

# 7. Hur betygsätts muntlig språkfärdighet i engelska? En studie av lärares resonemang kring bedömning av det nationella provet för årskurs 9

Maria Frisch

## Inledning

- Vad sätter vi på tjejens prestation, då? Jag...
- Du nämnde E.
- Ja.
- Jag hade C, men nedre skiktet då. För jag tyckte att hon var så pass kommunikativ, faktiskt.
- Jag tänkte i början att det var ett E och sedan så gick det upp till D.
- Ja. Ja, så gjorde jag med.
- Ah, ja, då kommer jag och sträcker ut mig ännu mer. Jag har varit uppe på C, fast övre.
- Oj, oj!

Så kan det initialt låta, när lärare från olika skolor och kontexter träffas och tillsammans diskuterar bedömning av ett inspelat elevsamtal från de nationella proven i engelska. Erfarenheterna varierar beroende på det lokala sammanhanget. Det kan leda till att lärare bedömer relativt, dvs. rangordnar utifrån egen erfarenhet av den aktuella elevgruppen istället för i förhållande till kunskapskrav och bedömningsfaktorer.

---

### Hur du refererar till det här kapitlet:

Frisch, M. (2021). Hur betygsätts muntlig språkfärdighet i engelska? En studie av lärares resonemang kring bedömning av det nationella provet för årskurs 9. I Bardel, C. et al. (Red.). *Forskarskolan FRAM — lärare forskar i de främmande språkens didaktik* (s. 157–175). Stockholm: Stockholm University Press. DOI: <https://doi.org/10.16993/bbg.h>. License: CC-BY 4.0.

För att stödja lärares bedömning finns ett omfattande lärarmaterial som levereras tillsammans med proven. Syftet med de nationella proven är att bidra till en rättvis och likvärdig bedömning och betygsättning i hela landet. För att detta syfte ska uppnås krävs att lärare ges möjlighet och tid att följa provens bedömningsanvisningar. Om inte nationella prov används för kalibrering vid bedömningen, kan likvärdigheten i bedömning av proven stå på spel.

Betyg är idag, förutom att vara ett kvitto på den enskilda elevens kunskaper, ett sätt att utvärdera hur lärare och skolor, kommuner och nationer lyckas i sina utbildningsuppdrag. Betygen har kommit att spela en allt viktigare roll i den allmänna debatten om skolan. I Sverige har denna debatt intensifierats ytterligare av delvis sjunkande resultat i internationella jämförelser som PISA, TIMSS och PIRLS samt av diskussioner om den ökande segregationen i samhället och skolan (se t ex Skolverket, 2018; Östh et al., 2013). Bedömnings- och betygs likvärdighet har ifrågasatts och undersökts (se t ex Skolinspektionen, 2013, 2018; Skolverket, 2019). Ofta har lärare gjorts ansvariga för bristande likvärdighet och därmed har pressen på den enskilde läraren ökat (Mickwitz, 2011, 2015).

För att motverka de sjunkande resultaten och den bristande likvärdigheten har en rad förändringar genomförts under de senaste tio åren. Nya läroplaner med nya kursplaner och en ny betygs-skala introducerades 2011 i avsikt att stärka kunskapsuppdraget och tydliggöra utbildningsmålen. Vidare har antalet nationella prov utökats och Skolverket har publicerat kommentarmaterial, bedömningsstöd och bedömarträning i olika ämnen för att ytterligare bistå lärare och främja en likvärdig tolkning av styrdokumenten (skolverket.se).

En fråga som kan ställas är om detta har påverkat lärares bedömningsarbete. För att närmare granska ett antal lärares arbete med bedömning och betygsättning undersöktes i den här beskrivna studien hur den muntliga språkfärdigheten i ämnet engelska prövas och bedöms i det nationella provet för årskurs 9 i grundskolan. Undersökningen gjordes i nära anslutning till införandet av den nya läroplanen, *Lgr 11* med tillhörande kursplaner. Det inledande citatet i texten återger ett utdrag ur ett samtal mellan informanter i studien.

Hur lärare uppfattar att muntlig språkfärdighet beskrivs och definieras i styrdokumentet och hur skrivningarna i dessa dokument påverkar deras praktik när det gäller bedömning och betyg-sättning var ämnet för studien. Forskningsfrågorna löd:

1. Hur organiserar, bedömer och betygsätter lärare den muntliga språkfärdigheten i det nationella provet i engelska?
2. Hur uppfattar lärare muntlig språkfärdighet?
3. Vad påverkar lärares bedömning av muntlig språkfärdighet?

## Det muntliga delprovet

Det nationella provet i engelska för årskurs 9 innehåller tre delprov, där delprov A utgör den muntliga delen. I detta delprov visar eleverna prov på muntlig produktion och muntlig interaktion genom att samtala med varandra. I enlighet med instruktionerna till provet delas eleverna in i par som kan förväntas fungera väl tillsammans. En lärare sitter med vid samtalet och fungerar som provledare, dvs. ger instruktioner och stödjer eleverna vid behov. Provledarens roll vid genomförandet är viktig, även om hen bör hålla sig i bakgrunden; stöd ska ges så att var och en får chans att visa så mycket som möjligt av sin förmåga.

I förberedelser inför provtillfället ingår, enligt lärarinformationen, att läraren sätter sig in i prov- och bedömningsunderlaget. Elevmaterialet såväl som inspelade exempelsamtal med tillhörande bedömningskommentarer bör studeras.

Inspelning av provsamtalen rekommenderas starkt i provens lärarinformation. Utifrån likvärdighetsaspekter och för att lärare i efterhand ska kunna kontrollera eller göra sin bedömning är detta viktigt. Inspelningen möjliggör också samråd med kollegor om bedömningen även av de muntliga elevprestationerna.

## Muntlighet och bedömning

Att tala och interagera på ett främmande språk kräver planering, tänkande, återkallande från minnet vad man lärt, anpassning till talsituation och samtalspartner samtidigt som man formulerar och till slut verbalt yttrar vad man har för avsikt att säga. Detta är en komplicerad process. "Learners must simultaneously attend

to content, morphosyntax and lexis, discourse and information structuring, and the sound system and prosody, as well as appropriate register and pragmalinguistic features” (Hinkel, 2006, p. 114). Dessa krav gör att talat språk tenderar att ha en vagare och mer allmän vokabulär och en enklare grammatisk struktur än skrivet språk (Luoma, 2004).

Vidare handlar muntlig språkfärdighet om att kommunicera med en eller flera samtalspartner. Att endast föra mer eller mindre parallella monologer i ett samtal anses inte visa prov på kommunikativ förmåga. Galaczi (2008) visar i en studie hur deltagarna i en dialog kan komma att gestalta samtalet på olika sätt. De fyra möjligheter som anges är följande:

- att tala parallellt med varandra utan att utbyta så mycket information
- att jämlikt utbyta åsikter och idéer
- att föra ett asymmetriskt samtal, där den ena parten dominerar och styr interaktionen
- att under samtalets gång växla mellan de olika formerna av samtal

Interaktionen med samtalspartnern är således också en del av den kommunikativa förmågan.

Att pröva talat språk innebär flera olika svårigheter. Den metod som här undersöks och som används i de nationella proven utgår från ett så kallat *performance test*, eller autentiskt prov, där eleverna samtalar med varandra i en så naturlig samtalsituation som möjligt. Tidigare forskning har visat att samtalsparterna i sådana dialoger eller gruppsamtal påverkar varandra på olika sätt, vilket därmed inverkar på de individuella talprestationerna (Brown, 2003; Davis, 2009). Elevers personliga läggning, om de till exempel är extroverta och självsäkra eller inte, har visat sig spela roll både för den egna prestationen och för det gemensamt uppbyggda samtalet (Nakatsuhara, 2011; Ockey, 2009). Hur väl eleverna känner varandra kan naturligtvis också vara av betydelse (Ikeda, 1998). Även samtalsparternas respektive färdighetsnivå påverkar det gemensamma samtalet (Gan, 2010). Tidigare forskning har också visat att bedömare av dessa skäl ibland finner det svårt att

fastställa vem som presterat vad i de gemensamt konstruerade konversationerna (Brooks, 2009).

Komplexiteten i bedömning av muntlig språkfärdighet ger i sig alltså upphov till möjliga skillnader i betygsättningen (Brown, 2000). Skillnader finns även mellan olika bedömare, med olika sätt att tänka, agera och självständigt fatta beslut (Borg, 2003). Provinstruktioner och bedömningskriterier tolkas olika av olika individer och ger variation i bedömningen (Brown et al., 2005; Orr, 2002). Forskning har visat att mer erfarna bedömare ibland tenderar att fokusera på ett mindre antal bedömningsfaktorer, ofta holistiska och intuitiva, och inte alltid på samtliga föreskrivna faktorer (Bejar, 2012; Joe, Harmes & Dickerson, 2011). En lärares individuella pedagogiska grundsyn och syn på undervisningsämnet påverkar också såväl undervisning, bedömning som betygsättning (Black & Wiliam, 1998; McMillan, 2003).

En annan faktor som påverkar lärares bedömning är det lärarkollektiv som bedömaren är en del av. Denna *community* kan vara utvecklande och stärkande (Davison, 2004; Wiliam, 2007) eller hindrande och begränsande, till exempel så att den ger upphov till grupptänkande, *group think* (Hargreaves, 1994) eller verka för att bevara lokalt etablerade tolkningar och traditioner.

I fallet med de svenska nationella proven finns bedömaren ofta med som provledare vid själva provtillfället. Om provledaren inte bara instruerar eleverna initialt, utan också eventuellt behöver stödja dem under samtalet, kan detta skapa osäkerhet kring graden av extern inblandning i provsamtalet och därmed påverka förutsättningar för en likvärdig bedömning (Sandlund & Sundqvist, 2011).

Med ovan nämnda aspekter som bakgrund har i den här beskrivna studien undersökts hur tolv lärare provat och bedömt muntlig språkfärdighet. I studien undersöktes också hur en form av sambedömning kan påverka lärares betygsättning av elevsamtal.<sup>1</sup>

---

<sup>1</sup> Den form för så kallad sambedömning som användes var den som rekommenderades i Lärarinformationen till det nationella provet i engelska våren 2013, dvs. lärarna diskuterade sina bedömningar av elevprestationerna i relation till bedömningsanvisningarna.

## Metod

För att besvara forskningsfrågorna intervjuades tolv lärare med erfarenhet av och särskilt intresse för bedömning. Lärarna var mellan 30 och 60 år gamla, hade arbetat som lärare mellan 5 och 25 år och kom från olika typer av skolor i två olika geografiska regioner i Sverige. Både kommunala skolor och fristående skolor var representerade. För att nå lärare på skolor av olika storlek, med olika typer av huvudmän och med olika elevunderlag togs kontakt med ett stort antal rektorer. Efter en mängd förfrågningar anmälde sig slutligen tolv lärare från åtta olika skolor i fyra kommuner, relativt jämt fördelade över olika typer av skolor, att delta i studien. Deltagandet i studien bygger således på så kallad självselektion.

Samtliga tolv lärare deltog i en första intervju. Sex av dem deltog därefter i ett sambedömningsmöte med efterföljande gruppdiskussion. Vid detta tillfälle lyssnade deltagarna på två inspelade elevsamtal och bedömde dem individuellt. Efter den individuella bedömningen diskuterade lärarna resultaten av sina bedömningar med varandra. Samtliga intervjuer och gruppdiskussioner spelades in, transkriberades och analyserades.

Intervjuerna gjordes i ett semi-strukturerat format, vilket innebär tydliga frågor, men med utrymme för spontana följdfrågor och kommentarer. Då intervjuer i sig innebär en relation mellan intervjuare och informant, gällde det för intervjuaren, som själv var en erfaren språklärare, att se till att hen inte lade in sina egna uppfattningar i frågorna eller läste in sina egna idéer i intervju svaren. Analyserna gjordes därefter enligt Spradleys (1979) etnografiska metod för analys av intervjuer. Detta innebär ett sökande efter mönster och teman som visar på den individuella grundsyn som informanterna ger uttryck för i sina uttalanden.

## Resultat

**Forskningsfråga 1.** Hur organiserar, bedömer och betygsätter de intervjuade lärarna den muntliga språkfärdigheten i det nationella provet i engelska?

När det gäller den första forskningsfrågan, om lärares arbete med den muntliga delen av det nationella provet, gav intervju svaren

vid handen att de deltagande lärarna, trots olika lång tid i yrket, var väl bekanta med provet och att de organiserat, genomfört och bedömt det under flera år. Den muntliga delen av provet i engelska har funnits sedan 1998 och har haft ett liknande format över åren. Anpassningar till gällande kursplan och läroplan har dock gjorts kontinuerligt.

Lärarna i studien menade att den muntliga delen av det nationella provet är ett bra delprov med utmärkta instruktioner och bedömningsexempel. De följde en rutin för genomförandet av provet som fanns på den egna skolan och uttryckte att den fungerade relativt väl. De upplevde att den lokala rutinen i stort sett följde anvisningarna och att de kunde påverka den till viss del. Däremot ansåg de inte att de hade tillräckligt med tid för att vara två bedömare vid provtillfället eller för att be en eller flera kollegor sambedöma samtliga inspelade samtal i efterhand. Hälften av deltagarna uppgav att de spelat in elevsamtalen. Två av dem lyssnade rutinmässigt på alla samtal en andra gång, via inspelningen de gjort, innan de slutgiltigt betygsatte elevprestationerna. Sex av de deltagande lärarna lyssnade endast en gång, vid själva provtillfället, på de samtal de bedömde.

Lärarna i studien visade stor medvetenhet om komplexiteten i att bedöma den mångfacetterade muntliga språkfärdigheten. De var väl förtrogna med det komplicerade samspelet mellan två personer som samtalar och deras inbördes påverkan på varandras prestationer, liksom på den gemensamma prestationen. De gav uttryck för att de funderar mycket över hur de kan skapa optimala provsituationer för eleverna, där de tar hänsyn till de olika elevernas språkliga nivå, deras sociala kompetens och personliga läggning samt provuppgifternas natur och svårighetsgrad.

Lärarens inblandning i samtalet, som provledare och därmed instruktör och inspiratör, diskuterades också. Att stödja, men inte störa elever som är för passiva under samtalet så att de deltar mer aktivt ansågs vara en svår avvägningsfråga.

Alla lärarna i studien bedömde sina egna elevers prestationer på de muntliga proven. Endast en av tolv hade en medbedömare med vid samtalet. Ingen annan hade medbedömare eller sambedömare till alla samtal, men samtliga hade kollegor med vilka de kunde diskutera svårare fall. Detta kan jämföras med situationen

för de skriftliga delarna av provet: Enligt egen utsägo sambedömde och diskuterade 11 av 12 av de deltagande lärarna de skriftliga delarna. (Denna stora skillnad mellan bedömning av det muntliga respektive skriftliga delprovet bekräftas av statistik från de lärarenkäter som medföljer de nationella proven och som finns tillgängliga på provprojektets hemsida vid Göteborgs universitet.<sup>2</sup>)

Samtliga informanter menade att de först gjorde en holistisk bedömning av elevernas prestationer i samtalen och sedan gick in på de olika bedömningsfaktorerna för att stämma av att allt fanns med på rätt nivå. Två av informanterna gjorde den analytiska bedömningen efteråt, när de lyssnade på de inspelade samtalen, övriga menade att de gjorde båda bedömningarna under tiden som samtalet pågick.

Lärarna i studien genomförde och organiserade det muntliga delprovet i stort sett enligt anvisningarna. De var väl insatta i bedömningens komplexitet. Aspekter som olika färdighetsnivåer och elevers olika personliga egenskaper och hur de påverkar testresultaten ansåg informanterna var av betydelse, varför de tog sig tid att gruppera eleverna på bästa möjliga sätt. Övriga förberedelser prioriterades dock inte. Detta ledde till att en del av informanterna inte lyssnat på de bedömda elevexemplen i provmaterialet inför provperioden eller tagit del av bedömningen av dem. Gemensamma betygsdiskussioner eller medbedömning av de muntliga prestationerna prioriterades generellt inte, enligt lärarna. Hälften av deltagarna i undersökningen spelade inte in elevernas samtal under den aktuella provperioden, vilket omöjliggjorde medbedömning eller sambedömning<sup>3</sup> av dessa samtal. Brister i förberedelse inför bedömningen av samtalen och uteblivna kollegiala bedömningsdiskussioner kan potentiellt ge upphov till onödiga variationer i bedömning lärare emellan.

**Forskningsfråga 2.** Hur uppfattar lärarna i studien muntlig språkfärdighet?

---

<sup>2</sup> <https://www.gu.se/nationella-prov-frammande-sprak/prov-och-bedomningsstod-i-engelska/engelska-arskurs-7-9/nationellt-prov-i-engelska-for-arskurs-9#Resultat>

<sup>3</sup> Medbedömning avser en annan persons bedömning, som i efterskott jämförs med lärarens bedömning; sambedömning innebär att lärare tillsammans diskuterar sig fram till bedömningen.



När lärarna i studien definierade muntlig språkfärdighet, använde de till stor del samma formuleringar som återfinns i läroplan och kursplan. På en direkt fråga om de nya styrdokumentet<sup>4</sup> menade samtliga deltagare att de arbetat aktivt med att sätta sig in i dem i god tid före införandet och att de därför kände sig förtrogna med dem. Alla hade tillsammans med kollegor på den egna skolan och i vissa fall även med kollegor från andra skolor, beretts möjlighet att läsa in sig på och diskutera läroplanen, *Lgr11*. Samtidigt ansåg samtliga att det finns tolkningsutrymme i dokumenten, speciellt när det gäller kunskapskraven, vilka ansågs allmänt hållna. Det nationella provet menade man dock ger god vägledning i tolkningen.<sup>5</sup>

Vid en närmare analys av informanternas svar på frågor om muntlig språkfärdighet, om kunskapskraven såväl som om undervisning och lärande i muntlig kommunikation, syns skillnader i den individuella pedagogiska grundsynen. En del av lärarna tenderade att lägga mer vikt vid elevens förmåga att hålla ett samtal igång genom ett utbyte av åsikter och fraser i dialogform. Andra fokuserade mer på elevens förmåga att förmedla ett innehåll i form av egna idéer eller åsikter så att motparten förstår. Ytterligare en grupp lärare menade att ett rikt ordförråd och språklig korrekthet borgar för att kommunikationen verkligen går fram på rätt sätt. De deltagande lärarna kunde delas in i en grupp som fäste större vikt vid själva *interaktionen*, en grupp som fäste större avseende vid *innehållet* som förmedlas i interaktionen och en tredje grupp som hade större fokus på den *språkliga korrektheten*. Dessa variationer i uppfattning om muntlig språkfärdighet ledde även till att informanterna delvis tolkade skrivningarna i styrdokumentet olika och i enlighet med den egna individuella grundsynen.

Variationerna i uppfattning kom till uttryck i de bedömningar informanterna gjorde av två inspelade elevexempel. I diskussionerna efter bedömningarna betonades olika aspekter av elevprestationerna och skilda sätt att bedöma och betygsätta dem. Här kan man till viss del se de olika individernas språksyn i värderingen av elevernas prestationer i samtalen.

<sup>4</sup> Intervjuerna genomfördes under våren 2013, då *Lgr11* och en ny betygs-skala nyligen börjat gälla.

<sup>5</sup> Att stödja tolkningen av styrdokumentet var vid den här tiden, 2013, ett av de nationella provens syften.

De vid tidpunkten för intervjuerna nya styrdokumentens formuleringar genomsyrade lärarnas beskrivningar av muntlig språkfärdighet. Det tyder på att de nya styrdokumentet vid den här tiden var välkända och delvis implementerade samt att de deltagande lärarna hade eller höll på att inkorporera dem i sin praxis. Det torde även betyda att lärarna accepterat de nya styrdokumentet och fann målbeskrivningarna och kravnivåerna för ämnet adekvata. Variationerna i språksyn och därmed tolkning av styrdokumentet berodde således inte på ett avståndstagande från de skrivningar som finns i läroplan och kursplan, utan på individuella skillnader i uppfattning om muntlig språkfärdighet. Lärarna efterlyste gemensamma bedömningssamtal för att nå större samsyn. Denna samsyn gällde inte bara det muntliga delprovets bedömningsanvisningar, utan de intervjuade lärarna påpekade att det krävs en samsyn när det gäller läroplanen generellt och i detta sammanhang kursplanen i engelska speciellt, med syften, centralt innehåll och kunskapskrav. De efterlyste fler samtal runt tolkningen av styrdokumentet på den egna skolan, men var även angelägna om att föra samtal med lärare från andra skolor.

**Forskningsfråga 3.** Vad påverkar lärarnas bedömning av muntlig språkfärdighet?

Vad som påverkar en lärares prioriteringar och beslut är naturligtvis en mångfacetterad och komplex fråga. När lärarna i studien själva talade om vad de anser påverkar deras och andras bedömningar uppehöll de sig vid det tolkningsutrymme som finns i styrdokumentet. De talade om olika bedömningskulturer bland lärare på olika skolor och de talade om de elevgrupper de själva möter. De menade vidare att man blir en bättre och säkrare bedömare och betygsättare med längre erfarenhet. Tid, eller snarare brist på tid, att förbereda, diskutera med kollegor och lyssna på exempelsamtalen påverkar också bedömningen, ansåg informanterna.

I lärarnas egna utsagor kan man skönja ett antal ramfaktorer som påverkar deras arbete med bedömning och betygsättning av elevers prestationer. Styrdokument, resurser, tid och elevgruppen är klassiska ramar inom vilka skolans arbete gestaltas (Lundgren, 1972). Emellertid kan man här även tala om att en lokal tradition på den enskilda skolan bidrar till en viss bedömningskultur

och således enligt lärarna utgör ytterligare en ramfaktor som påverkar bedömning.

En samsyn lärare emellan torde vara en förutsättning för likvärdig bedömning av elevers prestationer. Denna samsyn behöver, som flera lärare påpekade i intervjuerna, ske i flera led; styrdokumentet (*the intended curriculum*) behöver tolkas för planering av undervisning och urval av övningar och uppgifter (*the implemented curriculum*), elevprestationerna i provet behöver värderas och tolkas mot kunskapskrav och bedömningsanvisningar och där efter vägas samman för ett summativt betyg på elevens kunskaper (*the attained curriculum*) (Van den Akker, 2003; Allal, 2013). Dessa tolkningar görs inom vissa ramar, som de resurser lärare har till sitt förfogande, de elevgrupper de möter i sin vardag samt den tid de rent objektivt har för sitt uppdrag (Lundgren, 1972). Den lokala kulturen på skolan är ytterligare en ram (Gustafsson, 1999), som påverkar lärare i deras tolkningar och prioriteringar i det dagliga arbetet med planering, undervisning och bedömning, vilket lärarna i den här studien själva tog upp i intervjuerna.

Lärarnas olika individuella uppfattningar av muntlig språkfärdighet i engelska speglas i deras definitioner av färdigheten och i hur de uppfattade skrivningarna i styrdokumentet, och utgjorde ytterligare en påverkan på bedömningen. Samtliga aspekter som informanterna tog upp; interaktionen, innehållet i samtalet och språkriktigheten, finns med i kursplanen och som bedömningsfaktorer i anvisningarna till den muntliga delen av provet. Någon inbördes rangordning mellan dem finns inte i styrdokumentet eller bedömningsanvisningar, vilket gjorde det möjligt för de individuella bedömnarna att vikta dem olika, trots att de använde formuleringar och begrepp direkt ur dokumentet. Detta visar att det finns rum för de här beskrivna individuella tolkningarna utifrån individernas egna uppfattningar, eftersom de olika aspekterna inte utesluter varandra. De olika preferenserna är möjliga inom styrdokumentets gränser.<sup>6</sup>

---

<sup>6</sup> Detta tolkningsutrymme kan upplevas både som en tillgång och ett problem. På vilket sätt stärks bedömarkompetensen och likvärdigheten i bedömningen på bästa sätt? Diskussionen om detta har intensifierats efter det att den här aktuella studien genomfördes och har bland annat resulterat i

I diskussionerna i de grupper som bedömde två elevexempel kom samma variation i uppfattning om muntlig språkfärdighet till uttryck. Lärarna tog fasta på olika aspekter av samtalen, värderade dem olika, men endast i ett fall skiljde sig betygssättningen av elevprestationer mellan de två grupperna. Den subjektivitet som är en del av komplexa, kvalitativa bedömningar blev tydlig.

Lärarna i studien var medvetna om variationen i bedömning och menade att en del av skillnaderna kunde bero på det faktum att de var en del av ett lärarkollektiv, dvs. att de påverkades av den tradition som etablerats bland kollegorna på den lokala skolan. De identifierade sig med den egna skolan och kollegorna och blev själva en del av en lokal praxis, som eventuellt skiljer sig från praxis på andra skolor. Denna skolkultur kan påverka såväl undervisning som bedömning, men även hur den tid som finns till förfogande utnyttjas.

När lärarna reflekterade över sin egen bedömarkarriär, menade de att erfarenheten, att ha gjort proven många gånger och att ha bedömt många elevers muntliga språkfärdighet, ger viss säkerhet och större kompetens i bedömningen, något som också visats i andra studier (Forsberg & Wermke, 2012; Borgström & Ledin, 2014). Att diskutera med kollegor, både på den egna skolan och på andra skolor, är ytterligare en faktor som påverkar bedömarkompetensen positivt, enligt deltagarna i studien. Studier har också visat att bedömaröverensstämmelsen ökar, när lärare ges möjlighet att samverka kring bedömning (Erickson, 2009; Skolverket, 2009, 2014; Skolinspektionen, 2018).

## Sammanfattande slutsatser

Sammanfattningsvis bör påpekas att lärarna i den här aktuella studien själva valde att medverka och var erfarna och intresserade bedömare av muntlig språkfärdighet. Detta gör att resultaten måste tolkas med försiktighet och inte med någon självklarhet kan betraktas som representativa för hela gruppen lärare i engelska.

De nationella proven genomfördes av lärarna i studien i stort sett enligt de anvisningar som medföljer proven. Här tycks inga

---

att Skolverket (2018) publicerade Allmänna råd, i vilka den likvärdighetsbefrämjande aspekten betonas.

stora skillnader i de förutsättningar som gavs eleverna i provsituationen ha förelegat. Men lärarna ansåg inte att de haft möjlighet att förbereda och efterarbeta proven helt i enlighet med anvisningarna på grund av tidsbrist.

Deltagarna i studien beskrev muntlig språkfärdighet med ord och begrepp från styrdokumentet, med tyngdpunkten på tre olika aspekter: innehåll, interaktion och språkriktighet. Deras uppfattning om muntlig språkfärdighet skilde sig åt men låg inom det tolkningsutrymme som styrdokumentet ger. Lärarna menade själva att de påverkades av de elevgrupper de mötte, av de tidsramar de hade samt av de bedömningstraditioner som fanns på den egna lokala skolan när de bedömde elevers prestationer på de muntliga delproven.

En faktor som vid många tillfällen togs upp av deltagarna var bristen på tid. De menade att de behövde mer tid för förberedelse av proven, för att bättre sätta sig in i bedömningsexemplen. Även tid för medbedömning efterlystes – antingen genom att två bedömare gemensamt genomför provet eller genom att medbedömare i efterhand lyssnar på inspelningar av elevprestationer. Tid för sambedömning och diskussioner om bedömning rent allmänt önskades också. Deltagarna tycktes mena att det låg utanför de egna möjligheterna att prioritera dessa aktiviteter inom de ramar de hade. De ansåg att endast skolledning eller huvudman kan göra denna prioritering. Även diskussioner mellan olika skolor skulle kunna möjliggöras och organiseras av skolledningar och huvudmän.

## **Didaktiska reflektioner och implikationer**

Lärarna i studien efterlyste mer tid till förberedelse och bedömning av den muntliga delen av det nationella provet. Vid undersökningstillfället hade det nationella provet för första gången anpassats efter *Lgr11*. Trots detta hade inte samtliga informanter i studien haft möjlighet att lyssna igenom samtliga elevexempel inför provtillfället. En årlig ”kalibrering” med kommenterade elevexempel och bedömningsanvisningar är en möjlighet att minska individuella skillnader i bedömningspraxis och att stödja likvärdigheten. Eventuellt har utrymmet för dessa förberedelser ökat något, i och med att den muntliga delen av det nationella

provet i engelska sedan läsåret 2015/2016 förlagts till höstterminen, medan övriga delar fortsatt genomförs på vårterminen.

Den av de tolv lärarna i studien som hade möjligheten att ha ytterligare en bedömare med under provgenomförandet, var i sin tur bedömare tillsammans med en annan lärare vid provtillfällena där elever som hen inte undervisade deltog. Denna typ av medbedömning är ett sätt att stödja samsynen i bedömningen på skolnivå. Detta kan dock inte avhjälpa skillnader i bedömning av muntlig språkfärdighet skolor emellan.

Bedömningsdiskussionerna på den enskilda skolan behöver, enligt lärarna i den här studien, intensifieras. Medvetenheten om olikheter i uppfattningen om den muntliga språkfärdigheten individer emellan fanns hos de deltagande. De önskade därför diskutera mera för att belysa olikheter och minimera variationen i den slutgiltiga betygssättningen. Diskussioner och kollegialt lärande på skolor, men även i övergripande sammanhang skolor emellan, ger förutsättningar för att skapa en större bedömargemenskap, vilket i sin tur kan bidra till att öka likvärdigheten i bedömning.

Likvärdigheten och rättssäkerheten torde gynnas om bedömningsdiskussioner kontinuerligt förs på varje skola, men även i lärargrupper sammansatta av lärare från olika skolor. I den studie som här redovisas önskade samtliga lärare diskutera bedömning med kollegor från andra skolor. Föreställningen att bedömartraditionerna skiljer sig åt oroar lärarna och ger en osäkerhet. Möjligheten att diskutera bedömning, betygssättning och betygskriterier med kollegor kan dämpa oro och osäkerhet samtidigt som den kan bidra till att öka bedömarkompetensen och samsynen deltagande lärare emellan, en vidare bedömargemenskap kan uppstå. En utökning av tiden för förberedelse inför nationella prov och för bedömningsdiskussioner i samband med betygssättningen av proven behöver organiseras från skollednings- och huvudmannanivå. Detta skulle stärka rättssäkerheten i de prov och bedömningar som görs.

En implikation av den här rapporterade studien är att lärares bedömarkompetens behöver stärkas. Professionella, kunniga och erfarna lärares kvalitativa bedömningar av komplexa kompetenser och förmågor behövs i den svenska skolan, inte bara för bedömning av nationella prov. Muntlig språkfärdighet är en

mångfacetterad förmåga som inte lätt kan mätas med statistisk exakthet i matriser eller med enkla prov – här behövs erfarna språklärare med pedagogiskt kunnande och bedömarkompetens. Svensk skola behöver ställa erforderliga resurser, bland annat i form av tid, till förfogande för att minimera variationen i bedömningen och därmed öka likvärdigheten i utbildningen.

Denna studie visar att lång vana att genomföra de muntliga proven inte automatiskt stärker likvärdigheten i bedömning. Det räcker inte heller med att vara medveten om provformens och förmågans komplexitet. En medvetenhet om den egna bedömarkompetensens gränser och utvecklingspotential är en förutsättning för att variationen i de subjektiva bedömningarna av komplexa förmågor ska kunna minimeras och likvärdigheten i betygssättningen öka. På den enskilda skolan behöver därför betygssättningen vara ständigt pågående. De nationella proven skulle kunna ses som ett årligt avstämmningstillfälle, då ämnesgruppen tillsammans prövar sin bedömarkompetens mot de bedömda elevexempel som ställs till förfogande. En blandning av provgrupper, så att lärare är medbedömare vid samtal mellan elever som de inte undervisar, kan bidra till ökad likvärdighet i bedömningen. Diskussioner om bedömningen av samtalen efter själva provtillfället kan ytterligare bidra till att fördjupa erfarenheten och öka kompetensen att göra kvalitativa bedömningar och sätta likvärdiga betyg. Det är den enskilde lärarens ansvar att sträva efter att göra en likvärdig värdering av elevers prestationer, men det är en skollednings ansvar att ge lärarna möjligheten att utföra sitt uppdrag på bästa möjliga sätt.

## Referenser

- Allal, L. (2013). Teachers' professional judgement in assessment: A cognitive act and a socially situated practice. *Assessment in Education: Principles, Policy & Practice*, 20, 20–34. <https://doi.org/10.1080/0969594X.2012.736364>
- Bejar, I. I. (2012). Rater cognition: Implications for validity. *Educational Measurement: Issues and Practice*, 31, 2–9. <https://doi.org/10.1111/lj.1745-3992.2012.00238.x>
- Black, P. & Wiliam, D. (1998). *Inside the black box: Raising standards through classroom assessment*. Phi Delta Kappan. <https://doi.org/10.1177/003172171009200119>

- Borg, S. (2003). Teacher cognition in language teaching: A review of research on what language teachers think, know, believe, and do. *Language Teaching*, 36, 81–109. <https://doi.org/10.1017/S0261444803001903>
- Borgström, E. & Ledin, P. (2014). Bedömarvariation – Balansen mellan teknisk och hermenutisk rationalitet vid bedömning av skrivprov. *Språk & Stil*, NF 24, 133–165. <https://www.diva-portal.org/smash/get/diva2:800920/FULLTEXT01.pdf>
- Brooks, L. (2009). Interacting in pairs in a test of oral proficiency: Co-constructing a better performance. *Language Testing*, 26, 341–366. <https://doi.org/10.1177/0265532209104666>
- Brown, A. (2000). An investigation of the rating process in the IELTS oral interview. *IELTS research reports*, 3, 49–84. <https://search.informit.com.au/documentSummary;dn=905752862811172;res=IELHSS>
- Brown, A. (2003). Interviewer variation and the co-construction of speaking proficiency. *Language Testing*, 20, 1–25. <https://doi.org/10.1191/0265532203lt2420a>
- Brown, A., Iwashita, N., & McNamara, T. (2005). An examination of rater orientations and test-taker performance on English-for-Academic-Purposes speaking tasks. *Research Report – Educational Testing Service*, Princeton RR, 5. <https://doi.org/10.1002/j.2333-8504.2005.tb01982.x>
- Davis, L. (2009). The influence of interlocutor proficiency in a paired oral assessment. *Language Testing*, 26, 367–396. <https://doi.org/10.1177/0265532209104667>
- Davison, C. (2004). The contradictory culture of teacher-based assessment: ESL teacher assessment practices in Australian and Hong Kong secondary schools. *Language Testing*, 21(3), 305–334.
- Erickson, G. (2009). *Nationella prov i engelska – en studie av bedömarsamstämmighet*. <https://www.gu.se/nationella-prov-frammande-sprak/rapporter-och-skrifter>
- Forsberg, E., & Wermke, W. (2012). Knowledge sources and autonomy: German and Swedish teachers' continuing professional development of assessment knowledge. *Professional Development in Education*, 38, 741–758. <https://doi.org/10.1080/19415257.2012.694369>



- Galaczi, E. D. (2008). Peer-Peer Interaction in a Speaking Test: The Case of the "First Certificate in English" Examination. *Language Assessment Quarterly*, 5, 89–119. <https://doi.org/10.1080/15434300801934702>
- Gan, Z. (2010). Interaction in group oral assessment: A case study of higher- and lower-scoring students. *Language Testing*, 27, 585–602. <https://doi.org/10.1177/0265532210364049>
- Gustafsson, C. (1999). Ramfaktorer och pedagogiskt utvecklingsarbete. *Pedagogisk forskning i Sverige*, 4, 43. <https://open.lnu.se/index.php/PFS/issue/view/83>
- Hargreaves, A. (1994). *Changing teachers, changing times: Teachers' work and culture in the postmodern age*. Cassell.
- Hinkel, E. (2006). Current perspectives on teaching the four skills. *TESOL Quarterly*, 40, 109–131. <https://doi.org/10.2307/40264513>
- Ikedá, K. (1998). The paired learner interview: A preliminary investigation applying Vygotskian insights. *Language Culture and Curriculum*, 11, 71–96. <https://doi.org/10.1080/07908319808666542>
- Joe, J. N., Harnes, J. C., & Hickerson, C. A. (2011). Using verbal reports to explore rater perceptual processes in scoring: A mixed methods application to oral communication assessment. *Assessment in Education: Principles, Policy & Practice*, 18, 239–258. <https://doi.org/10.1080/0969594X.2011.577408>
- Lundgren, U. P. (1972). *Frame factors and the teaching process: a contribution to curriculum theory and theory on teaching*. Almqvist & Wiksell. <http://hdl.handle.net/2077/12809>
- Luoma, S. (2004). *Assessing speaking*. Cambridge University Press.
- McMillan, J. H. (2003). Understanding and improving teachers' classroom assessment decisionmaking: Implications for theory and practice. *Educational Measurement: Issues and Practice*, 22, 34–43. <https://doi.org/10.1111/j.1745-3992.2003.tb00142.x>
- Mickwitz, L. (2011). *Rätt betyg för vem?: Betygsättning som institutionaliserad praktik*. [Licentiatuppsats, Stockholms universitet]. <https://www.diva-portal.org/smash/get/diva2:396277/FULLTEXT03>
- Mickwitz, L. (2015). Den professionella lärarens möjlighetsvillkor. *Utbildning och demokrati*, 24, 33–54. <https://www.oru.se/global>

assets/oru-sv/forskning/forskningsmiljoer/hs/humus/utbildning-och-demokrati/2015/nr-2/larissa-mickwitz---den-professionella-lararens-mojlighetsvillkor.pdf

- Nakatsuhara, F. (2011). Effects of test-taker characteristics and the number of participants in group oral tests. *Language Testing*, 28, 483–508. <https://doi.org/10.1177/0265532211398110>
- Ockey, G. J. (2009). The effects of group members' personalities on a test taker's L2 group oral discussion test scores. *Language Testing*, 26, 161–186. <https://doi.org/10.1177/0265532208101005>
- Orr, M. (2002). The FCE speaking test: Using rater reports to help interpret test scores. *System*, 30, 143–154. [https://doi.org/10.1016/S0346-251X\(02\)00002-7](https://doi.org/10.1016/S0346-251X(02)00002-7)
- Sandlund, E., & Sundqvist, P. (2011). Managing task-related trouble in L2 oral proficiency tests: Contrasting interaction data and rater assessment. *Novitas-ROYAL*, 5, 91–120. <http://www.acarindex.com/dosyalar/makale/acarindex-1423909825.pdf>
- Skolinspektionen. (2013). Olikheterna är för stora. Omrättning av nationella prov i grundskolan och gymnasieskolan, 2013. *Redovisning av regeringsuppdrag Dnr U2011/6544/GV*. <https://docplayer.se/31275321-Olikheterna-ar-for-stora.html>
- Skolinspektionen. (2018). Ombedomning av nationella prov 2017. Fortsatt stora skillnader. *Redovisning av regeringsuppdrag Dnr U2014/7535/GV, U2017/05038/GV*. <https://skolinspektionen.se/globalassets/02-beslut-rapporter-stat/ombedomning-nationella-prov/ombedomning-nationella-prov-2018.pdf>
- Skolverket (2009). *Redovisning av regeringsuppdrag att ge förslag på hur det nationella provsystemet bör utvecklas och utformas Dnr U2009/1671/S*. <https://www.skolverket.se/publikationer?id=2200>
- Skolverket (2014). *Sambedomning i skolan – exempel och forskning*. <https://www.skolverket.se/publikationer?id=3172>
- Skolverket (2018). *Allmänna råd om betyg och betygssättning* <https://www.skolverket.se/publikationer?id=4000>
- Skolverket (2018). *Analyser av familjebakgrundens betydelse för skolresultaten och skillnader mellan skolor*. En kvantitativ studie av

- utvecklingen över tid i slutet av grundskolan. <https://www.skolverket.se/publikationer?id=3927>
- Skolverket (2019). *Analys av likvärdig betygssättning mellan elevgrupper och skolor*. Jämförelser mellan betyg och nationella prov i årskurs 9. <https://www.skolverket.se/publikationer?id=4035>
- Spradley, J. P. (1979). *The ethnographic interview*. Holt, Rinehart and Winston.
- Van den Akker, J., Kuiper, W., & Hameyer, U. (2003). *Curriculum perspectives: An introduction*. Kluwer Academic Publishers.
- William, D. (2007). Content then process: Teacher learning communities in the service of formative assessment. In D. B. Reeves (Red.), *Ahead of the curve: The power of assessment to transform teaching and learning* (s. 183–204). Solution Tree.
- Östh, J., Andersson, E., & Malmberg, B. (2013). School choice and increasing performance difference: A counterfactual approach. *Urban Studies*, 50, 407–425. <https://doi.org/10.1177/0042098012452322>