

# Bedömarperspektiv på muntliga nationella prov i engelska

Linda Borger

Den muntliga delen av det nationella provet i engelska uppskattas av lärare, som anser att provet ger ett bra stöd vid betygssättning. Men det är komplext att bedöma muntlig språkfärdighet och muntliga prov kräver även tid och resurser för det praktiska genomförandet. Detta aktualiserar frågor om bedömarsamstämmighet, likvärdighet och praktiska konsekvenser, vilket blev utgångspunkten för mitt avhandlingsprojekt<sup>1</sup>. I denna artikel ger jag en bakgrund till avhandlingen och berättar om några av huvudresultaten.

## Bakgrund

I mitt avhandlingsprojekt har jag undersökt det muntliga nationella delprovet i engelska på gymnasiet utifrån ett bedömningsperspektiv. Jag har en bakgrund som gymnasielärare i engelska och svenska och tyckte att det var intressant att fördjupa mig i just det muntliga delprovet. Bedömning av muntlig språkfärdighet är ett relativt underutforskat område och det finns få studier som undersöker storskaliga muntliga språkprov i en skolkontext. I Sverige har vi haft ett obligatoriskt muntligt delprov i engelska med en elevinteraktiv provmodell sedan 1998 i grundskolan (åk 9) och 2000 i gymnasiet. Detta är tämligen unikt i ett europeiskt perspektiv där många länder antingen inte prövar muntlig förmåga alls i sina nationella prov, eftersom det är komplext och tidskrävande, eller har en lösning där de muntliga proven utformas och genomförs på lokal nivå (Europeiska kommissionen/EACEA/Eurydice, 2015).

Huvudsyftet med de nationella proven är att vara ett stöd för lärare att göra likvärdiga bedömningar och därmed bidra till en rättvis betygssättning av en elevs kunskaper. Skolinspektionens omdömmingar av nationella prov (se

t.ex. Skolinspektionen, 2018) har dock visat att samstämmigheten brister för de delprov som innehåller längre elevproducerade svar, till exempel uppsatsdelar.<sup>2</sup> Detta har lett till en debatt om nationella prov och bedömarvariationens betydelse för en rättvis och likvärdig bedömning. Men de muntliga delarna av de nationella proven har inte undersökts, eftersom det inte är obligatoriskt att spela in och dokumentera dem.

## Det muntliga provformatet

Den muntliga delen av de engelska nationella proven prövar muntlig produktion och interaktion i enlighet med kurs- och ämnesplaner och genomförs som ett samtal mellan två elever, där de får uttrycka, utveckla och diskutera ett innehåll utifrån ett givet tema. Provet är ett exempel på ett så kallat *performance-prov* som innehåller uppgifter som ska leda till en så autentisk målspråksanvändning som möjligt (McNamara, 1996). Bedömning av autentiska provuppgifter är komplex och förutsätter tolkning i relation till bedömningskriterierna, vilket kan leda till bedömareffekter. Det finns olika sorters bedömareffekter och bland de vanligaste är att bedömare har olika stränghetsgrad. Exempel på andra faktorer som kan påverka bedömningen i autentiska

prov är att bedömare tolkar eller viktar bedömningskriterierna på olika sätt.

Forskning har visat att par- och gruppsamtal genererar ett mer autentiskt interaktionsmönster än intervjuformatet som genomförs med en provtagare och en intervjuare/examinator. Men det finns utmaningar med par- och gruppsamtal ur bedömningssynpunkt. Till exempel kan olika bakgrundsfaktorer hos provtagarna såsom personlighet, kön och språklig nivå inverka på samtalspartnerns prestation (O’Sullivan, 2002). De studier som har gjorts visar dock inga entydiga svar i frågan om eller hur dessa så kallade *interlocutor effects* påverkar bedömningen. En ytterligare utmaning vid par- och gruppsamtal är att individuella betyg sätts på en gemensamt skapad prestation, så kallad *co-constructed interaction*, varför det kan vara svårt att fullt ut separera prestationerna.

### Material, deltagare och genomförande

Avhandlingen innehåller tre studier där den gemensamma nämnaren är bedömarperspektivet. De två första studierna bygger på ett material av sex ljudinspelade elevsamtal, det vill säga tolv individuella prestationer, som kommer från en utprovning i kursen Engelska 6 i gymnasiet. Samtalen hade valts ut för att spegla olika färdighetsnivåer och bedömdes av 17 gymnasielärare i engelska (fyra män och 13 kvinnor) från elva olika skolor, både kommunala och fristående skolor. De svenska lärarna använde de nationella kunskapskraven och de analytiska bedömningsfaktorer som bifogas i lärarinstruktionerna till provet. Förutom de svenska lärarna deltog även 14 internationella bedömare, sju från Finland och sju från Spanien.<sup>3</sup> De bedömde samma sex elevsamtal men utifrån referensnivåerna

i Gemensam europeisk referensram för språk (GERS) (Skolverket, 2009). Våra nationella kurs- och ämnesplaner i språk är som bekant jämförda med språknivåerna i GERS, framför allt med hjälp av textuella jämförelser, men det finns få studier som empiriskt har validerat detta samband.

Data till de två första studierna samlades in under endagsseminarier som hölls med de olika

bedömargrupperna där de lyssnade på de sex samtalen och bedömde dem individuellt. Både de svenska lärarna och de internationella bedömarna gav ett helhetsbetyg för varje elevprestation och skrev ett sammanfattande omdöme, där de motiverade betyget och förklarade vilka aspekter av den muntliga förmågan som de särskilt tog hänsyn till i bedömningen. Den tredje studien är en enkätstudie där jag gick ut till 150 gymnasieskolor med en

enkät som bland annat handlar om det praktiska genomförandet av det muntliga delprovet. Enkäten besvarades av 267 gymnasielärare i engelska. Av dessa var 75 % kvinnor och medelåldern var 47 år (med en spridning på 26-68 år). Undervisningserfarenheten var i genomsnitt 16 år (1-42 år) och en stor majoritet av lärarna i urvalet (96 %) uppgav att de hade lärarlegitimation i engelska. Urvalet var representativt i förhållande till geografisk spridning och fördelning av kommunala och fristående skolor.

### Några huvudresultat

Avhandlingen kom fram till att det muntliga delprovet ger möjlighet att pröva elevers kommunikativa muntliga förmåga på ett brett och representativt sätt i förhållande till kurs- och ämnesplanen, men det går att utveckla arbetet

*” Det finns olika sorters bedömareffekter och bland de vanligaste är att bedömare har olika stränghetsgrad. Exempel på andra faktorer som kan påverka bedömningen i autentiska prov är att bedömare tolkar eller viktar bedömningskriterierna på olika sätt. ”*

med bedömersamstämmighet och det praktiska genomförandet. I det följande återges och diskuteras några av huvudresultaten.

### Aspekter av muntlig förmåga

För att systematisera analysen av de svenska lärarnas och GERS-bedömarens skriftliga kommentarer till elevprestationerna utvecklades ett kodningsschema som bygger på beskrivningen av *kommunikativ språklig kompetens* i GERS. Enligt GERS består den kommunikativa kompetensen av tre komponenter: en *lingvistisk* (rent språkliga aspekter såsom ordförråd, grammatik och uttal), en *sociolingvistisk* (förmåga att anpassa språket till sociala konventioner och normer) och en *pragmatisk* kompetens (den funktionella användningen av språkliga resurser, t.ex. att organisera och strukturera ett innehåll, att tala med flyt och att anpassa det man säger till situation och mottagare). I GERS utgör även språkliga strategier en viktig del av kommunikation och inläring. I muntlig produktion och interaktion är framför allt *interaktionsstrategier* (t.ex. användandet av turtagningsmarkörer och förmågan att knyta an sitt bidrag till den andra talarens bidrag) och *produktionsstrategier* (t.ex. omformuleringar och korrigeringar av språkliga fel) framträdande.

Innehållsanalysen visade att bedömarna tog hänsyn till en stor bredd av aspekter som ingår i den kommunikativa muntliga språkkompetensen, vilket också är i linje med beskrivningen i kurs- och ämnesplanerna i engelska. De lingvistiska och pragmatiska aspekterna och elevernas interaktionsstrategier var mest framträdande. Däremot kommenterade bedömarna elevernas sociolingvistiska kompetens i mindre utsträck-

ning. Detta kan ha att göra med att provet inte specifikt möjliggör att visa upp förmågan att använda sociokulturella aspekter av språkanvändning, såsom språkliga markörer i sociala relationer och artighetskonventioner. Liksom i tidigare studier inom muntlig bedömning utgjorde de lingvistiska aspekterna av elevernas kommunikativa förmåga den största gruppen av kommentarerna, även om pragmatiska aspekter och interaktions- och produktionsstrategier också kommenterades flitigt. Detta betyder nödvändigtvis inte att bedömarna lade större vikt vid lingvistiska aspekter än andra delar av den kommunikativa förmågan men det pekar på att de har en framträdande roll. Tidigare forskning har också visat att det är vanligt att de lingvistis-

**” Liksom i tidigare studier inom muntlig bedömning utgjorde de lingvistiska aspekterna av elevernas kommunikativa förmåga den största gruppen av kommentarerna, även om pragmatiska aspekter och interaktions- och produktionsstrategier också kommenterades flitigt. ”**

ka aspekterna får större tyngd i bedömningen, eftersom de kanske är lättare att ’kvantifiera’ än andra aspekter av kommunikativ kompetens. Detta är alltså något som bedömarna bör vara medvetna om så att man uppmärksammar en så stor bredd av elevernas muntliga prestationer som möjligt.

Det fanns många exempel i bedömarkommentarerna på hur eleverna använde *interaktionsstrategier*, och i delstudie två

identifierade jag tre huvudkategorier:

- *strategier för att utveckla samtalets innehåll* – genom att till exempel fördjupa diskussionen och bygga vidare på partnerns samtalsämnen; och att ställa frågor och följdfrågor som för samtalets innehåll vidare
- *interaktiva lyssnandestrategier* – att lyssna aktivt på partnern genom att ge verbal bekräftelse och uppbackningar, och genom att förtydliga och omformulera vid behov
- *turtagningsstrategier* – användandet av turtagningsmarkörer (att ta initiativ, inleda och hålla igång samtalet) och hur naturligt flödet i

samtalet var, vilket på engelska benämns som 'conversational fluency'.

Bedömarna upplevde det positivt när dessa strategier användes på ett kollaborativt och ömsesidigt sätt, då eleven aktivt avlyssnade och svarade på den andra talarens bidrag. Det uppfattades således mer negativt om eleven saknade strategier för att föra samtalet framåt, genom att till exempel inte bygga vidare på eller bekräfta det partnern sa. Som framgår är interaktionskompetensen komplex att bedöma eftersom den skapas tillsammans av provtagarna i provsituationen. Särskilt svårbedömt visade sig ett samtal vara där den ena eleven tog mer plats än den andra. En del bedömare menade att den elev som tog mindre plats hjälptes av den dominanta eleven medan andra tyckte att asymmetrin påverkade den passiva elevens prestation negativt. Att asymmetrisk interaktion är speciellt svår ur ett bedömningsperspektiv har bekräftats av tidigare studier (se t.ex. May, 2011) och det är alltså en fråga som kräver speciell uppmärksamhet, inte minst i bedömningsanvisningarna.

### **Bedömarsamstämmighet**

Som nämnts ovan medför det elevinteraktiva provformatet en viss grad av oförutsägbarhet och resultaten visade också på en viss bedömarvariation. Det fanns en tydlig skillnad i stränghet när jag tittade på de svenska lärarnas bedömningar, vilket också framkommer i internationell forskning kring samma tema. I synnerhet var skillnaden stor mellan enstaka mycket stränga respektive milda bedömningar men på det stora hela var bedömarna tämligen överens och den övergripande reliabiliteten var god, även om det fanns utrymme för förbättring. Jag gjorde även en tentativ jämförelse av några elevprestationer

med lite större bedömarvariation och fann att det kunde förklaras av att bedömarna antingen uppmärksammade samma aspekter av elevens muntliga förmåga men värderade dem olika – som mer positivt eller negativt – eller att de lade tyngdpunkt på delvis olika aspekter. Detta kan bero på att bedömarna tolkar bedömningskriterierna (dvs. kunskapskraven i den svenska kontexten) på olika sätt i förhållande till elevprestationerna eller att de lägger olika stor vikt vid olika aspekter i den holistiska bedömningen. Sammantaget pekar resultaten på att det är viktigt att arbeta med sätt att stärka både samstämmigheten i lärarnas bedömningar och samsynen kring tolkning av bedömningskriterierna.

### **Det praktiska genomförandet**

Vad gäller genomförande av de muntliga delproven beskrivs på Skolverkets hemsida att den organisation som är lämpligast "kan se olika ut på olika skolor", och det poängteras att de muntliga proven är "hela skolans angelägenhet".<sup>4</sup> Det framgår också att det ytterst är "rektorn som ansvarar för att organisationen gynnar både elever och lärare". I enkäten ville jag därför undersöka hur det praktiska genomförandet av det muntliga delprovet gick till. Det framkom att det var vanligast att genomföra de muntliga proven på lektionstid, vilket av många upplevdes som stressigt och tidskrävande, inte minst eftersom läraren enligt Skolverkets instruktioner bör säkerställa att eleverna får undervisning även under den tid då provet genomförs. Lärare som arbetade på skolor där proven organiserades mer centralt, som en schemabrytande aktivitet, uttryckte att de var nöjda med den lösningen. En annan stressfaktor i samband med det muntliga provet var att hitta lokaler som passar. Hälften av lärarna i enkäten uppgav att det inte fanns tillräckligt med lokaler

*” Som framgår är interaktionskompetensen komplex att bedöma eftersom den skapas tillsammans av provtagarna i provsituationen. Särskilt svårbedömt visade sig ett samtal vara där den ena eleven tog mer plats än den andra.”*

för att genomföra provet på ett effektivt sätt. Här handlar det bland annat om att det krävs flera rum samtidigt eftersom eleverna har femton minuters förberedelsestid som ofta utförs parallellt med att ett annat elevpar genomför provet. Upplägget med förberedelsestid kan därför behövas över, dels i förhållande till praktiska konsekvenser och dels i förhållande till hur den påverkar autenticiteten i elevernas tal. En ytterligare aspekt av provets genomförande är inspelning, vilket rekommenderas i lärarinstruktionerna eftersom det möjliggör sam- och medbedömning och att man kan lyssna på samtalen igen. Närmare hälften av lärarna i urvalet uppgav att de spelade in de muntliga delproven, medan ungefär 40 % inte gjorde det. Det var framförallt brist på tid, både vad gäller att lyssna igenom samtal igen efteråt och att sambedöma med kollegor, som lyftes fram som den främsta anledningen till att inte spela in.

I lärarinstruktionerna rekommenderas att man ska arbeta med sambedömning, vilket innebär att lärare tillsammans diskuterar och bedömer elevprestationer utifrån bedömningsanvisningarna. Ett annat tillvägagångssätt som nämns är medbedömning då två eller fler lärare oberoende av varandra bedömer samma elevprestation och sedan diskuterar bedömningen. Lärarna som besvarade enkäten var generellt positiva till sam- och medbedömning, men resultaten visar att 42 % av lärarna bedömde de muntliga proven i engelska på egen hand, främst på grund av tidsbrist och hög arbetsbelastning. Så studien visar att det finns ett behov av att ge lärare bättre förutsättningar att vidareutveckla arbetet med sam- och medbedömning, speciellt i det muntliga delprovet där detta troligtvis tar mer tid att organisera än för den skriftliga uppsatsdelen. Dessutom framkom

i enkäten att 62 % av lärarna upplevde att de inte fick tillräckligt stöd från skolledningen för att organisera och genomföra de muntliga proven, vilket pekar på ett behov av mer administrativt stöd.

### Slutkommentar

Om man ser till det muntliga nationella delprovets styrkor och svagheter framkommer två tydliga slutsatser i avhandlingen. För det första möjliggör det muntliga provet att pröva elevers kommunikativa muntliga förmåga på ett brett och representativt sätt. Den tydliga kopplingen mellan det muntliga delprovet och det fokus på muntlig produktion och interaktion som finns i

ämnes- och kursplaner stärker validiteten. Svårigheten ligger dock i att bedöma muntlig förmåga på ett reliabelt sätt, det vill säga med så god samstämmighet som möjligt. En konsekvens är därför att det när man har öppna provuppgifter som att tala och skriva krävs att det tillsätts extra resurser för att arbeta med sam- och medbedömning, då elevprestationer diskuteras i förhållande till kunskapskrav, exempel och övrigt bedömningsmaterial.

Tidigare forskning har tydligt visat att bedömningsansvar och reliabilitet ökar då fler än en bedömare är involverad. Sam- och medbedömning förstärker en gemensam syn på kvaliteter i elevernas prestationer och med hjälp av flera ögon kan man bli uppmärksam på aspekter som man kanske inte hade sett annars (Erickson, 2018). Avhandlingen belyser även det decentraliserade ansvaret för genomförandet av de nationella proven där skolledningen har en central roll. I enlighet med detta är det viktigt att organisationen av de muntliga delproven blir en angelägenhet för hela skolan och inte lämnas åt enskilda lärare att ta hand om.

*” Så studien visar att det finns ett behov av att ge lärare bättre förutsättningar att vidareutveckla arbetet med sam- och medbedömning, speciellt i det muntliga delprovet där detta troligtvis tar mer tid att organisera än för den skriftliga uppsatsdelen.”*

## Referenser

Erickson, G. (2009). *Nationella prov i engelska – en studie av bedömersamstämmighet*. Hämtad från [https://nafsgu.se/digitalAssets/1319/1319572\\_bed\\_np-eng-g.erickson-2009.pdf](https://nafsgu.se/digitalAssets/1319/1319572_bed_np-eng-g.erickson-2009.pdf)

Erickson, G. (2018). *Extern bedömning av kunskap eller Om värdet av att se med andras ögon*. Artikel inom ramen för Skolverkets webbkurs om likvärdighet. Hämtad från <https://www.skolverket.se/skolutveckling/kompetensutveckling/artiklar-till-webbkurser-om-bedomning-och-betygsattning>

Europeiska kommissionen/EACEA/Eurydice. (2015). *Languages in Secondary Education: An Overview of National Tests in Europe – 2014/15*. Eurydice Report. Luxemburg: Publications Office of the European Union.

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*, 25(1), 69–87.

May, L. (2011). Interactional Competence in a Paired Speaking Test: Features Salient to Raters. *Language Assessment Quarterly*, 8(2), 127–145.

McNamara, T. (1996). *Measuring Second Language Performance*. London: Longman.

O’Sullivan, B. (2002). Learner acquaintanceship and oral proficiency test pair-task performance. *Language Testing*, 19(3), 277–295.

Skolinspektionen. (2018). *Ombedömning av nationella prov 2017 – Fortsatt stora skillnader*. Stockholm: Skolinspektionen.

Skolverket. (2009). *Gemensam europeisk referensram för språk: lärande, undervisning och bedömning*. Stockholm: Skolverket.

## Fotnoter

<sup>1</sup> Borger, L. (2018). *Investigating and Validating Spoken Interactional Competence: Rater Perspectives on a Swedish National Test of English* (Doktorsavhandling, Gothenburg Studies in Educational Sciences, 424). Göteborg: Acta Universitatis Gothoburgensis. Hämtad från <http://hdl.handle.net/2077/57946>

<sup>2</sup> Det ska dock nämnas att den metod som Skolinspektionen använt har kritiserats av bland annat Gustafsson & Erickson (2013) och det finns andra studier som har visat på mera positiva resultat i engelska (Erickson, 2009).

<sup>3</sup> För mer information om resultaten av GERS-bedömningen, se min avhandling.

<sup>4</sup> <https://www.skolverket.se/undervisning/gymnasieskolan/nationella-prov-i-gymnasieskolan/genomfora-och-bedoma-prov-i-gymnasieskolan#h-Muntligadelprov>



LINDA BORGER är universitetslektor vid Institutionen för pedagogik och specialpedagogik vid Göteborgs universitet, där hon undervisar på lärarutbildningen och på kurser i bedömning och språkdidaktik. Hon arbetar även med analys och provutveckling i projektet Nationella prov i främmande språk (NAFS). Linda Borger har en bakgrund som gymnasielärare i engelska och svenska.