

Textbedömning i svenskämnet – attityder, erfarenheter och variation

E Aldrin

Sammanfattning

Artikeln undersöker praktiserande svensklärares attityder och erfarenheter kring textbedömning, samt betydelsen av sociokulturella faktorer för bedömarvariation. Materialet består av en webbenkät som besvarats av 135 svensklärare i Västsverige verksamma i årskurs 7–9. Frågorna behandlade textbedömning generellt, det nya betygssystemet, samt bedömning och beskrivning av en autentisk elevtext. Resultaten visar flera goda exempel och erfarenheter, men också behov av ökad information från Skolverket på flera punkter. Vidare framkommer en påfallande bedömarvariation, som föreslås hänga samman med en stor variation i synen på textkvalitet bland deltagarna. Bedömarvariationens samspel med den sociala kontexten (läraren, dennes erfarenhet samt skolmiljön) är begränsad men visar ett intressant mönster som efterfrågar fortsatt forskning.

Nyckelord: bedömning, elevtexter, svensklärare, attityder, variation



Emilia Aldrin

Emilia Aldrin är lektor i svenska språket och verksam inom ämneslärarutbildningen i svenska på Högskolan i Halmstad. Hennes forskning rör sig främst inom sociolingvistik och behandlar relationen mellan språk och samhälle. I den här studien intresserar hon sig för elevtextbedömning ur ett socialt perspektiv.

Aldrin

Introduktion

Elevers skrivande och skrivutveckling har en central plats inom svenskämnet genom hela grundskolan och gymnasiet. En stor del av svensklärares arbete handlar om att läsa, bedöma och ge utvecklande respons på olika slags texter. Kraven är stora; både samhälle, elever och föräldrar förväntar sig likvärdighet och rättvisa i bedömningen. Samtidigt är textbedömning, som all kvalitativ bedömning, en komplex och grannlaga uppgift där många aspekter ska sammanvägas. Landets svensklärarytbildningar har länge fått kritik för att innehålla för lite träning i bedömning (Östlund-Stjärnegårdh 2002, s. 191) och statliga utredningar av de nationella proven har visat på betydande skillnader i bedömningen mellan enskilda lärare, skolor och grupper. Sedan 2011 finns också ett nytt betygssystem som får konsekvenser för textbedömningen.

Syfte

Hur ser då praktiserande svensklärare själva på den betydande del av arbetet som handlar om bedömning av elevers texter? Trots att textbedömning både nationellt och internationellt är ett relativt väl utforskat område vet vi inte så mycket om lärares egna attityder och erfarenheter. I denna artikel redovisas en explorativ studie med syfte att utforska vilka attityder och erfarenheter som finns bland svensklärare, dels vad gäller bedömning av elevtexter generellt och dels i förhållande till det nya betygssystemet. Ett underordnat syfte är att bidra till diskussionen om bedömarvariation vid bedömning av elevers skrivande genom att undersöka graden av samstämmighet när ett stort antal lärare bedömer en och samma text samt i vilken mån bedömningen påverkas av sociokulturella faktorer (såsom den bedömande lärarens kön, arbetserfarenhet samt typ av skolmiljö). Undersökningen har finansierats av Lärarytbildningsnämnden vid Högskolan i Halmstad och förhoppningen är att denna artikel ska vara till nytta för svensklärarytbyråer.

Tidigare forskning

Lärares bedömning av elevers skrivande har stått i fokus för en mängd forskning. I det här avsnittet diskuterar jag studier om variation i bedömning och bedömningsgrunder samt studier om textens och kontextens betydelse för bedömningen. Avslutningsvis tar jag även upp forskning om bedömning som utgår från lärarens eget perspektiv.

Variation i bedömning och bedömningsgrunder

Mycket forskning har ägnats åt att studera i vilken mån elevtextbedömning varierar mellan olika lärare. I Sverige visar resultaten entydigt att sådan variation finns och kan vara relativt stor. Östlund-Stjärnegårdh (1999) lät bedömargrupper om fyra lärare i varje grupp betygsätta gymnasietexter och fann en genomsnittlig variation på ett betygssteg (mellan de dåvarande betygen IG/G/VG/MVG). Författaren drar slutsatsen att "lärarna inte är överens, men att de heller inte tar ut så stora svängningar i bedömningen" (s. 30). Ciolek Laerum (2009) undersökte hur berättande elevtexter från årskurs 9 bedömdes av minst 4 olika lärare. Här visade resultaten att trots att en ge-

mensam bedömningsmatris användes, så kunde samma text i något fall betygsättas med upp till 3 olika betyg (dåvarande G/VG/MVG). Sammantaget var lärarna enbart överens om bedömningen av 35 % av texterna. Även i Norge, där mycket skrivforskning genomförts, visar undersökningar av nationella skrivprov relativt stor oenighet i bedömningen (Berge 1996, 2002), då nästan en tredjedel av de 60 bedömarna avvek systematiskt från övriga i sina bedömningar av elevtexterna. Fasting m.fl. (2009, s. 632) tolkar det som att de norska lärarna har olika förväntningar på ungdomarnas skrivkompetens. En internationell studie (Brown, Glasswell, & Harland 2004) där 17 lärare fick särskild undervisning och handledning i bedömning före bedömningen visade emellertid att deltagarna uppvisade "substantial agreement" (s. 114).

Studier av omrättning av nationella prov i Sverige har uppmärksammat att externa bedömare ofta är strängare och ger lägre betyg än den ordinarie läraren, särskilt när det gäller bedömning av längre texter (Östlund-Stjärnegårdh 2002, Östlund-Stjärnegårdh 2009, Skolverket 2009, Skolinspektionen 2012, Skolinspektionen 2013, Lundgren 2013). Liknande resultat har visats internationellt när det gäller bedömning av längre texter (McKinstry m.fl. 2004). Gibbons & Marshall (2010, s. 36) fann dock i en engelsk studie ingen sådan tydlig tendens, utan där bedömde den externa bedömare ibland strängare och ibland snällare. Östlund-Stjärnegårdh (2009, s. 11) förklarar skillnaden mellan externa bedömares och ordinarie lärares betyg dels med att läraren använder all sin kunskap om eleven och hans eller hennes sätt att uttrycka sig liksom elevens övriga prestationer på ett sätt som en extern bedömare inte kan, dels med att det kan finnas en tendens att den egna läraren "hellre friar än fäller", medan externa bedömare kan vilja "visa att de minsann vet vad som krävs för de olika betygen". Gustafsson & Erickson (2013) som har framfört omfattande kritik av studierna kring nationella provets omrättning påpekar att det också kan handla om att urvalsprocessen för externa bedömare medfört att fler stränga personer funnits i denna grupp än vad som är representativt för lärarkollegiet som helhet (s. 76–77).

Det har föreslagits att vissa texter ger upphov till större variation i bedömningen än andra. Löfquist (1989, s. 89) menar att texter som alla lärare är överens om är en företeelse som "inte alls är typisk utan snarare extrem". Östlund-Stjärnegårdh (1999) finner att majoriteten av elevtexterna i hennes material tillhör en så kallad mellangrupp, där bedömningen varierar men ändå inte särskilt mycket. Hon menar också att det är någonstans i denna mellangrupp som den svårdragna gränsen mellan två betyg bör gå (1999, s. 31). I en senare studie (Östlund-Stjärnegårdh 2002, s. 71) finner hon att lärare är mest överens om elevtexter av lägre kvalitet (dåvarande betyget IG). De statliga granskningarna av nationella prov i svenska visar också att det främst är vid de högsta betygsstegen som den ordinarie lärarens bedömning och en extern bedömares skiljer sig åt (Skolinspektionen 2012, s. 21; 2013, s. 16).

En viktig orsak till svårigheten att uppnå bedömaröverensstämmelse för elevtexter i svenska är att texter och skrivande är en komplex verksamhet som kan bedömas ur flera olika aspekter vilka både kan uppfattas, prioriteras och vägas samman på olika sätt av olika läsare. Dysthe m.fl. (2002, s. 38 jfr Hillocks 1987) har beskrivit texters komplexitet som en upp-och-nervänd triangel indelad i olika skikt eller nivåer: skriv-

Aldrin

situation (textens anpassning till syfte och mottagare), innehåll, struktur, meningar, ordval samt rättskrivning och formalia. Vissa textaspekter såsom "skrivsituationen" rör sig på en mer global nivå (är övergripande och påverkar stora delar av texten) och har därför placerats i triangelns bas, medan andra textaspekter såsom "rättskrivning och formalia" är på en mer lokal nivå (påverkar delar och detaljer av texten) och har placerats i triangelns spets. Under skrivandets gång rör man sig ofta från globala till alltmer lokala nivåer, även om vana skribenter också kan arbeta på flera nivåer samtidigt (Dysthe m.fl. 2002, s. 39). Palmér & Östlund-Stjärnegårdh (2005) som har utarbetat en holistisk modell för bedömning av elevtexter framhåller också betydelsen av en helhetsbedömning där helhetsintrycket avgör omdömet, även om de olika textnivåerna beaktas som delar av helheten (s. 19, 27, 152). Dessutom ingår texter och skrivande i en social praktik som formas på olika sätt i olika kulturer och sammanhang och som också påverkar bedömningen (Berge 2002, s. 462).

Trots texters och textbedömningens komplexa natur visar flera studier en generell samstämmighet när det gäller de bedömningsgrunder lärare utgår från. En undersökning om lärares gränsdragning mellan IG och G (Östlund-Stjärnegårdh 1999) visade att faktorer relaterade till texten som helhet var allra vanligast, följt av disposition, innehåll, språk och stil. Bergman-Claesson (2003) har visat stora likheter även i lärares elevrespons. Mest frekventa var kommentarer till textens yta följt av textens innehåll, medan det som kommenterades minst var läsbarhet, effekt och berättarperspektiv (Bergman-Claesson 2003, s. 51). Lärares skrivpedagogiska ställningstagande och arbetssätt avspeglades istället mer i hur de kommenterade (s. 64–67). Även Kronholm-Cederberg (2009, s. 283) finner att de flesta av lärarnas kommentarer till elevtexter rör textens yta och den lokala textnivån.

Den sociala kontextens betydelse för bedömningen

Synen på bedömning som en socialt situerad praktik är relativt ny inom skrivbedömningsfältet (se t.ex. McNamara 2001, Berge 2002). I Sverige har endast ett fåtal studier relaterat textbedömning till sociokulturella faktorer eller en lokal kontext (se Skar 2013 för en översikt). Några studier har undersökt sociokulturell variation och funnit att flickor ofta får högre betyg än pojkar (Klapp Lekholm 2008 s. 83) och att flickors texter oftare än pojkars fått högsta betyg av den ordinarie läraren, men ett lägre betyg av externa bedömare (Östlund-Stjärnegårdh 2009, s. 8; Skolinspektionen 2013, s. 24; Lundgren 2013 s. 15). En studie visar att elever med utländsk bakgrund samt elever vars föräldrar saknar högskoleutbildning oftare än andra har fått ett lägre betyg av den ordinarie läraren än av omrättaren på nationella provet (Tyrefors & Vlachos 2012, s. 111–112). Andra studier visar att sociokulturella faktorer i elevens hemmiljö och lärarens (omedvetna) förväntningar på den aktuella eleven påverkar bedömningen (Klapp Lekholm 2008 s. 31–33; 82–87). Allas förväntningar är inte heller lika; lärarens kön, ålder och behörighet har visat sig ha en viss, om än liten, inverkan (Tyrefors & Vlachos 2012, s. 124–125, Klapp Lekholm 2008 s. 86). Den omgivande skolkontexten skulle också kunna påverka bedömningen. Lundgren (2013, s. 15) finner att texter skrivna inom yrkesförberedande program oftare fått en strängare bedömning vid

omrättningen än texter skrivna inom högskoleförberedande program, vilket kan tyda på att skolans kultur och elevsammansättning också påverkar lärarens förväntningar. Det finns dock ett stort behov av fler studier om hur textbedömningen påverkar och påverkas av det sociala sammanhang och den sociala praktik som den ingår i.

Bedömning ur lärarperspektiv

Merparten av forskningen om textbedömning har utgått från ett externt, granskande perspektiv. Praktiserande svensklärares egen syn på och erfarenhet av textbedömning är däremot mindre utforskat, även om några studier finns. Mickwitz (2011) har i en samhällsvetenskaplig licentiatavhandling intervjuat svensklärare angående deras generella syn på betygsättning i svenskämnet. Hon finner att ett av de teman som genomsyrade lärarnas berättelser var "betygens likvärdighet". Lärarna såg det som omöjligt att åstadkomma en betygsättning som är helt oberoende av lärare och skola, även om det ansågs vara ett ideal att sträva efter (s. 85). Samtidigt beskrevs problemet med betygens likvärdighet i första hand som något som rörde "andra lärare" och "andra skolor" (s. 86). Korp (2006) har i en utbildningsvetenskaplig avhandling intervjuat gymnasielärare i bland annat svenska om deras inställning till betygsättning och nationella prov. Hon finner att många svensklärare använde sig av en s.k. "analytisk modell" (s. 219), där betyg på kursen inte avgörs av betyg på enskilda uppgifter utan av en helhetsbedömning av elevernas uppfyllelse av kursplanemålen utifrån ett brett bedömningsunderlag. Detta innebar att enskilda uppgifter inte alltid betygsattes för att undvika att sända felaktiga signaler till eleverna. Östlund-Stjärnegårdh (2002, s. 175–181) har genomfört mindre intervjuer med fyra svensklärare i gymnasiet och Komvux. Hon finner att lärarna tycker att bedömning både är intressant och betungande och att det som framför allt upplevs underlätta bedömningsprocessen är sambedömning samt i någon mån det nationella provets bedömningsunderlag. I en norsk undersökning har Berge (1996, 2002) intervjuat modersmåls lärare om bedömning av elevtexter. Han finner att de förhåller sig till mer eller mindre uttalade textnormer i sin bedömning, som innebär att redan elevers val av vissa genrer och visst innehåll uppfattas vara mer prestigefyllt och förknippat med högre textkvalitet (Berge 2002, s. 464, 467, 483).

Material och metod

I syfte att få insyn i praktiserande svensklärares egna attityder till och erfarenheter av textbedömning och samtidigt kunna undersöka den sociala kontextens betydelse för bedömarvariationen, utan att ge avkall på möjligheten till viss generalisering, skapades en digital webbenkät (se bilaga 1). Information om studien samt en länk till enkäten skickades via e-post till ett stort antal lärare i Västsverige efter godkännande från skolans rektor. Enkäten skickades ut i början av våren 2014, då de nya betygen hade tillämpats under några terminer. Eftersom en stor del av den tidigare forskningen om bedömning i svenskämnet berört gymnasienivå, riktades denna studie till årskurs 7–9. Av de totalt 341 skolor som kontaktades deltog till sist endast lärare från 55 skolor (16 %). Av de 191 lärare som fick enkäten var det dock hela 71 % (135 st.) som

Aldrin

valde att delta, vilket antyder att ämnet ansågs relevant och angeläget. Det finns även ett visst internt bortfall, som är koncentrerad till de öppna frågorna som kräver ett egenformulerat svar. Därför varierar det totala antalet svarande mellan olika frågor i resultatredovisningen nedan.

Eftersom deltagarurvalet skett slumpmässigt finns både mer och mindre erfarna lärare med i studien, liksom lärare från olika typer av skolor. En majoritet av deltagarna (76 %) identifierar sig som kvinnor, vilket också är fallet i grundskollärargruppen generellt i landet. En majoritet av lärarna har också arbetat länge (medianantalet år som svensklärare i årskurs 7–9 är 13 år), vilket även det stämmer ganska väl med den nationella bilden. Samtidigt finns också nyare lärare med i studien (17 % har arbetat max 5 år). När det gäller olika typer av skolor anser de flesta lärare att det varken är särskilt många eller få högpresterande elever på deras skola (medelsvaret är 2,7 på en skala från 1–5). En mindre del (14 %) anser dock att det finns många högpresterande elever på skolan, medan bara ett fåtal (7 %) anser att det finns få högpresterande elever på skolan. Gällande flerspråkighet är den generella uppfattningen att det finns ganska få flerspråkiga elever på skolorna (medelsvaret är 3,74 på en skala från 1–5). Det finns dock också deltagare (12 %) som anser sig arbeta på en skola med många flerspråkiga elever. Sammantaget bedömer jag deltagargruppens representativitet som relativt god trots att urvalsgruppen är liten.

Enkäten innehöll inledningsvis tre frågor om attityder och erfarenheter av bedömning och tre frågor om attityder och erfarenheter av det nya betygssystemet. Attitydbegreppet definieras här i enlighet med den språkvetenskapliga attitydforskningen (se t.ex. Garrett 2010) som en relativt varaktig värderande inställning till ett visst fenomen. Attityder består av en komplex kombination av föreställningar om attitydobjektet, känslor och värderingar kring det samma, samt agerande i förhållande till attitydobjektet utifrån dessa aspekter. Denna studie fokuserar dock främst på den känslomässiga aspekten, som ofta också betraktas som mest central (Garrett 2010, s. 23). Upplägget på frågorna innebar att en första fråga gällde lärarens generella känslomässiga inställning till textbedömning (enkelt - svårt) respektive det nya betygssystemet (positiv - negativ). På så vis ville jag skapa en inledande bild av hur lärarna upplever textbedömningspraktiken. Svaret angavs på en kvantitativ skala i 5 steg där ytterpolerna markerades med kvalitativa beskrivningar (t.ex. "Det är mycket enkelt"/"mycket svårt", "Jag är mycket positiv"/"mycket negativ") för att hjälpa deltagarna att tolka svarsskalan (se Nyström Höög 2005). Attityder uttrycks inte bara som starka värderingar utan också som likgiltighet eller ambivalens (Garrett 2010), varför skalan hade ett ojämnt antal svarsalternativ som möjliggör ett mittensvar som kan tolkas som "varken eller" alternativt "osäker/vet ej". Efter dessa inledande kvantitativa frågor följde för varje tema två öppna frågor där deltagarna ombads beskriva sina erfarenheter med egna ord, i syfte att skapa en fördjupad och mer praktisknära bild av lärarnas upplevelser av textbedömning.

Vidare ingick i enkäten en autentisk elevtext från årskurs 8 som läraren ombads bedöma och motivera sin bedömning av. Tidigare forskning har ofta låtit ett mindre antal lärare bedöma flera texter för att skapa s.k. bedömarprofiler och undersöka mer

generell samstämmighet i tolkning av kriterier etc. I den här studien vill jag istället undersöka vilken variation eller samstämmighet som finns i tolkningen av texters kvalitet och har därför låtit ett stort antal lärare bedöma en och samma text. Elevtexten (se bilaga 2) var en mycket kort (<500 ord), relativt känslös, berättande text skriven på temat "Barndom"¹. Skrivuppgiften ingick i en skolklass ordinarie verksamhet och syftade enligt läraren i första hand till att öva förmågan att uttrycka tankar och känslor samt skapa inlevelse och engagemang. Berättande texter är vanligt förekommande i hela skolsystemet (Palmér & Östlund-Stjärnegårdh 2005, s. 62) och är också en vanlig genre vid det nationella provet i svenska (Östlundh Stjärnegårdh 2009, s. 10). Den berättande texten "kräver både tankeförmåga och språklig förmåga" (Palmér & Östlund-Stjärnegårdh 2005, s. 29). Ju högre upp i åldrarna genren används, desto fler olika typer av berättande texter får eleverna skriva och desto mer komplicerade och varierade blir de i sin struktur (Palmér & Östlund-Stjärnegårdh 2005, s. 62, 64). Den valda elevtexten hade en varierande textkvalitet för att undvika en förenklad bedömning. Till sin hjälp vid bedömningen fick lärarna dels Skolverkets aktuella kunskapskrav, dels en uppmaning om att utgå från ett antal givna bedömningsaspekter (helhetsintryck, genreanpassning, disposition, styckeindelning och bindning, meningsbyggnad, ordval samt språkriktighet) som anknyter till befintliga teorier och modeller för textbedömning (Dysthe m.fl. 2002, Palmér & Östlund-Stjärnegårdh 2005, Skolverket 2011). I övrigt formulerades motiveringen av bedömningen fritt. Undersökningsupplägget inkluderade även ett socialt experiment som innebar att elevtexten försågs med ett av tre olika personnamn med varierande sociokulturella associationer som skulle kunna påverka lärarens bedömning (se Aldrin u.u.). För att belysa detta infogades även några specifika frågor om lärarens upplevelse av texten som skulle kunna påverkas av sociokulturella fördomar, vilka därför inte diskuteras här.

För att möjliggöra en sociokulturell analys efterfrågades även vissa bakgrundsfakta (lärarens kön, ålder och typ av skolmiljö). Webbenkäten var öppen under två månaders tid och formatet tillät deltagarna att pausa och återvända till enkäten flera gånger, men inte ändra redan registrerade svar. Resultaten har analyserats kvantitativt och kvalitativt. Frågor med bundna svarsalternativ har analyserats med hjälp av kvantitativa beräkningar och statistiska signifikanstest. En del resultat som inte är statistiskt signifikanta kommer ändå att redovisas, då de uppvisar ett intressant mönster. Öppna frågor har analyserats kvalitativt, ur ett främst tematiskt perspektiv.

Resultat

Lärarnas attityder till textbedömning

Ett första tema som ska undersökas är lärarens generella känslomässiga inställning till bedömning av elevers texter. De flesta upplever detta som vare sig särskilt svårt eller enkelt (alternativt både och). På en skala från 1–5 (där 1 motsvarar "mycket enkelt" och 5 motsvarar "mycket svårt") har ungefär hälften av de 133 lärarna (52 % eller 69 st.)

¹ Såväl elev som vårdnadshavare, lärare och rektor har godkänt textens medverkan i studien.

Aldrin

angett mittenalternativet (3). Medelsvaret hamnar också kring mitten (2,92). Bland de lärare som valt att polarisera sitt svar, är det något fler som lutar åt den enklare ytterpolen av skalan (27 % eller 36 st.) än åt den svårare (21 % eller 28 st.). Bara enstaka personer (4 st.) har angett något av de yttersta svarsalternativen ("mycket enkelt" eller "mycket svårt"). Sammantaget har lärarna i enkäten alltså inte någon extrem eller entydig uppfattning om bedömning av elevtexter, vilket inte är förvånande med tanke på textbedömningens komplexitet.

Inställningen till elevtextbedömning varierar delvis utifrån den sociokulturella faktorn lärarens arbetserfarenhet. Lärare som har arbetat 16–20 år har uppfattat textbedömningen som allra svårast (medelvärdet är här högst: 3,08), men undantaget denna grupp tycks lärarna uppfatta bedömning som svårare ju kortare tid de har arbetat, vilket också visas av tabell 1. Sambandet är visserligen mycket svagt och inte statistiskt signifikant², men det ligger i linje med internationella forskningsresultat som pekat på att tilliten till den egna textbedömningen är större bland lärare med längre erfarenhet (Hunter m.fl. 2006, s. 60). Det är inte konstigt att bedömningen kan upplevas enklare med ökad erfarenhet och träning. Med tanke på textbedömningens komplexitet och kritiken mot bristande träning inom svensklärarytbildningarna är det snarare påfallande att skillnaderna inte är större. Kan här finnas ett mörkertal bland mindre erfarna lärare som inte vill visa att man finner detta svårt eller rör det sig rent av om en omedvetenhet om bedömningens problematik?

Arbetserfarenhet, antal år	Attityd till bedömning, medelvärde skala 1–5
Max 5 år	2,96
6–10 år	2,90
11–15 år	2,87
16–20 år	3,08
Mer än 20 år	2,84

Tabell 1. Generell inställning till bedömning av elevtext, relaterat till lärarens arbetserfarenhet (n 136).

Upplevda svårigheter

I en särskild fråga tillfrågades lärarna om huruvida de ser några svårigheter med bedömning av elevers texter. Frågan var öppet formulerad, men som en tolkningshjälp angavs inom parentes några exempel ("Det kan t.ex. vara att bedömning av en särskild texttyp är särskilt svår, gränsdragning mellan särskilda betygssteg, avvägning av olika bedömningsgrunder, att ge respons, eller något helt annat"). Endast 4 personer (av 123 som besvarat frågan) svarade att de inte upplever några svårigheter alls. Övriga beskrev en eller flera svårigheter, vilka sammanfattas i tabell 2.

² Sambandet är inte statistiskt signifikant. En χ^2 -analys ger p 0,629.

Svårighet	Andel svarande (antal)
Gränsdragning mellan betygssteg	45 % (56)
Ge respons	28 % (34)
Sammanvägning av olika aspekter av texten	20 % (25)
Tolka styrdokumentens kunskapskrav	15 % (18)
En särskild texttyp	11 % (14)
Likvärdighet och objektivitet	11 % (14)
Tidsbrist	5 % (6)
Övriga svar	5 % (6 st)

Som framgår av tabell 2 har inget svar angetts av mer än hälften av lärarna i enkäten. Även om de flesta upplever någon svårighet med bedömning är det alltså stor variation i vad man upplever som svårt. Vanligast förekommande är att det upplevs svårt att dra gränser mellan olika betygssteg (angett av 45 % eller 56 st.). Intressant nog är det inte något enskilt betygssteg som här nämns oftare än andra, vilket skiljer sig från tidigare forskning (Östlundh-Stjärnegård 2002, s. 175), utan alla gränsdragningar i det nya betygssystemet framstår som lika problematiska. Eftersom det nya betygssystemet (A-F) vid enkätens genomförande endast hade tillämpats en kortare tid är detta inte förvånande och kan vara av övergående karaktär. Det är också vanligt att lärarna inte alls preciserar vilket/vilka betygssteg de syftar på. En lärare skriver:

Svårigheten ligger i att avväga när en text blir 'välutvecklad'. Vad exakt gör texten 'välutvecklad'? Steget från C till B/A således.

Som framgår av citatet hänger avgränsningen av betygsstegen för många samman med en svårighet att tolka styrdokumentens kunskapskrav och värdeord. Det förekommer dock även att lärarna nämner denna svårighet separat utan koppling till gränsdragning mellan betyg (angett av 15 % eller 18 st.). En lärare skriver:

Bedömningar i och för sig är inte svåra, dvs. att visa vad som skulle kunna höja kvaliteten i texten, svårigheterna kommer när den skall relateras till ett betygssteg vars definitioner av nödvändighet blir abstrakta och svårtolkade.

Den näst vanligaste svårigheten som lärarna nämner är att ge respons (angett av 28 % eller 34 st.). Många tar här upp svårigheten med att ge konstruktiv och utvecklande respons som inspirerar eleven till att faktiskt ta till sig kommentarerna och förbättra sitt skrivande. Många tar också upp att de upplever det svårt att få eleven att förstå responsen. Det antyds i flera fall att läraren gärna vill använda sig av vetenskapliga termer eller begrepp ur styrdokumentet för att precisera och motivera sina kommentarer, men att detta förhindras av att eleven inte förstår dessa. Några lärare nämner svårigheter med när respons ska ges samt hur mycket och på vilket sätt. Enstaka upplever att det är särskilt svårt att ge respons på mycket svaga texter (utan att riskera att signalera "ändra allt"), medan andra upplever det särskilt svårt med starka texter (där det inte upplevs finnas något att vidareutveckla). En lärare skriver:

Aldrin

Det är inte helt lätt att arbeta formativt. Elever behöver få respons, men på olika sätt. Det är tidskrävande. Ibland kan jag tycka att det är en svår balansgång mellan det formativa och det summativa. Elever behöver få förståelse för hur de ligger till i relation till kunskapskraven samtidigt behöver de få respons för att utvecklas i sitt lärande.

Precis som i citatet ovan har en del lärare tagit upp tidsbristen inom yrket som en särskild försvårande omständighet för att ge god respons. Ett fåtal har också angett tidsbristen separat (5 % eller 6 st.). Trots allt finner jag det överraskande att inte fler har tagit upp tidsbrist som en svårighet med textbedömning.

Den tredje vanligaste svårigheten handlar om att väga samman kvaliteter och brister på olika textnivåer och vilka delar som då ska ges mest betydelse (angett av 20 % eller 25 st.). En lärare skriver:

Innehåll kontra språkliga brister. Svårt då att veta vilket betyg som ska väga tyngst - språket el innehåll?

Många nämner här just en upplevd motsättning mellan språklig form, det vill säga textens mest lokala ytnivå, och textens övriga aspekter, där man menar att det ena ofta visar en högre kvalitet än det andra. Några beskriver mer specifikt en motsättning mellan språk och innehåll, men det förekommer också att man talar om språk kontra genreanpassning eller språk kontra kreativitet. Intressant nog uttrycker nästan alla lärare i studien en syn på texten som uppdelad i just två dimensioner. Det är bara ett fåtal som beskriver texter som mer komplexa och flerdimensionella (då i termer av "språk, struktur och innehåll" eller "språk, genreanpassning och dramaturgi"). Om detta ger en korrekt bild av den generella synen på texter, tycks här finnas en grannliga uppgift för landets lärarutbildningar i att förmedla moderna teorier om texters sammansatthet och olika nivåer (t.ex. Dysthe m.fl 2002; Palmér & Östlund-Stjärnegårdh 2005).

En upplevd svårighet som nämns mindre ofta är att någon särskild texttyp är särskilt svår att bedöma. Endast 11 % (14 st.) tar upp detta. Ingen enskild texttyp dominerar i lärarnas svar, utan allt från berättande texter och dikter till krönika, faktatext och argumenterande texter nämns i ungefär samma utsträckning. Detta är intressant eftersom tidigare forskning pekat på att framför allt skönlitterära texter anses svåra och skapar oenighet i bedömningen (Östlund-Stjärnegårdh 2002, s. 72, Berge 1996, s. 514, 516). Däremot stämmer det väl överens med granskningen av den senaste omgången av nationella provet (Skolinspektionen 2013, s. 24-25) där bedömarvariationen visade sig vara likartad för alla texttyper.

En annan svårighet som nämns mindre ofta är att bedöma "likvärdigt" och "objektivt". Endast en liten grupp (11 % eller 14 st.) har nämnt detta. Det som i den nationella debatten lyfts fram som det största problemet med bedömning och det som i en tidigare studie (Mickwitz 2011) visade sig vara ett genomgående tema i svensklärarnas samtal om betyg och bedömning, är alltså långt ifrån den mest framträdande svårig-

het som dessa enkätdeltagare upplever. Kanske är en orsak till detta att många ser på likvärdighetsproblematiken som något som främst rör "andra lärare" och "andra skolor" (jfr Mickwitz 2011, s. 85). Insikten om att det är svårt att uppnå likvärdighet i bedömningen är kanske svår att acceptera för den lärare som med bibehållen stolthet vill fortsätta utöva sin yrkespraktik. En annan möjlig orsak kan dock vara att temat upplevs självklart och uttjat efter alla offentliga debatter, eller att det upplevs mindre problematiskt med det nya betygssystemet. De enkätdeltagare som ändå tar upp detta problem uttrycker sig ofta försiktigt i termer av att man som lärare kan bli "hemmablind" utifrån sin elevgrupp eller "fylla i" textluckor utifrån sin kunskap om eleven. Andra kommentarer är att en elevs personliga stil kan appellera mer till en viss lärare, att det är svårt att bedöma elever som utvecklats mycket och att det blir svårare att bedöma likvärdigt ju fler kunskapskrav som ska examineras samtidigt i uppgiften.

Slutligen förekommer också andra svar som bara har nämnts av någon enstaka person (totalt 5 % eller 6 st.). Här beskrivs exempelvis att det är svårt att sammanväga bedömningar av olika uppgifter till ett slutbetyg och att det är svårt att bedöma texter skrivna av nya elever som man inte känner så väl.

Erfarenheter av faktorer som underlättar

Lärarna tillfrågades också om de har erfarenhet av något som kan underlätta bedömningen av elevtexter. Även denna fråga var öppet formulerad, men utan förslag på ingångar. Något färre besvarade denna fråga (115 st.) och variationen i svaren är också något mindre. En del valde att i svaret fokusera på befintliga strategier som de i nuläget använder sig av, andra har istället lyft fram behov och önskemål om hjälp som de inte har tillgång till. Ingen svarade nej på frågan. Svarsfördelningen visas i tabell 3.

Underlättar	Andel svarande (antal)
Sambedömning/gemensamma samtal om bedömning	50 % (57)
Matriser med tydliga bedömningskriterier	35 % (40)
Exempeltexter med omdöme och motiveringar	23 % (27)
En tydlig skrivuppgift	17 % (20)
Goda tidsramar	9 % (10)
Formativ bedömning	8 % (9)
Erfarenhet	4 % (5)
Enstaka bedömningskriterier för varje uppgift	4 % (5)
Övriga svar	10 % (11)

Det allra vanligaste svaret är att sambedömning och gemensamma diskussioner om bedömning och textkvalitet med andra lärare underlättar bedömningen av elevers texter. En lärare skriver:

Sambedömning i kollegiet. Diskussion av vad som är en "god" text.

Intressant nog är det ändå enbart hälften av lärarna som nämnt detta (50 % eller 57

Aldrin

st.), trots att det varit ett återkommande mönster att majoriteten tillämpat sambedömning vid nationella proven (Östlund-Stjärnegårdh 2002, s. 176; 2009, s. 16) och att det ständigt rekommenderas av de statliga utredningarna (Skolinspektionen 2012, s. 25; Skolinspektionen 2013, s. 33). De flesta formulerar sitt svar som att sambedömning är något man redan nu ägnar sig åt, men en del lärare lyfter fram att de i nuläget inte har tillräckliga tidsramar för att kunna ägna sig åt detta så mycket eller ofta som de skulle behöva. Kanske är detta en anledning till att inte fler nämner fördelarna med sambedömning.

Det näst vanligaste svaret är att man upplever att matriser och mallar med tydliga bedömningskriterier underlättar (angett av 35 % eller 40 st.). Här nämns nationella matriser från Skolverket och via Nationella provet, men också egna matriser som läraren eller lärarlaget själva ställer samman. En lärare skriver:

Att man gör en bedömningsmall innan man rättar texten, med vad som ska finnas med (vad gäller innehåll). Precis som de gör till nationella provet.

Ett tredje vanligt svar, som ofta tycks hänga nära samman med användningen av matriser men som också nämns separat, är tillgång till exempeltexter med både omdöme och motivering (angett av 23 % eller 27 st.). Dessa används dels som stöd i lärarens bedömning, dels som stöd för att motivera och förklara bedömningskriterierna för eleverna. Återigen handlar det både om att man utnyttjar tillgängliga exempel från Skolverket och via Nationella Provet och att läraren eller lärarlaget tagit fram egna exempel. En lärare skriver:

Fler exempeltexter med bedömning. Korta exempeltexter som belyser olika aspekter av textbedömning. Det material som finns för bedömningsstöd handlar om att tolka budskap i text etc. Förvisso finns gamla NP att tillgå men det skulle vara önskvärt med ett bedömningsstöd där man resonerar kring varför man bedömt på ett särskilt sätt. Jag vill gärna kunna förklara varför jag bedömt dem som E, C, A.

I detta sammanhang är det många som påpekar att det finns ett behov av fler exempeltexter från Nationella Provet och Skolverket, som inte bara anger omdöme utan också innehåller motiveringar, i synnerhet för svårbedömda texter samt för texter från årskurs 7 och 8.

Ett fjärde återkommande svar från lärarna är att bedömningen underlättas av en tydlig skrivuppgift med tydligt syfte, tydliga bedömningskriterier och tydliga instruktioner till eleverna (angett av 17 % eller 20 st.). Ju mer planering och information som görs i början av skrivmomentet, desto enklare blir bedömningen i slutet. En lärare skriver:

Förklara syftet med texten. Behöver inte vara några matriser eller fördjupande LPP:er utan en enkel och öppen uppstartsdiskussion i början av ett arbete. Krävs dock att man återkommande släpper fram denna diskussion under hela perioden.

Ett svar, som återigen intressant nog inte är särskilt frekvent, är behovet av goda tidsramar (9 % eller 10 st.). Här nämns både tid för att planera skrivuppgiften och tid för själva bedömningen. Ett annat mindre vanligt svar är att formativ bedömning underlättar (8 % eller 9 st.). Lärarna lyfter här fram betydelsen av såväl tidig respons från läraren, som kamratrespons och självbedömning. Jag uppfattar att det är det kontinuerliga arbetet med att höja textens kvalitet som upplevs minska arbetet i slutbedömningen. Vidare har några har påpekat att bedömningen blir enklare ju längre erfarenhet man har (4 % eller 5 st.). Några har påpekat att det underlättar att välja ut ett eller ett par kriterier för bedömning i en viss uppgift, snarare än att bedöma allt i alla texter (4 % eller 5 st.). Till sist förekommer också enstaka övriga svar som inte är återkommande (totalt angett av 10 % eller 11 st.), såsom att bedömningen underlättas av att eleverna skriver på dator, att bedömningen sker muntligt, att läraren får möjlighet att följa elevers skrivutveckling över tid samt att läraren har tillgång till ett lämpligt ordförråd för att beskriva texters kvalitet.

Lärarnas attityder till det nya betygssystemet

Undersökningens andra tema är svensklärarnas inställning till textbedömning utifrån det nya betygssystemet. Deltagarnas generella attityd till de nya betygen kan beskrivas som relativt positiv (medelvärde är 2,22 på skalan 1-5 där 1 motsvarar "mycket positiv"). Ungefär hälften (49 % eller 61 av 125 st.) har angett svaret (2) motsvarande en ganska positiv attityd. En fjärdedel (25 % eller 31 st.) har angett mittensvaret (3) och är alltså vare sig positiva eller negativa (eller både och). En femtedel (20 % eller 24 st.) har angett ytterpolssvaret (1) motsvarande en mycket positiv attityd. Endast ett fåtal lärare (7 % eller 9 st.) har angett att de är ganska eller mycket negativa till det nya betygssystemet. Inställningen samspelar inte med de undersökta sociokulturella faktorerna i någon större mån.

Fördelar med det nya betygssystemet

En särskild fråga uppmanade lärarna att med egna ord beskriva vad de ser som den största fördelen med det nya betygssystemet. Frågekonstruktionen i sig förutsatte alltså att det finns fördelar. Nästan alla (utom 3 st.) svarade att de kunde se fördelar och som framgår av tabell 4 är det framför allt en aspekt som dominerar svaren.

Fördelar	Andel svarande (antal)
Fler nivåer	78 % (96)
Ökad tydlighet	24 % (30)
Övriga svar	6 % (7)

Tabell 4. Fördelar med det nya betygssystemet (n = 125).

En överväldigande majoritet har lyft fram fördelen med det nya betygssystemets fler

Aldrin

nivåer (78 % eller 96 av 123 st. svarade detta). Av kommentarerna framgår att många uppfattar detta som mer rättvist och att särskilt det tidigare betyget Godkänd var alltför omfattande. Flera lyfter också fram fördelen med att nu slippa sätta ut plus- och minustecken i kanten. Vidare påpekar många att de fler betygsnivåerna möjliggör för eleven att snabbare se en positiv utveckling i omdömet när hen förändrar sin arbetsinsats och att detta underlättar för formativ bedömning. En lärare skriver:

Fler steg gör att det känns mer rättvist. På den gamla betygsskalan var spannet för G alltför stort. En elev som precis uppnådde målen kunde få detta betyg liksom en som låg väldigt nära VG.

En annan upplevd fördel är ökad tydlighet (24 % eller 30 st. har svarat detta), vilket delvis, men inte enbart, verkar hänga samman med de fler betygsnivåerna. Både kunskapskraven och progressionen mellan betygsnivåerna anses ha blivit tydligare och en del menar också att detta gjort det lättare att motivera bedömningen för eleverna. En lärare skriver:

Tydlighet. Klarare krav på vad som krävs för de olika betygen. Lite tydligare.

Utöver dessa vanliga svar förekommer också några andra upplevda fördelar (6 % eller 7 st.) som bara angetts av en eller två personer, såsom att det nya betygssystemet ställer högre krav i och med att alla kunskapskrav måste vara uppnådda för ett högre betyg, att betygen blir mer internationellt jämförbara och gångbara och att man uppskattar att det finns tydliga kopplingar mellan kunskapskrav, centralt innehåll och syfte i läroplanen.

Nackdelar med det nya betygssystemet

Lärarna fick också beskriva vad de ser som den största nackdelen med det nya betygssystemet. Nästan alla (utom 8 st.) svarade att de även ser nackdelar och återigen är det ett litet antal svar som återkommer. Variationen är dock något större här, vilket framgår av tabell 5.

Nackdelar	Andel svarande (antal)
Vaga och otydliga kunskapskrav	43 % (52)
Alltför hårda krav	29 % (35)
För många kunskapskrav	6 % (7)
Elever och föräldrar ser C som förväntad norm	5 % (6)
Ej möjligt vikta mellan kunskapskraven	3 % (4)
Övriga svar	18 % (22)

Den vanligast upplevda nackdelen med det nya betygssystemet är att det innehåller vagheter och otydligheter i kunskapskraven (43 % eller 52 av 121 har svarat detta).

Särskilt problematiska upplevs de värdeord som används för att signalera progression mellan betygsnivåerna (god - mycket god, till viss del - till övervägande del etc.), vilka beskrivs som subjektiva. Detta medför enligt många att bedömningen inte kan bli likvärdig över landet. En lärare skriver:

Omöjligt att tolka kriterierna på ett rättvist och likvärdigt sätt. Luddiga och otydliga formuleringar som kan betyda i princip vad som helst.

En annan nackdel som lyfts fram av många är att det nya betygssystemet upplevs som alltför hårt (29 % eller 35 st. nämner detta). Det handlar framför allt om ett missnöje med kraven för D och B, som innebär att kraven för nästa nivå (C respektive A) måste vara uppnådda "till övervägande del". Många menar att formuleringen "till övervägande del" är oklar. Någon lärare uttrycker att: "ganska duktiga elever får ganska dåliga betyg". I detta sammanhang beskrivs också ett problem med elever som har en ojämn kunskapsprofil vilka drabbas hårt av det nya betygssystemet, det vill säga elever som ligger högt på flera mål men dras ner av att de misslyckats med något enda. Man upplever här att betygssystemet inte tillåter eleverna att göra misstag. Det förekommer också att lärare upplever det alltför svårt att få högsta betyg. En lärare skriver:

Att kraven är hårdare, att man ska uppnått mål över de betyg som man får för att få det lägre. Svårt att förklara och motivera. Eleven får göra färre misstag.

Ett anslutande svar är att man upplever det problematiskt att alla kunskapskrav ses som lika viktiga och att det inte är möjligt att vikta mellan dessa (angett av 3 % eller 4 st.). Några upplever det faktiskt oklart om man kan vikta eller ej. Ett mindre vanligt svar är att man upplever att kunskapskraven är för många (angett av 6 % eller 7 st.). Ett annat mindre vanligt svar är att elever och föräldrar ofta upplever E som ett dåligt betyg och ser C som den självklara normen (motsvarande ungefär det tidigare Godkänd), vilket skapar orealistiska förväntningar (angett av 5 % eller 6 st.). Dessutom förekommer en relativt stor mängd övriga svar som enbart angetts av en eller två personer (18 % eller 22 st.). Här nämns bland annat att det är oklart om hela betygsskalan kan användas vid enskilda prov eller endast vid terminsslut, att det saknas kunskapskrav för årskurs 7 och 8, att betygssystemet är oflexibelt, att nivåerna är för många, att det spär på betygsinflationen och att det tar tid att ständigt implementera och vänja sig vid nya system. Mängden övriga svar för denna fråga tyder på att det, naturligt nog, finns många tankar kring förbättringsmöjligheter av det nya betygssystemet.

Variation i textbedömningen

I enkäten fanns också en autentisk elevtext från årskurs 8 som lärarna fick ge ett omdöme (enligt det nya betygssystemet A-F) och motivera. Texten fick det genomsnittliga omdömet 3,9 (i princip motsvarande omdömet D). En närmare analys visar dock

Aldrin

en relativt stor variation i bedömningen³. Som framgår av tabell 6 är det en mycket liten andel (3 % eller 4 st.) som har satt omdömet B, medan fördelningen mellan övriga omdömen (C, D, E) är i princip jämn. Ingen lärare har gett texten det allra högsta eller lägsta omdömet (A respektive F). Standardavvikelsen är 0,872 vilket innebär en genomsnittlig variation på nästan ett helt betygssteg. I jämförelse med tidigare forskning (t.ex. Östlund-Stjärnegårdh 2002, s. 65) är bedömarvariationen ändå mindre här. En total samstämmighet i bedömningen är naturligtvis orimligt att förvänta när 113 lärare bedömer en och samma text. Det som ändå är anmärkningsvärt är att det inte finns någon tydlig centrering av omdömen kring något enskilt betyg och att variationen spänner jämnt över hela tre olika betygssteg. Det finns alltså ingen majoritetsuppfattning om texten i gruppen. Detta är mycket ovanligt i tidigare forskning (se studier anförda i avsnitt 1.2.1 ovan). Det är också intressant att det, till skillnad från flera tidigare studier (Östlund-Stjärnegårdh 2002, s. 71; Skolinspektionen 2012, s. 21; 2013, s. 16), inte är de högsta betygsstegen som det här råder oenighet om, utan snarare mitten av betygsskalan. Att det rör sig om en elevtext av varierande kvalitet är säkert en bidragande orsak till detta, liksom att uppgiftspecifika betygskriterier för varje omdöme saknades. Det nya betygssystemet och lärarnas ovana med detta kan också ha haft någon betydelse i sammanhanget. Ändå finner jag bedömarvariationens mönster påfallande.

Omdöme	Andel svarande (antal)
A	-
B	3 % (4)
C	32 % (36)
D	34 % (39)
E	30 % (34)
F	-

Variation i bedömningsgrunderna

Lärarna fick även motivera sin bedömning. Vid första anblicken är dessa motiveringar relativt lika. De flesta har gjort en ganska utförlig bedömning som inkluderar flera aspekter av texten. En lärare skriver:

Helhetsintrycket och genreanpassning stämmer väl. Dispositionen och styckeindelningen störs av underrubriker som ex. Två veckor senare... Meningarna är varierade med ett relevant ordval men det förekommer exempel på brist i ändelse på verb, val av pronomen och bortfall av skiljetecken. Jag tycker att texten precis når upp till ett betyg C med tanke på innehållet, trots några få språkliga missar.

I princip utgår lärarna från de textaspekter som föreslogs i instruktionen: helhets-

3

Denna kvarstår till stor del inom de 3 textversionerna med olika namn.

intryck, genreanpassning, disposition, styckeindelning och bindning, meningsbyggnad, ordval samt språkriktighet. Många har också kommenterat textens innehåll samt förmåga att väcka engagemang och intresse, vilket nämns i uppgiftinstruktionen. Även om lärarna tittat på samma aspekter av texten, har de förhållit sig till dessa på väldigt olika sätt. Det är inte alls ovanligt att precis samma aspekt beskrivs visa särskilt hög kvalitet av någon lärare samtidigt som det beskrivs som särskilt bristande av någon annan. Exempelvis uppvisar materialet följande motsägelsefulla beskrivningar av texten⁴:

- ”Eleven håller sig väl inom genren berättande text” | ”Något tillyxad genreanpassning”
- ”En tydlig tankegång som förstärks av en konsekvent styckeindelning” | ”Bristande disposition, styckeindelning”
- ”Meningsbyggnaden fungerar” | ”Dock finns problematik gällande meningsbyggnad”
- ”Ordval och språkriktighet fungerar” | ”Ordval och språkriktighet lämnar en del kvar att önska”
- ”Innehållsmässigt utvecklad berättelse” | ”Däremot skulle innehållet kunnat utvecklas betydligt mer”

Det tycks därmed som att textens kvalitet ur olika hänseenden faktiskt uppfattas väldigt olika av olika lärare. Detta tyder på att det saknas vad Hultman & Westman (1977/1992, s. 25) kallar ”en gemensam skolnorm” vad gäller textbedömning (se även Berge 1996, s. 258). Ett liknande resonemang förs av Skar (2013, s. 59) som i sin forskningsgenomgång drar slutsatsen att elevtextbedömning ofta domineras av individuella (och inte gemensamma) föreställningar om det skrivkonstrukt som ska mätas. Svensklärarna i denna undersökning verkar inte ha en gemensam syn på vad som utgör högre eller lägre kvalitet ur olika hänseenden i en text av detta slag. Eftersom deltagarurvalet inte består av specialutbildade bedömare utan vanliga, intresserade lärare finner jag denna bristande samsyn viktigt att uppmärksamma.

Betydelse av sociokulturella faktorer

Lärarnas bedömning av texten tycks också hänga samman med externa sociokulturella faktorer, om än i liten utsträckning. Kvinnorna har i genomsnitt gett ett något sämre omdöme på texten än männen (medelvärdena är 4,01 respektive 3,70). Detta samband är också statistiskt signifikant⁵. Eftersom andelen män i undersökningen är liten bör resultatet ändå tas med försiktighet.

4 Denna kvarstår till stor del inom de 3 textversionerna med olika namn.

5 En χ^2 -analys, där de högsta och lägsta omdömena B och F inte har inkluderats då färre än 5 personer av varje kön valt dem, ger $p < 0,011$.

Aldrin

Omdöme	Kvinnor andel (antal)	Män andel (antal)
A	-	-
B	2 % (2)	7 % (2)
C	25 % (21)	48 % (13)
D	42 % (35)	15 % (4)
E	31 % (26)	30 % (8)
F	-	-
Totalt	100 % (84)	100 % (27)

Skolans sociokulturella miljö tycks också ha en viss betydelse, åtminstone när det gäller andel flerspråkiga elever. Lärare som arbetar på en skola som de uppfattar har många flerspråkiga elever har nämligen gett ett något bättre omdöme på texten än andra (medelvärde är här 3,31). Skillnaderna är små, men nästan statistiskt signifikanta⁶. Sambandet är dock inte linjärt, vilket också framgår av tabell 8. Lärare som arbetar på en skola med ”ganska många” flerspråkiga elever har gett texten det sämsta genomsnittliga omdömet (medelvärde är här 4,25). Möjligen är det lärarnas självuppskattning av skolans andel flerspråkiga elever som blir problematisk, eftersom samma antal kan tolkas på olika sätt av olika personer när det ska omformuleras i relativa termer av ”många eller få”. Att det bästa medelomdömet på texten finns bland lärare som uppfattar sig arbeta på skolor med många flerspråkiga elever är ändå anmärkningsvärt. Kanske kan en flerspråkig skolmiljö påverka lärarens uppfattning om genomsnittlig textkvalitet och öka risken för att eleverna får ett högre betyg än vad prestationen motsvarar. En annan möjlighet är att lärare som arbetar i flerspråkiga miljöer är mer benägna att ha överseende med brister i textens ytnivå.

Mängd flerspråkiga elever	Medelomdöme, A–F motsvaras här av 1–6
Många	3,31
Ganska många	4,25
Varken många eller få	3,91
Ganska få	4,14
Få	3,94

Slutligen har texten fått ett något bättre omdöme av lärare som arbetat i max 10 år än

⁶ En χ^2 -analys där enbart gruppen ”många flerspråkiga elever på skolan” jämförs med övriga ger $p = 0,065$. Kruskal-Wallis-testet ger $p = 0,081$.

av dem som arbetat längre (se tabell 9), vilket ligger i linje med en tidigare studie om betyg i svenskämnet (Klapp Lekholm 2008 s. 86). Bland de få lärare som gett texten det bästa omdömet (B) har samtliga (4 st.) arbetat mindre än 10 år som lärare. Skillnaderna är dock små och inte statistiskt signifikanta⁷.

Arbetserfarenhet, antal år	Medelomdöme, A–F motsvaras här av 1–6
Max 5 år	3,89
6–10 år	3,84
11–15 år	3,97
16–20 år	4,09
Mer än 20 år	4,07

Diskussion och slutsatser

Den här studiens omfång är starkt begränsat såväl till innehåll som till antal deltagare. Trots detta menar jag att den ger intressanta inblickar i svensklärares attityder till och erfarenheter av textbedömning och en antydning om den sociala kontextens betydelse för bedömningen som efterfrågar vidare forskning. Eftersom deltagarurvalet visar en relativt god representativitet kan vissa generella slutsatser dras. En första viktig slutsats är att svensklärarna i studien överlag har en nyanserad inställning till textbedömning och ser både svårigheter och faktorer som kan underlätta. Flera av de svårigheter som nämns hänger samman med det nya betygssystemet, men det framkommer också att nästan var tredje deltagare upplever det svårt att ge god respons och att var femte deltagare upplever det svårt att väga samman olika aspekter av texters kvalitet vilket efterfrågar mer fokus på detta inom lärarutbildningen. Studien visar också på flera goda exempel och idéer när det gäller strategier för att underlätta textbedömning. Förutom sambedömning, vilket är en dokumenterat effektiv strategi för ökad likvärdighet, framhäver många vikten av tydlighet i skrivuppgiften (som kan ta sig uttryck i såväl formulering av instruktioner som urval av enstaka kriterier för bedömning i varje uppgift, utarbetning av egna uppgiftsspecifika bedömningsmallar och exempeltexter med motivering av omdöme). Samtidigt efterlyses även fler och bättre bedömningsmatriser och exempeltexter med motiveringar från Skolverket. Den generella inställningen till de nya betygen är relativt positiv och det som framför allt uppskattas är de fler betygsnivåerna. De vanligast upplevda nackdelarna (svårtolkade bedömningskriterier och hårdheten i systemet) efterlyses också ett ökat informationsarbete från Skolverket.

Den kvantitativa undersökningen av bedömarvariation pekar på en relativt påfallande brist på samstämmighet som framför allt visar sig i att lärarnas bedömningar

⁷ En χ^2 -analys, där det högsta omdömet B inte har inkluderats då färre än 5 personer valt detta, ger p 0,300. Kruskal-Wallis-testet ger p 0,716.

Aldrin

fördelar sig jämnt över hela tre olika betygssteg. Resultatet bör dock tas med försiktighet eftersom lärarna i denna uppgift saknade flera av sina vanliga bedömningshjälpmedel (såsom en bedömningsmatris anpassad efter uppgiften). Mer anmärkningsvärt är den kvalitativa undersökningens resultat som pekar på en bristande samsyn bland lärarna vad gäller textkvalitet och hur den ska tolkas i ett konkret textexempel, vilket yttrade sig i att samma aspekt av texten ofta kunde anses uppvisa hög respektive låg kvalitet av olika lärare. Detta uppmärksammar behovet av ett ökat medvetet arbete för utvecklingen av en gemensam bedömningsnorm bland svensklärare och pekar återigen på vikten av sambedömning och betydelsen av att sådana rutiner kan ges utrymme inte bara i samband med det nationella provet utan också i lärarnas vanliga vardag.

Studien visar slutligen också att skillnader i lärares textbedömning inte enbart varierar slumpmässigt, utan i någon mån hänger samman med den sociala kontext som den äger rum i och som inkluderar läraren, hans eller hennes erfarenhet, samt den miljö som finns på skolan. Eftersom studien är begränsad och resultaten här inte alltid statistiskt signifikanta behöver dessa tendenser undersökas ytterligare i framtida forskning för att öka kunskapen om hur de eventuellt påverkar likvärdigheten i bedömning av elevtexter.

Referenser

- Aldrin, E. (under utg.). Assessing names? Effects of name-based stereotypes on teachers' evaluation of pupils' texts. (planerad utg. i *Names: A Journal of Onomastics*).
- Berge, K. L. (1996). *Norsksensorernas tekstnormer og doxa: En kultursemiotisk og sosioteknologisk analyse*. Trondheim: Universitetet Trondheim.
- Berge, K. L. (2002). Hidden norms in assessment of students' exam essays in Norwegian upper secondary schools. *Written Communication* vol. 19, nr. 4, ss. 458–492.
- Bergman-Claesson, G. (2003). *Tre lärare – tre världar. Lärarkommentarer till elevtexter i tre gymnasieklasser*. Uppsala: Uppsala universitet.
- Brown, G.T.L., Glasswell, K. & Harland, D. (2004). Accuracy in the scoring of writing: Studies of reliability and validity using a New Zealand writing assessment system. *Assessing Writing*, vol. 9, nr. 2, ss. 105–121.
- Ciolek Laerum, B. (2009). *Elever skriver och lärare bedömer – en studie av elevtexter i åk 9*. (Svenska i utveckling 25; FUMS Rapport 226.) Uppsala: Uppsala universitet.
- Dysthe, O., Hertzberg, F. & Løkensgard Hoel, T. (2002). *Skriva för att lära*. Lund: Studentlitteratur.
- Fasting, R.B., Thygesen, R., Berge, K.L., Evensen, L.S. & Vagle, W. (2009). National Assessment of Writing Proficiency Among Norwegian Students in Compulsory Schools. *Scandinavian Journal of Educational Research*, vol. 53, nr. 6, ss. 617–637.
- Garrett, P. (2010). *Attitudes to language*. Cambridge: Cambridge University Press.
- Gibbons, S. & Marshall, B. (2010). Assessing English: A trial collaborative standardised marking project. *Assessing English: A trial collaborative standardised marking project*, vol. 9, nr. 3, ss. 26–39.

- Gustafsson, J.-E. & Erickson, G. (2013). To trust or not to trust?—teacher marking versus external marking of national tests. *Educational Assessment, Evaluation and Accountability*, vol. 25, nr. 1, ss. 69–87.
- Hillocks, G. (1987). Synthesis of Research on Teaching Writing. *Educational Leadership*, vol. 44, nr. 8, ss. 71–82.
- Hultman, T. G. & Westman, M. (1977/1992). *Gymnasistsvenska*. (Nytryck i nordiska språk och svenska NNS 3). Stockholm: Stockholms universitet / Edsbruk: Akademitryck.
- Hunter, D., Mayenga, C., & Gambell, T. (2006). Classroom assessment tools and uses. Canadian English teachers' practices for writing. *Assessing Writing*, vol. 11, nr. 1, ss. 42–65.
- Klapp Lekholm, A. (2008). *Grades and grade assignment: effects of student and school characteristics*. Diss. Göteborg: Göteborgs universitet.
- Kronholm-Cederberg, A. (2009). *Skolans responskultur som skriftpraktik : gymnasis- ters berättelser om lärarens skriftliga respons på uppsatsen*. Diss. Åbo: Åbo Akademi University.
- Lundgren, P. (2013). *Om bedömning av skrivdelen i nationella provet i svenska i årskurs 9 och Sv1 vårterminen 2012*. Stockholm: Utbildningsförvaltningen, Stockholms stad.
- Löfquist, G. (1989). *Lärare bedömer uppsatser*. (Pedagogisk-psykologiska problem 510.) Lund: Lunds universitet.
- McKinstry, B., Cameron, H., Elton, R. & Riley, S. (2004). Leniency and halo effects in marking undergraduate short research projects. *BMC Medical Education*, vol. 4, nr. 28.
- McNamara, T. (2001). Language assessment as social practice: challenges for research. *Language testing*, vol. 18, nr. 4, ss. 333–349.
- Mickwitz, L. (2011). *Rätt betyg för vem? Betygsättning som institutionaliserad praktik*. (Studier i språkdidaktik 3.) Lic. Stockholm: Stockholms universitet.
- Nyström-Höög, C. (2005). What do people actually think? On scale measuring and personal narratives in attitude studies. *Acta Linguistica Hafniensia*, vol. 37, ss. 193–215.
- Palmér, A. & Östlund-Stjärnegårdh, E. (2005). *Bedömning av elevtext: en modell för analys*. Stockholm: Natur och kultur.
- Skar, G. (2013). *Validitetsperspektiv på skrivbedömning. Fallstudier av skrivbedömning i svenskundervisning på gymnasiet*. Diss. Stockholm: Stockholms universitet.
- Skolverket (2009). *Bedömaröverensstämmelse vid bedömning av nationella prov*. (Dnr 2008: 286) Stockholm: Skolverket.
- Skolverket (2011). *Språket på väg. Ett kartläggningsmaterial i svenska och svenska som andraspråk för grundskolans åk 7-9. Reviderad upplaga enligt Lgr 11. Del 2*. Stockholm: Fritzes.
- Skolinspektionen (2012). *Lika för alla? Omrättning av nationella prov i grundskolan och gymnasieskolan under tre år*. (Dnr U2009/4877/G) Stockholm: Skolinspektionen.

Aldrin

- Skolinspektionen (2013). *Olikheterna är för stora. Omrättning av nationella prov i grundskolan och gymnasieskolan, 2013*. (Dnr U2011/6544/GV). Stockholm: Skolinspektionen.
- Tyrefors Hinnerich, B. och Vlachos, J. (2012). Systematiska skillnader mellan interna och externa bedömningar av nationella prov. I: Resultatbilaga till *Lika för alla? Omrättning av nationella prov i grundskolan och gymnasieskolan under tre år*, Skolinspektionen, ss. 63–122.
- Östlund-Stjärnegårdh, E. (1999). *Principen och praktiken. En enkätundersökning av lärares syn på bedömning av gymnasieelevers texter*. (Svenska i utveckling 12.; FUMS Rapport 195.) Uppsala: Uppsala universitet.
- Östlund-Stjärnegårdh, E. (2002). *Godkänd i svenska? Bedömning och analys av gymnasieelevers texter*. (Skrifter utgivna av Institutionen för nordiska språk vid Uppsala universitet 57.) Diss. Uppsala: Uppsala universitet.
- Östlund-Stjärnegårdh, E. (2009). *Bedömaröverensstämmelse – ämnet svenska*. Uppsala: Uppsala universitet.