzy*, dvs. oklara och otydliga, snarare än *sharp*, klara och precisa. Han menar att de bedöms enligt en glidande skala och är en abstrakt mental konstruktion som gör det svårt att uttrycka dessa exakt i ord och att förstå utanför kontexten. Det innebär även att förståelse av kriterier skapas över tid och att förståelse på detta sätt utvecklas hos praktiker. En annan utgångspunkt är antagandet att kriterier bör kompletteras med exempel på elevarbeten som illustrerar de kännetecknen som kriterierna representerar.

Vidare utgör Säljö (2000) beskrivning av kunskap som knuten till argumentation och handling i en social kontext till följd av att man försöker se, förstå och hantera omvärlden på ett bestämt sätt ytterligare en utgångspunkt. Han menar att kunskap inte är en objektiv sanning utan att den fås genom att var och en intar ett visst perspektiv till den verklighet som studeras och beroende på vilket perspektiv man intar så framträder olika saker. Han kritiserar ett förhållningssätt som innebär att kunskap är direkt överförbart från en person till en annan och som förmedlas, mottas och lagras för att sedan kunna tillämpas. Säljö (2000) beskriver lärande som ett samspel mellan människa och omvärld vilket sker genom socialt och kulturellt betingade artefakter. Artefakter kan vara både fysiska och intellektuella och de artefakter som finns att tillgå i ett visst sammanhang möjliggör olika sätt att handla. Genom att använda artefakter medieras (överförs) ”verkligheten” i olika praktiker (Säljö 2000) och de har därmed ingen inneboende mening utan denna skapas i interaktionen mellan individerna och artefakterna i en specifik situation, kontext. En bedömningsmatris kan i detta perspektiv ses som en intellektuell artefakt där kunskaper om kriterier och hur bedömning går till har ”byggts in”.

Ytterligare utgångspunkt har varit att se bedömning med bedömningsmatriser som instructed actions (Garfinkel 2002), vilket kan beskrivas som ansvarskrävande handlingar. Det innebär att människan inte ses som en slav under regler och för att utföra en handling utifrån en instruktion krävs förståelse för det sammanhang där handlingen utförs. Goodwin (1994, 2000) menar att utvecklandet av ett kompetent/professionellt seende består i att deltagare tillsammans försöker förstå en praktik genom samtal och diskussion. Bedömning med bedömningsmatris/instruktion ses här som ett exempel på en sådan praktik. Instruktionen skapas då genom diskussioner där exempel både hämtas från den aktuella situationen och från vidare resonemang, i vilka den aktuella situationen är en del. Instruktionen blir därmed det som syns i handlingen snarare än det som står i instruktionen. Genom att noga studera det som blir synligt vid sådan interaktion i förhållande till den specifika situation där den utförs kan man få en känsla för hur lärare skapar mening kring sin uppgift att bedöma.

Utifrån dessa utgångspunkter har lärares bedömning med bedömningsmatriser studerats. En handling, bedömning med bedömningsmatris, studeras således bäst i den kontext där den utförs och genom hur den blir synlig i handling. Det finns antaganden att uppsatta kriterier inte nödvändigtvis är så styrande vid bedömning som det görs anspråk på i flera sammanhang och att instruktionen tolkas i den lokala kontexten. Detta väcker frågor kring hur lärare skapar mening kring sin uppgift med att bedöma med stöd av bedömningsmatriser, vilka överväganden de gör samt vilka andra resurser lärare använder sig av vid bedömning. Då hanteringen av instruktioner vid bedömning kan ses som en del i utvecklandet av en gemensamt förhandlad professionell kompetens studeras detta bäst genom hur instruktionen blir synlig i handling.

### **Studiens uppläggning och datainsamling**

I denna studie används en del av det nationella ämnesprovet för åk 9 i matematik från 2009. Det är sex lärare som deltar i studien. Dessa arbetar på tre olika skolor i samma kommun och

är samtliga lärarutbildade och behöriga att undervisa i matematik i åk 9. Lärarna har videofilmats när de tillsammans två och två bedömer en uppgift, kallad B2, utifrån en uppgiftsspecifik bedömningsmatris med tillhörande bedömda autentiska elevarbeten. Uppgiften som studeras ”prövar elevens förmåga att lösa problem, reflektera över och tolka sina resultat samt bedöma deras rimlighet” samt ”elevens förmåga att uttrycka sina tankar skriftligt, dra slutsatser och generalisera” (Skolverket 2009). Lärarna bedömer mellan tolv och femton elevlösningar per par och varje film är mellan 35 och 60 minuter långa. Det insamlade materialet består av tre filmer som är inspelade i autentiska bedömersituationer i samband med att provet genomförts i maj 2009. På filmerna kan man se lärarna som bedömer, elevlösningarna och matrisen vilket har gjort det möjligt att vid analysen gå tillbaka och se hur lärarna gjort markeringar och hur de pekat i materialet. Elevlösningarna är hämtade från de i studien medverkande lärarnas egna undervisningsgrupper vilket innebär eleverna även undervisats av dessa. Det insamlade materialet består dessutom av kopior av elevlösningarna och en ifylld bedömningsmatris för varje elev på vilken lärarna markerat sin bedömning. De inspelade filmerna har sedan transkriberats och analyserats med stöd av datorprogrammet Transana. Vid tillfället för studiens genomförande undervisade jag själv i en åk 9 och genomförde då det aktuella ämnesprovet.

### *Nationella ämnesprov*

De nationella ämnesprovets syfte är att stödja lärares bedömning av hur väl eleverna uppnått styrdokumentens mål, ge stöd för betygssättning samt bidra till likvärdig bedömning över landet (Skolverket 2007). Provet prövar inte alla kursplanens mål och för att möjliggöra bedömning av elevers matematikkunskaper utifrån målen och betygskriterierna behövs ett brett underlag (Skolverket 2007). Ämnesproven i matematik i åk 9 är därför fördelat på tre delprov som skiljer sig åt beträffande kunskapsinnehåll, arbetssätt, redovisnings- och bedömningssätt. Delprovets olika uppgifter avser att mäta elevernas kunskaper i förhållande till de i kursplanen uppsatta målen för åk 9 samt mot betygskriterierna. För att tydliggöra de kvalitativa nivåerna i mål och betygskriterier ges olika poäng vid bedömningen. G-poäng som syftar till kunskaper som kopplas till mål att uppnå och VG-poäng som syftar till betygskriterierna för VG och/eller MVG. Några uppgifter ger utrymme att visa kvalitet i lösningarna som hänförs till kriterierna för MVG. Det handlar om uppgifter som kan lösas på olika sätt, med generella metoder eller med ett mer utvecklat matematiskt språk. Dessa uppgifter är markerade med symbolen  $\infty$ . Delproven ges vid olika tillfällen under vårterminen och när alla delprov är utförda summeras antal G- och VG-poäng samt antal  $\infty$  och det resulterar i ett provbetyg.

### *Uppgift och matris*

Uppgiften som ska lösas innebär att två olika modeller, modell A och B, för katters åldrande ska jämföras. Modell A innebär att katten åldras 7 kattår för varje människoår. Modell B innebär att kattens ålder ökar med 15 kattår under det första levnadsåret och med 10 kattår under andra levnadsåret. Därefter åldras katten 4 kattår för varje människoår. Den inledande deluppgiften (a) går ut på att bestämma åldern hos en kattunge som idag är tre år enligt de båda modellerna. I nästa deluppgift (b) ska en tabell fyllas i där man bestämmer kattens ålder enligt båda modellerna för ett antal år och sedan ska modellerna prickas in som två grafer i ett koordinatsystem. I den tredje deluppgiften (c) ska eleven så exakt som möjligt bestämma den tidpunkt då de båda modellerna ger samma ålder och slutligen i den fjärde deluppgiften (d) ska de båda modellernas rimlighet jämföras och eleven ska motivera vilken modell som anses mest rimlig. Den inledande deluppgiften är förhållandevis enkel och kan resultera i svar som antingen är rätt eller fel medan de följande deluppgifterna blir allt mer komplexa.



Uppgiften bedöms med utgångspunkt i en analytisk, uppgiftsspecifik bedömningstris som består av tre kriterier/kunskapsaspekter, *Förståelse och metod*, *Genomförande och analys* samt *Redovisning och matematiskt språk*. För varje kriterie finns beskrivningar på tre olika kvalitativa nivåer. Eftersom det nationella ämnesprovet är ett prov som används i summativt syfte där resultaten på de olika delproven ska räknas samman till en totalpoäng finns det för varje kvalitativ nivå angivet det antal G- och/eller VG-poäng som lösningen kan resultera i. På den högsta nivån finns även angivet om elevlösningarna kan uppvisa  $\alpha$ , MVG-kvalitet. Totalt kan uppgiften resultera i fyra G-poäng, fyra VG-poäng samt fyra olika MVG-kvaliteter. Som stöd, utöver bedömningstrisen, har lärarna även sju bedömda autentiska elevarbeten av olika kvalitet.

## Resultat och analys

De inledande deluppgifterna där eleverna ska bestämma kattens nuvarande ålder enligt de två modellerna samt fylla i en tabell kan inte ses som komplexa utan de svar som krävs kan vara antingen rätt eller fel. Utifrån den ena modellen är katten tjuogoett år och utifrån den andra tjugonio. När lärarna ska bedöma dessa uppgifter blir det inga diskussioner och de kan i stort sett pricka av i trisen om eleven löst uppgiften eller inte. I de fall eleverna har lyckats att lösa uppgifterna på ett lämpligt sätt kan lärarna konstatera att så är fallet, markera detta i trisen och gå vidare. I trisen uttrycks att *”eleven ska visa förståelse för problemet genom att tolka modellerna och bestämma kattens ålder korrekt enligt båda modellerna för några år”*. Har det gjort detta får de en g-poäng.

- L1: mmmm. Ska vi se här. tjuogoett och...  
 L2: ...tjuogoett och tjugonio ja  
 L1: Tabellen ifylld...  
 L2: i förlängd version  
 L1: Precis. Sen ritar han diagram här. Där dock skärningspunkten hamnar utanför y-axelns längd  
 L2: mmm  
 L1: Dock har han ju fortsatt göra det rätt även när diagrammet tar slut  
 L2: Ja  
 L1: han har gått upp...ja Fortsätt (vänder blad)  
 L2: Ja. Helt riktigt  
 L1: Fem komma sju år enligt koordinatsystem uppskattningsvis. Då är han ju där då (pekar i trisen). Han visar..  
 L2: Han har hittat skärningspunkten  
 L1: ...skärningspunkten. ett ett

I dialogen ovan kan man se att lärare 2 faller lärare 1 i talet och fortsätter meningen där denne slutade, vilket kan tolkas som att det finns ett rätt svar och att detta inte behöver förhandlas. I den fortsatta diskussionen har eleven även, utifrån sina tabellvärden, ritat ett koordinatsystem med två grafer så att denne kunnat identifiera skärningspunkten mellan dessa båda och därmed angett ett rimligt värde för när de ger samma ålder. Här menar jag att lärarna har lätt att tolka elevlösningen i förhållande till instruktionen eftersom kriterierna är *”sharp”* (Sadler, 1989). Formuleringen i trisen uttrycker att eleven ska *visa ”förståelse”*, vilket kan ses som *”fuzzy”*, men är sedan kompletterad med ett förtydligande om hur eleven ska göra det, *”genom att ... bestämma kattens ålder korrekt...”*. Av detta uppfattar jag att det lärarna gör här inte kan ses som kvalitativ bedömning.

I nästa deluppgift ska eleven göra ett koordinatsystem samt rita två grafer, en för varje modell för kattens åldrande. I trisen uttrycks att *”punkterna för minst en modell ska vara korrekt inprickade i ett acceptabelt koordinatsystem”*. Det finns matematiska konventioner som anger hur ett korrekt ritat koordinatsystem ska se ut och när eleverna håller sig inom dessa blir det heller ingen diskussion mellan lärarna och de behövde inte heller vända sig till de bedömda

elevarbetena. Flera koordinatsystem blev däremot föremål för överväganden kring huruvida de var acceptabla eller inte. Det kan t.ex. vara att eleverna har vänt på x- och y-axel, graderat axlarna på olämpligt sätt, prickat in något värde fel, prickat in värdena rätt men utifrån ett felaktigt beräknat värde, dragit linjer mellan punkterna på ett felaktigt sätt eller utelämnat pilar vid axlarnas slut. De bedömda autentiska elevarbeten som finns är tänkta att visa elevarbeten på olika kvalitativa nivåer (Skolverket, 2010) och är sju till antalet. Endast ett av elevarbetena visar exempel på koordinatsystem som bedömts inte vara acceptabelt och i det fallet har eleven inte ritat något koordinatsystem alls. I dessa situationer tvingas lärarna att själva tolka vad som kan anses vara ett acceptabelt koordinatsystem och vad som inte är det. De resonemang de lutar sig mot när de gör sina avgörande bedömningar kommer då ibland att handla om sådant som ligger utanför matrisens formuleringar, vilket ger en antydning om att instruktionen inte ger dem stöd att bortse från sådana faktorer. Denna brist på stöd får till följd att lärarna får svårt att vara konsekventa och använder det frirum för tolkning som skapas till att lyfta in andra faktorer i bedömningen. En av de faktorer som lyfts in är den kunskap lärarna har sedan tidigare om eleven. Följande exempel är hämtat från en situation där lärarna uttrycker att eleven de bedömer är svag. Eleven har inte tagit hänsyn till att det i deluppgiften stod att de båda graferna skulle ritas in i ett koordinatsystem utan har gjort två stycken. Eleven har dessutom vänt på x- och y-axel vilket får lärarna att överväga om koordinatsystemet är acceptabelt eller ej.

L3: Ehhh "Punkterna för minst en modell korrekt inprickade i ett acceptabelt koordinatsystem".  
 L4: mm Men om du om det är en elev som är svag här...  
 L3: Ja det är det  
 L4: ...som du menar, så tycker jag att de e ju bara detta att han har fått ihop det här med x och y  
 L3: ja  
 L4: han har skrivit. Och att där är pilar har han inte glömt bort. Men sen har han vänt  
 L3: Han har vänt på dom ja  
 L4: ja och det kan man...det är väl ingen som... lag...  
 L3: nej  
 L4: ...som säger att det ska va det ena eller det andra  
 L3: Acceptabelt koordinatsystem...  
 L4: Ja, tycker jag

Lärarna ger uttryck för att eleven är svag, men ändå har tagit till sig något av undervisningen och det resulterar i att eleven får ett g-poäng trots att de är tveksamma till det. Det fortsatta resonemanget kring att eleven vänt på axlarna kan även tyda på att de inte följer eller är medvetna om de instruktioner som fanns till uppgiften där det tydligt angavs vad som skulle vara på x- respektive y-axel. När lärarna vänder sig till de bedömda elevlösningarna så finner de inga exempel som ger dem stöd i hur de ska hantera lösningar där eleverna vänt på axlarna, vilket möjliggör att de själva tolkar om det är acceptabelt eller inte. I den diskussion som sedan följer i samband med att lärarna ska bedöma elevens redovisning och matematiska språk framgår det att de även här tar fasta på att det är en svag elev de bedömer:

L3: Han har inget matematiskt språk över huvud taget.  
 L4: Nä alltså det som han inte...  
 L3: Alltså det känns ju inte som han e där och ska ha en poäng till (pekar i matrisen). Han har ju...  
 L4: Samtidigt som jag tycker att e detta en svag elev...  
 L3: mmm  
 L4: som har lyckats ändå så här...  
 L3: Ja det har han  
 L4: ...få till det så e det...  
 L3: Så att han har ju ändå...  
 L4: Skriv ett frågetecken.  
 L3: Det är om han ska ha den (1 g-poäng)  
 L4: mmm men annars blir det  
 L3: Redovisningen omfattar hela uppgiften. Det gör den ju i stort sett

L4: mm det är ju det där att han inte har kopplat ihop dom två, men det var det ju ingenting som stod heller utan det var ju bara en slutledning...  
 L3: mmm  
 L4: ...som skulle göra det  
 L3: att man skulle ha dom i samma ja  
 L4: Ja det tycker jag.  
 L3: ja men då sätter vi där  
 L4: mmm. Men då blir det i så fall 4 nej 3  
 L3: en två tre ja. Ja.

För att få ett g-poäng under kriteriet *Redovisning och matematiskt språk* ska redovisningen enligt bedömningsmatrisen vara "möjlig att följa" och det matematiska språket ska vara "acceptabelt". Dessutom ska redovisningen omfatta samtliga deluppgifter. Inledningsvis anser lärarna inte att eleven har något matematiskt språk och att denne därför inte ska få poängen. De uttrycker att "det känns" inte som att han ska ha denna poäng. Längre fram lyfter de fram att eleven ritat graferna i två koordinatsystem som en anledning att inte utdela poänget. De konstaterar att det egentligen inte stod i uppgiften, vilket det faktiskt gjorde. De anser att redovisningen omfattar hela uppgiften och enas om att eleven ska få poänget och lyfter åter in att det är en svag elev som lyckats att få till något som liknar ett koordinatsystem. Eleven får slutligen tre g-poäng på hela uppgiften, varav två av dem varit föremål för överväganden och där matrisen och de bedömda elevlösningarna inte givit lärarna det stöd de kanske hade behövt för att göra en tolkning som hjälpt dem att se förbi det faktum att de visste hur eleven presterade i vanliga fall. Båda poängen var lärarna tveksamma till att ge eleven utifrån att de uttryckte brister i lösningen. I båda fallen lyfter de in resonemang om att det är en svag elev och utgår inte från lösningens egentliga förtjänster. De resonerar fram och tillbaka och bedömningsmatrisens vaga formuleringar ger stöd för att utdela poäng och de vänder sig till de bedömda elevlösningarna. Jag upplever att lärarnas resonemang liknar de som beskrivs av både Sadler (1998) och Wyatt-Smith et al. (2010) att lärare vid bedömning utifrån kriterier tar in andra källor i bedömningen. När lärarna vid senare tillfällen ska bedöma liknande elevlösningar där eleverna har ritat båda graferna i samma koordinatsystem får inte dessa några poäng för ett acceptabelt koordinatsystem. Detta sker i samförstånd och utan några övriga kommentarer vilket kan tolkas som att lärarna egentligen är överens om att ett koordinatsystem där eleven tvärtemot instruktionen vänder på x- och y-axel inte är att se som acceptabelt. En av dessa elever har även tecknat formler för de båda modellerna och satt dem lika med varandra samt löst ut tidpunkten för när de båda modellerna ger samma ålder. Enligt kriteriebeskrivningarna i bedömningsmatrisen visar detta förfarande vid beräkning av tidpunkten på högre kvalitet i lösningen. Den ena formeln är däremot felaktig så det svar som faller ut är inte korrekt trots att eleven löst den ekvation som tecknats på ett korrekt sätt. Denna elev får därmed inte heller poäng för redovisning och matematiskt språk och missar följaktligen de två poäng som förra eleven fick. Det framgår inte av samtalet om den senare eleven är svagpresterande eller inte.

Sadler (1989) menar att professionell bedömning inte enkelt kan avgränsas och skrivas in i kriterier och nivåer och att denna därför antagligen kommer att förbli mottaglig för andra kunskapskällor. Jönsson & Svingby (2007) konstaterar att validiteten vid bedömning med bedömningsmatris inte självklart blir högre och jag menar att detta exempel stödjer det antagandet. Trots att matrisen är kompletterad med exempel på bedömda elevlösningar räcker inte dessa till för att ge lärarna stöd att bortse från vilken elev det är de bedömer. När instruktionerna inte lyckas ge lärarna stöd att avgöra vad som är ett acceptabelt koordinatsystem och vad som inte är det tvingas lärarna att konstruera egna gränser för detta, vilket kan vara en svår uppgift. Att så är fallet kan illustreras med följande korta utdrag:

L1: och det är ju alltså det är illa att du inte ser att den är proportionell här ju. Den ser ju ut som där går den så och sen går den brantare upp mellan två och tre. Ehhh det är snällt att säga att det är acceptabelt men det kanske är tufft att säga att det inte är det.

L2: Det får väl va acceptabelt men precis

När lärarna ska bedöma elevlösningarna vänder de sig inte bara till matrisen och de bedömda elevlösningarna för att få stöd. De vänder sig även till sig själva och sina egna erfarenheter av bedömning och vad de tycker en elev borde kunna. I en situation noterar en av lärarna att skärningspunkten mellan de båda graferna hamnar utanför y-axeln. Eleven har löst uppgiften på ett sätt som gett denne alla poäng uppgiften kan ge, 4 g-poäng och 4 vg-poäng, samt även två  $\pi$ . Eleven har satt upp formler för de båda modellerna, satt dessa lika med varandra och löst ut den tidpunkt när de båda modellerna ger samma ålder. Eleven får till ca 5,67 som sedan avrundas till 5,7. Lärarna diskuterar om eleven ska få ytterligare en  $\pi$  för sin redovisning och sitt matematiska språk, men stannar upp vid att skärningspunkten hamnat utanför koordinatsystemet samt vid hur eleven hanterat närmevärdet 5,67. Den totala poäng en elev kan få för redovisning och matematiskt språk är 1 g- och 1 vg-poäng samt en  $\pi$ .

L1: mm Ehhh redovisning tycker vi att den är klar och tydlig, matematiska språket lämpligt, båda graferna är acceptabelt ritade i ett koordinatsystem med lämplig skala.

L2: Och då har han ju valt... Det han inte gör så bra det är att han kommer utanför y-axeln

L1: mmm

L2: Ska vi anse att han bör få ett då?

L1: Ahhh

L2: Eller ska han få

L1: tycker väl...

L2: eller han ska ha sin mvg-poäng trots att den ser ut så

L1: njaaa. Kanske kan känna att ett det känns bra. Det är inte helt. Jag tycker han kunde han kunde varit mer klinisk här också

L2: Formulerat som att det skulle varit fem år och

L1: Ja, att han konstaterar att detta är ju faktiskt svaret

L2: mm

L1: Nu menar han ju att ja fem komma sju var ju en bra gissning

L2: mmmm

L1: han borde tatt det värdet och sett att detta är det exakt korrekta och liksom svarat så att det är det korrekta. Han löser ju ut att det är det värdet men, ja

L2: det spelar ingen roll att han valt fem komma sju istället för fem åroch åtta månader

L1: Ja, det tänkte jag också men jag tror att här är exempel på bägge delar (bläddrar bland de bedömda elevlösningarna)Bläddrar fram fullpoängsidan för en gångs skull

L2: Det avgör inte hur många poäng eller mvg... om han ska ha, tror du det?

L1: Nä. Ja det kan det ju va i och för sig. Där är en som säger fem år och åtta månader. Och då har man ju dom stjärnorna också. Ehhh

L2: kommer här (hänvisar till nästa sida i mallen) Där stjärna på alla.

L1: Han svarar så... Fem och två tredjedels år. Den är ju i och för sig rätt lika

L2: Det är lika det..

L1: Det är bara att han har decimaler där men det kan ju kvitta.

L2: Det kan kvitta ja

L1: Det är väl just i så fall jag kan tycka att han går utanför sin egen...

L2: mm

L1: ...sitt diagram där om man nu ska.....Men det kanske är gnälligt?

L2: Är det gnälligt?

L1: Ahh, jag tycker det är en ett etta

L2: ett etta

L1: Då får han fyra fyra och två stjärnor

Diskussionen resulterar i att lärarna inte ger eleven ytterligare en  $\pi$  för sin redovisning. De ger uttryck för att det finns något i elevens redovisning som inte stämmer överens med deras bild av hur en "klinisk" redovisning ska se ut. Det är två saker de hänger upp sig på, hanteringen av närmevärdet och skärningspunkten som hamnat utanför koordinatsystemet. Formuleringen i matrisen för att få ett g- och ett vg-poäng lyder: "Redovisningen är klar och tydlig. Det matematiska språket är lämpligt och båda graferna är acceptabelt ritade i ett koordinatsystem

med lämplig skala.”. För att dessutom få en  $\alpha$  ska redovisningen vara ”mycket god”. När lärarna inte får stöd i matrisen för att avgöra om eleven ska ha  $\alpha$  eller inte vänder de sig till de bedömda elevarbetena men inte heller där hittar de det stöd de behöver för att enkelt avgöra detta. I elevarbetena finns två exempel där närmevärdet hanterats på ett jämförbart sätt som i deras elevlösning och de väljer att acceptera elevens svar som likvärdigt. Det som kvarstår är skärningspunkten som hamnat utanför y-axeln och det är denna som faller avgörandet till elevens nackdel med motiveringar som ”Det känns bra” och ”jag tycker att det är en etta”. Vad denna känsla grundas i går inte att svara på med det underlag som finns här men jag tolkar det som att det rör sig om tidigare erfarenheter lärarna har av att bedöma liknande uppgifter samt deras uppfattningar av vad som är ett korrekt matematiskt språk, vilket stöds av Sadler (1989) och Wyatt-Smith et al. (2010). Det som avgör blir en detalj som i matematiskt avseende inte självklart kan anses som felaktigt. Däremot väljer de att godta att eleven hanterat närmevärdet på ett sätt som möjligtvis kunde ha uttryckts mer korrekt. I detta beslut tar de stöd av de bedömda elevarbetena. Konsekvensen för eleven blir en utebliven  $\alpha$  som kan spela en avgörande roll för denne i sammanräkningen av det totala provbetyget där det krävs minst 7 sådana för att få provbetyget MVG. Lärarna har återigen hamnat i en situation där vaga formuleringar, ”lämplig” och ”acceptabel”, ger utrymme för egna tolkningar och där de instruktioner de har till sitt förfogande ger begränsat stöd i bedömningen. Sadler (1987) uttrycker att det är osannolikt att instruktioner tillsammans med exempel helt och på ett enkelt sätt skulle kunna kompensera för den tysta kunskap, grundad i människors värderingar, eftersom externa formuleringar inte kan bli så uttömmande och täcka in alla olika möjligheter. Han menar att det inte ens skulle vara önskvärt. Mot bakgrund av Goodwin (1994, 2000) tolkar jag det som att det är i interaktionen mellan lärarna, vid den kvalitativa bedömningen som möjliggörs i friutrymmet som professionaliteten skapas och utvecklas.

Vid analysen av materialet blev det även synligt att lärarna ibland hamnade i situationer när de förhandlade om poäng mellan de olika aspekterna i bedömningsmatrisen. I exemplet nedan sker detta när lärarna uttrycker att eleven varken redovisat uppgiften på ett tillfredställande sätt eller ritat ett acceptabelt koordinatsystem. I bedömningsmatrisen uttrycks att elevlösningen inte ska ge några poäng om ”Redovisningen omfattar endast delar av problemet och det matematiska språket kan vara knapphändigt”. Ett g-poäng ska ges när ”Redovisningen är möjlig att följa och det matematiska språket är acceptabelt.” och ”Redovisningen omfattar hela problemet”. Formuleringarna öppnar återigen för lärarna att själva tolka var gränsen går mellan ett ”knapphändigt” och ett ”acceptabelt” koordinatsystem. I de bedömda elevlösningarna finns två exempel som kan ge lärarna vägledning kring hur de kan resonera för att avgöra detta. Det ena exemplet visar en lösning som endast omfattar två av fyra deluppgifter och i det andra har eleven i stort sett endast skrivit svaren och inte redovisat hur denne kommit fram till dessa samt utelämnat koordinatsystemet.

L1: Om man sätter två noll där men... (på Genomförande och analys)

L2: håller henne på noll

L1: ..håller henne på noll noll kanske på redovisning

L2: mmm Det saknas ju där det är ju inte så himla bra

L1: Nej

L2: och den är ju helt fel ju

L1: Ja. Ahh det tycker jag. Två noll, noll noll. Då hamnar hon på tre ett.

I detta fall har lärarna tidigare diskuterat att elevens koordinatsystem är slarvigt ritat eftersom linjal inte använts vilket har medfört att elevens tolkning av när de båda modellerna ger samma ålder inte blivit helt korrekt. De anser dessutom att elevens gradering av x- och y-axel är felaktig och är därför tveksamma till att ge eleven de två möjliga g-poängen för

genomförande och analys eftersom de uttrycker att ”*det är bra betalt*”. I övrigt har eleven redovisat hela uppgiften och dragit slutsatser. Här fattar de beslutet att trots allt ge eleven två g-poängen men att inte ge denne några poäng för redovisning och det matematiskt språk, trots att detta är fullt likvärdigt med flera andra elevlösningar som fått poänget. För elevens totalsumma spelar det ingen roll om denne får poäng för det ena eller för det andra men bedömningsmatrixens användning som bedömningsverktyg vid kvalitativ bedömning undergrävs.

## Diskussion

Resultatet av studien visar att kriterierna i bedömningsmatrixen är formulerade på ett sätt som bidrar till att det skapas ett friutrymme för lärarna att själva göra tolkningar av dem. Det leder till att lärarna tvingas att göra överväganden som får till följd att bedömningen inte blir konsekvent och att lärarna tar in andra resurser i bedömningen. Förutom de skriftliga instruktionerna (bedömningsmatrix och bedömda elevarbeten) använder även lärarna sk. tyst kunskap, outtalade kriterier, som grundar sig i lärarnas egna uppfattningar och erfarenheter av att bedöma vad som är god kvalitet i ett elevarbete. Även vetskapen om hur eleverna brukar prestera i vanliga fall får konsekvenser för bedömningen.

Att bedömningsmatrixen innehåller uttryck som ”acceptabelt”, ”rimligt” och ”lämpligt” gör att det skapas ett friutrymme för lärarna att tolka vad som ska räknas som t.ex. ett acceptabelt koordinatsystem och vad som inte räknas som acceptabelt. I detta friutrymme skapas möjlighet för lärarna att utveckla och använda sin professionalitet då inte heller de bedömda elevarbetena, på olika kvalitativa nivåer, exemplifierar alla de olika varianter som förekommer bland de elevlösningar lärarna ska bedöma. Sadler (1989) för fram att kvalitativ bedömning kännetecknas av att de kriterier som bedöms inte enkelt går att definiera och exemplifiera och att förståelse för dem utvecklas i den lokala kontexten, vilket är jag menar är vad lärarna i denna studie gör genom att gång på gång tvingas jämföra och ta ställning till om elevernas lösningar ska anses rimliga, acceptabla, lämpliga etc. eller ej. Goodwin (1994, 2000) uttrycker att instruktionen skapas i samspel och diskussion och att den blir det som blir synligt i handling snarare än det som står uttryckt i själva instruktionen. Han menar även att det är just i själva samspelet när aktörerna tolkar och försöker förstå instruktionen som det professionella seendet utvecklas. Med det sättet att förhålla sig förstår jag det som att lärarna i denna studie är mitt uppe i en process där de samtidigt som de bedömer även utvecklar sin bedömarkompetens i att kunna bedöma komplexa kunskaper hos elever. Vikten av att utveckla professionell bedömarkompetens lyfts även fram i andra studier. Rezaei et al. (2010) hävdar att användandet av bedömningsmatrixer inte nödvändigtvis ökar tillförlitligheten i bedömningen om inte bedömare är väldigt kompetenta, både i att använda och att designa bedömningsmatrixer, vilket de blir genom att använda dem. Säljö (2000) menar att kunskap inte kan överföras direkt utan kunskapsutveckling sker i samspelet mellan människan och omvärlden genom socialt och kulturellt betingade artefakter. Jag tolkar det som att bedömningsmatrixen är en artefakt som medierar handlingen att bedöma mellan konstruktören och brukaren, men då kunskapen inte kan lagras och direkt överföras (Säljö 2000) måste denna kunskap ”bli till” i samspelet mellan bedömare, elevlösningarna, bedömningsmatrixen och de bedömda elevarbetena. Vilka artefakter som är tillgängliga möjliggör olika handlingssätt (Säljö 2000) vilket innebär att med andra bedömningsanvisningar hade utfallet av bedömningen kunnat bli ett annat. Det är alltså den lokala kontexten och de artefakter som finns att tillgå som blir avgörande. Bedömningsmatrixen blir alltså ett verktyg som hjälper lärarna att se men vad de ser avgörs av den lokala kontexten. Tidigare genomförda studier (Jönsson och Svingby 2007) lyfter fram att bedömningsmatrixer ger ökad reliabilitet vid

bedömning om dessa är analytiska, uppgiftsspecifika och kompletterade med elevarbeten på olika kvalitativa nivåer. Bedömningsmatrisen som används i denna studie stämmer in på denna beskrivning vilket gör det rimligt att anta att om lärarna skulle ha bedömt dessa elevlösningar *utan* stöd av en bedömningsmatris hade utfallet blivit mindre tillförlitligt.

En normalisering av elever som starka och svaga och att sk. tyst kunskap lyfts in i bedömningen kan innebära en risk eftersom elevers lösningar då inte blir bedömda utifrån sina förtjänster utan bedöms på andra grunder vilket gör att bedömningen inte blir valid. Sadler (1985) menar att kvalitativ bedömning inte blir trovärdig endast genom att man förlitar sig på kriterier eftersom innebörden av dessa inte lätt kan formuleras. Därför kommer kvalitativ bedömning antagligen att förbli mottaglig för påverkan av annan kunskap och andra kompetenser vilket blir synligt i denna studie. I denna studie visar det sig att det kan vara en fördel för eleven att lärare "ser" och belönar förtjänster i lösningen som inte enkelt låter sig beskrivas, motiveras eller uttryckas men även en nackdel när de väljer att inte ge poäng då lösningarna inte når upp till den standard lärarna "vet med sig" men inte heller kan uttrycka, just eftersom det är svårt att formulera. Även resultatet av den omrättning som genomfördes under 2009 av ett urval av de nationella ämnesproven i grundskolan och gymnasieskolan (Skolinspektionen, 2010) visar på att det är stora skillnader mellan hur olika lärare bedömer. De visade sig att det fanns omfattande avvikelser mellan de ursprungliga bedömningarna och ombedömningarna i uppgifter där eleverna gavs möjlighet att uttrycka sig fritt och inte med enstaka ord eller numeriska värden. Den uppgift som ligger till grund för denna studie omfattades av ombedömningen och det visade sig att det fanns stora avvikelser och att dessa både var positiva och negativa, även om de till övervägande del var negativa. Även tidigare ombedömningar har gett liknande resultat (Olofsson 2006; Skolverket 2009). Detta väcker frågor kring förhållandet mellan sk. tyst kunskap, som lärts genom erfarenhet och som inte lätt kan uttryckas och den medvetna, explicita, kunskapen som kan uttryckas och kommuniceras på ett klart och entydigt sätt (den som finns uttryckt i matris och elevarbeten) (O'Donovan, Price och Rust 2004). O'Donovan et al. (2004) menar att utmaningen i att göra instruktioner transparenta ligger i att förstå att förhållandet mellan den tysta kunskapen och den explicita och att det i all meningsfull kunskap finns en dimension av tyst kunskap. Sadler (1987) anser att sådan tyst kunskap kan kommuniceras och spridas genom att lärare gör gemensamma bedömningar och att det kan bidra till mer tillförlitlig kvalitativ bedömning med tiden. Utifrån detta resonemang drar jag slutsatsen att den tysta kunskapen som lärare bär med sig, grundad i tidigare erfarenheter, är svår att komma ifrån vid kvalitativ bedömning, även om bedömningen sker med bedömningsmatris. Det blir då viktigt att vara medveten om förekomsten av denna och att ta fasta på att tillförlitligheten kan öka när vi gör bedömningen till en social process.

Denna studie har gett en del insikt i de svårigheter som är förbundna med att utföra kvalitativ bedömning av komplexa uppgifter. En bedömningsmatris löser inte av sig själv dessa svårigheter och det verkar som att de dilemman som uppstår vid bedömning av komplexa uppgifter även uppstår när man matrisbedömer. Studien ger dock en anledning till närmare studera hur användandet av bedömningsmatriser i social interaktion kan påverka kvaliteten på den lokala bedömningspraktiken.

## Referenser

- Arter, J, och Jay McTighe. 2001. *Scoring Rubrics in the Classroom*. Thousand Oaks, CA: Corwin Press.
- Broadfoot, P. 1995. Performance assessment in perspective: international trends and current English experience. I Torrance (red.) *Evaluating Authentic Assessment*. Buckingham, Philadelphia: Open University Press.
- Broadfoot, P. 1996. *Education, assessment and society – a sociological analysis*. Buckingham: Open University Press.
- Cunningham, G. 1998. *Assessment in the Classroom, Constructing and Interpreting Tests*. London och Washington D.C.: The Falmer Press.
- Eisner, E. 1991. Taking a Second Look: Educational Connoisseurship Revisited. I M. McLaughlin & D. Phillips (red.), *Evaluation and education at quarter century* (s. 169-187). Chicago: The National Society for the Study of Education.
- Garfinkel, H. 2002. *Ethnomethodology's program: Working out Durkheim's aphorism*. Lanham, MD: Rowman & Littlefield.
- Gipps, C. 1994. *Beyond Testing – Towards a Theory of Educational Assessment*. London: The Falmer Press.
- Gipps, C. 2001. Sociocultural aspects of assessment. I G. Svingby & S. Svingby (Red.), *Bedömning av kunskap och kompetens*. Stockholm: Lärarhögskolan i Stockholm, PRIM-gruppen.
- Goodwin, C. 1994. Professional vision. *American Anthropologist* 96, no.3: 606-633.
- Goodwin, C. 2000. Action and embodiment within situated human interaction. *Journal of Pragmatics* 32: 1489-1522.
- Jönsson, A. och G. Svingby. 2007. The use of scoring rubrics: Reliability, validity and educationan consequences. *Educational Research Review* 2: 130–144.
- Kjellström, K. 2000. Bedömningsmatris. *Nämnnaren*, nr 1: 45-51.
- Korp, H. 2003. *Kunskapsbedömning – hur, vad och varför*. Stockholm: Myndigheten för skolutveckling.
- O'Donovan, B., M. Price, och C. Rust. 2004. Know what I mean? Enhancing student understanding of assessment standards and criteria. *Teaching in Higher Education* 9, no. 3: 325-334.
- Olofsson, G. 2006. *Likvärdig bedömning? En studie av lärares bedömning av elevarbeten på ett nationellt prov i matematik kurs A*. Stockholm: Lärarhögskolan i Stockholm: PRIM-gruppen.
- Rezaei, A. R, och M. Lovorn. 2010. Reliability and validity of rubrics for assessment through writing. *Assessing Writing* 15: 18–39.
- Sadler, D.R. 1985. The origins and functions of evaluative criteria. *Educational Theory* 35, no. 3: 285–97.
- Sadler, D.R. 1987. Specifying and promulgating achievement standards. *Oxford Review of Education* 13, no. 2: 191–209.
- Sadler, D.R. 1989. Formative assessment and the design of instructional systems. *Instructional Science* 18, no. 2: 119–44.
- Sadler, D.R. 1998. Formative assessment: Revisiting the territory. *Assessment in Education* 5, no. 1: 77–85.
- Scriven, M. 1991. Beyond Formative and Summative. I M.W. Mc Laughlin & D.C. Phillips (Red.), *Evaluation and Education: At a quarter Century*. Chicago: Chicago University Press.



- Shepard, L. 2000. *The Role of Classroom Assessment in Teaching and Learning*. CSE Technincal Report.
- Skolinspektionen. 2010. *Kontrollrättning av nationella prov i grundskolan och gymnasieskolan*.  
<http://www.skolinspektionen.se/Documents/Regelbunden-tillsyn/kontrollrattning/redovisning-regeringsuppdrag.pdf?epslanguage=sv>
- Skolverket. 1998. *Bedömningsanvisningar till ämnesprovet i matematik*. Stockholm: Skolverket.
- Skolverket. 2000. *Kursplaner och betygskriterier 2000. Grundskolan*. Stockholm: CE Fritzes.
- Skolverket. 2007. *Provbetyg ? Slutbetyg ? Likvärdig bedömning? En statistisk analys av sambandet mellan nationella prov och slutbetyg i grundskolan, 1998-2006*. Stockholm: Skolverket.
- Skolverket. 2009. *Bedömningsanvisningar till nationella ämnesprovet i matematik*. Stockholm: Skolverket.  
[http://www.prim.su.se/matematik/tidigare\\_9.html](http://www.prim.su.se/matematik/tidigare_9.html)
- Skolverket. 2009. *Bedömaröverensstämmelse vid bedömning av nationella prov*. Stockholm: Skolverket.
- SOU 1992:94. *Skola för bildning*. Betänkande av läroplanskommittén. Stockholm: Norstedts tryckeri AB.
- Säljö, R. 2000. *Lärande i praktiken – ett sociokulturellt perspektiv*. Stockholm: Bokförlaget Prisma.
- Utbildningsdepartementet (1998). *Läroplan för det obligatoriska skolväsendet, förskoleklassen och fritidshemmet. LPO94*. Stockholm: CE Fritzes.
- Wyatt, C., V. Klenowski, och S. Gunn. 2010. The centrality of teachers' judgement practice in assessment: a study of standards in moderation. *Assessment in education: Principles, Policy & Practice* 17, no. 1: 59-75.

<http://www.skolinspektionen.se/Documents/Nyheter/U2009-4877-G.pdf?epslanguage=sv>  
100405 (Lund)

## Bilaga

### Missiv

Bilaga till Självständigt arbete i Masterprogrammet i utbildningsvetenskap vid Högskolan Kristianstad.

Ett acceptabelt koordinatsystem: om bedömning av nationella prov med stöd av bedömningsmatriser

Tänkt tidskrift: *Assessment in education: Principles, Policy & Practice*.

Tidskriftens inriktning:

Recent decades have witnessed significant developments in the field of educational assessment. New approaches to the assessment of student achievement have been complemented by the increasing prominence of educational assessment as a policy issue. In particular, there has been a growth of interest in modes of assessment that promote, as well as measure, standards and quality. These have profound implications for individual learners, institutions and the educational system itself.

*Assessment in Education* provides a focus for scholarly output in the field of assessment. The journal is explicitly international in focus and encourages contributions from a wide range of assessment systems and cultures. The intention is to explore both commonalities and differences in policy and practice.

*Assessment in Education* is the official journal of the International Association for Educational Assessment (IAEA).

Hämtat från tidskriftens websida 101012

<http://www.tandf.co.uk/journals/titles/0969594x.asp>

Tidskriftens riktlinjer

Artikeln ska omfatta mellan 5-7000 ord. Min artikel är på drygt 7 000 ord, inkl. utdrag ur mina transkript. Har inte hittat någon information om dessa ska räknas in i antal ord eller inte. Däremot har jag kollat andra artiklar som innehåller liknande utdrag och då hamnar även dessa artiklar över 7000 ord vilket jag har tolkat som att de *inte* ska räknas in. I så fall håller artikeln sig inom ovanstående gränser.