

Genomförande och bedömning av nationella muntliga prov i engelska: en pilotstudie

Pia Sundqvist

Lektor i engelsk lingvistik

pia.sundqvist@kau.se

Erica Sandlund

Lektor i engelsk lingvistik

erica.sandlund@kau.se

Lina Nyroos

Post-doc i nordiska språk

lina.nyroos@nordiska.uu.se

Peter Wikström

Doktorand i engelsk lingvistik

peter.wikstrom@kau.se

Abstract

Den här artikeln rapporterar resultat från en pilotstudie inom ramen för projektet Testing Talk (<http://www.kau.se/testing-talk>). Ett av projektets övergripande syften är att belysa komplexiteten i att bedöma muntlig engelska, ett annat att undersöka hur lärare går till väga när de förbereder och genomför muntliga prov. Fokus ligger specifikt på den muntliga delen av det nationella provet i engelska i årskurs 9. Pilotstudien omfattar två delar: en lärarenkät (N = 21) och lärarintervjuer (N = 3). Enkäten var webbaserad och innehöll förutom bakgrundsfrågor frågor om lärarens inställning till provet samt dess genomförande. Enkätsvaren har analyserats kvantitativt. Intervjuerna var semi-strukturerade och omfattade liknande frågeställningar som enkäten. Intervjuerna har analyserats kvalitativt. Resultaten visar bland annat att majoriteten av lärarna (90,5 %) tycker att det nationella provet fungerar bra alternativt mycket bra med avseende på att mäta muntlig språkfärdighet. Det är vanligt att man låter sina elever öva på gamla prov som en del av förberedelsearbetet. Lärarnas metoder beträffande hur man väljer att gruppera eleverna under själva provtillfället skiljer sig markant åt och bevekelsegrunderna för gruppindelning är också starkt differentierade. I artikeln diskuterar vi resultaten i förhållande till standardisering av prov och bedömning av muntlig färdighet.

Keywords: bedömning, muntlig färdighet, engelska, nationella prov, standardiserade prov

Inledning

Under 1980-talet fick den kommunikativa språkundervisningen sitt genombrott och i och med det följde en rad forskningsstudier om verbal interaktion, muntlig färdighet, bedömning av muntlig färdighet och muntliga prov (se t.ex. Brown & Abeywickrama, 2010; de Jong & van Ginkel, 1992; Lindblad, 1992; Mehnert, 1998; Porter, 1986; Reed, 1992; Riggenbach, 1998; Skehan & Foster, 1999; Young & Milanovic, 1992). Vi följer denna tradition inom projektet Testing Talk, där ett av projektets övergripande syften är att belysa komplexiteten i att bedöma muntlig engelska och ett annat att undersöka hur engelsklärare går till väga när de förbereder och genomför muntliga prov.¹ Vårt fokus ligger specifikt på den muntliga delen av det nationella provet i engelska i årskurs 9

I föreliggande artikel vill vi särskilt lyfta några resultat från en nyligen genomförd pilotstudie bestående av intervjuer samt en webbaserad enkät med verksamma engelsklärare. Pilotstudien är en del i förberedelserna inför en större datainsamling kring det muntliga nationella provet i engelska som genomförs under 2014 och utöver att fungera som instrumentutprovning syftade pilotstudien till att identifiera faktorer som kan spela in för provets genomförande och bedömning. Vi redovisar och diskuterar här de resultat som befanns vara av särskilt intresse för den större studien samt lyfter frågan kring standardisering och lika villkor för elever och lärare när det gäller de nationella proven i allmänhet och de muntliga proven i engelska i synnerhet.

Standardisering av prov – nationellt och internationellt

I Sverige hade vi haft standardiserade prov i engelska på gymnasienivå sedan 1960-talet, men det dröjde tjugo år innan en muntlig del inkluderades i dessa. Det muntliga provet var frivilligt till en början, men utfallet av en studie vid Göteborgs Universitet i slutet av 1980-talet och början av 90-talet (Lindblad, 1992) bidrog till att det blev obligatoriskt, bland annat eftersom man hoppades på en positiv *washback*-effekt (Bachman & Palmer, 1996, s. 30-31) på undervisningen beträffande användningen av talad engelska i klassrummet, vilket för övrigt var helt i linje med den kommunikativa språksyn som eftersträvades. Kort därefter infördes standardiserade nationella prov i engelska även på högstadienivå (obligatoriska sedan 1998). Sett ur ett internationellt perspektiv infördes våra nationella ämnesprov relativt sent. I USA hade man exempelvis börjat med ett standardiserat så kallat *Scholastic Aptitude Test* (sedermera *Scholastic Assessment Test*, *SAT*) redan i slutet av 1920-talet som en del i antagningsprocessen till eftergymnasiala studier; numera skapas även prov för lägre nivåer i skolan precis som i Sverige

¹ Testing Talk är finansierat av Vetenskapsrådet (Dnr 2012-4129). Svensk titel: Testa tal – tala under test. Projektthemsida: <http://www.kau.se/testing-talk>.

(<http://www.collegeboard.org/>). Till skillnad från våra nationella prov som genomförs separat i olika ämnen på olika dagar innebär ett SAT att eleven testas vid ett och samma tillfälle inom flera ämnesområden.

Forskningen på standardiserade prov är omfattande. Bland de fördelar som framkommit nämns ofta det faktum att standardiserade prov är validerade och redo att användas direkt av läraren, som alltså slipper spendera tid på att utveckla egna prov (Brown & Abeywickrama, 2010). Dessutom kan man administrera proven till en stor grupp testtagare på rimlig tid och korrekta svar/lösningar är ofta givna, vilket gör behov av tolkning överflödigt. Ibland används exempelvis flervalsfrågor som kan maskinrättas, vilket är synnerligen tidsbesparande. Bland de nackdelar som framkommit finns till exempel det faktum att det händer att lärare använder övergripande standardiserade färdighetstester (*proficiency tests*) som kunskapstester (*achievement tests*), ett förfarande som självklart är olyckligt (Brown & Abeywickrama, 2010). Vidare har forskning bland yngre elever visat att storskaliga standardiserade tester kan ha negativ påverkan på deras språkinläring (McKay, 2006). Det finns dock lysande undantag, såsom det prov som togs fram i Norge på 1990-talet där ingen formell inrapportering av resultat avkrävdes och lärarna samtidigt erbjöds kompetensutveckling om just bedömning; de kunde dra nytta av det standardiserade materialet i sin engelskundervisning (Hasselgren, 2000).

Sett från en något annorlunda synvinkel så föreligger andra typer av standardiseringsproblem just för *muntliga* provdelar. Det kan verka självklart att delar av prov med större tolkningsutrymme och med avsaknad av ett entydigt rätt svar också är svårare att standardisera, men det kan vara värt att dröja lite vid några av anledningarna till detta. Forskning om standardisering i andra sammanhang där muntlig interaktion är central, exempelvis i forskningsintervjuer och diagnostiska tester (se t.ex. Antaki, 1999; Houtkoop-Steenstra & Antaki, 1997; Marlaire & Maynard, 1990), har dock tydligt visat att det är komplicerat att uppnå standardisering i vad som samtidigt är ett socialt möte och ett samtal. Bland annat har samtalsforskning om standardiserade forskningsintervjuer visat att en fråga som i förväg formulerats i en intervjumanual sällan kan ställas på exakt samma sätt i alla intervjuer – och om de gör det så förloras också en del av vitsen med den talade intervjun till skillnad från en skriftlig enkät (Potter & Hepburn, 2005; Suchman & Jordan, 1990). Forskning om medarbetarsamtal i organisationer där man använder sig av frågeformulär i syfte att uppnå samma information i samtliga samtal har visat på just denna problematik, där samma förformulerade fråga levererades av chefer med små betydelsefulla skillnader i inramning, intonation och ordval, vilket fick direkta konsekvenser för vilken typ av svar medarbetaren gavs utrymme att ge (Nyroos & Sandlund, under utgivning). I en forskningsrapport (Seedhouse & Egbert, 2006) där interaktionen under så kallade *oral proficiency interviews* studeras (det vill säga standardiserade

muntliga prov i språk i intervjuformat som genomförs med en infödd talare och en testtagare) diskuterar författarna också hur examinatorernas små avvikelser från instruktionerna för samtalet skapar validitetsproblem och motverkar syftet med standardisering. Vår egen tidigare forskning om muntliga prov i engelska har också visat på skillnader exempelvis i hur elever respektive lärare tolkar en förutbestämd diskussionsuppgift (Sandlund & Sundqvist, 2012, 2013) och vi har kunnat påvisa att lärarens sätt att behandla elevernas bidrag till provsamtalet har stor betydelse för hur samtalet utvecklas. Standardisering av bedömningskriterier och dessutom dess praktiska tillämpning har naturligtvis också stor betydelse för huruvida man uppnår jämförbara resultat på exempelvis nationell nivå (jfr t.ex. Erickson, 2009a; Erickson, 2009b; Hasselgren, 2005).

Utöver de rent sociala och interaktionella aspekterna av standardisering i muntliga prov tillkommer förstås en rad praktiska förutsättningar (tidsåtgång, antal elever/prov osv.) som också rimligen kan förväntas få genomslag på hur provsamtalet ser ut som färdig produkt inför bedömning, och naturligtvis också hur själva bedömningarna görs.

Nationellt prov i engelska: fokus på muntligt delprov i årskurs 9

Som nämnts ovan har vi haft obligatoriska nationella prov i engelska på högstadienivå sedan 15 år tillbaka. Dessa konstrueras vid Göteborgs Universitet inom det så kallade NAFS-projektet (NAtionella prov i Främmande Språk). När det nationella provet i engelska först infördes i årskurs 9 år 1998 var *Lpo 94* den gällande läro- och kursplanen. Den var då tämligen ny och hade inneburit en stor förändring för skolan i och med sin målstyrning. Ett viktigt syfte med det nationella provet i engelska var och är att det skulle utgöra ett stöd för likvärdig bedömning och betygssättning i den nya målstyrda skolan (Erickson & Börjesson, 2001). Från och med vårterminen 2013 genomförs provet inom ramen för den senaste kursplanen, *Lgr 2011*, och i takt med ett allt starkare intresse för och fokus på resultat och måluppfyllelse i skolan har syftet för provet utökats. Alltsedan 2011 ska det utöver att stödja lärarens bedömning och betygssättning även ge underlag som medger analys av i vilken utsträckning kunskapsmålen nås på skol-, huvudmanna- samt nationell nivå (Velling Pedersen, 2013). Vidare sägs det också att det nationella provet kan bidra till att, likt tidigare, konkretisera gällande kursplanemål och betygskriterier .

Det nationella provet i engelska i årskurs 9 är ett typiskt *proficiency test* (se t.ex. Brown & Abeywickrama, 2010), det vill säga provet bygger inte på tidigare känt eller bearbetat stoff utan elevens allmänna språkfärdighet är vad som testas. Provet består av tre delar: Delprov A (tala/samtala), Delprov B1 (läsa och förstå) plus B2 (lyssna och förstå) samt

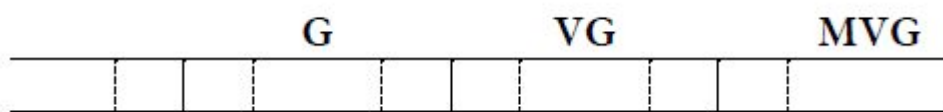
Delprov C (skriva). Delprov A, det muntliga provet, är inte tidsbundet utan genomförs någon gång under vårterminen från det att provmaterialet kommit till skolan (vanligtvis i januari) till och med provperiodens slut (vecka 22), medan de övriga delproven genomförs på två fasta dagar i april. Provkonstruktörerna lyfter fram att man ständigt har för avsikt att skapa prov där eleverna får en chans att visa vad de kan snarare än vad de inte kan (Velling Pedersen, 2013). Provet utvärderas kontinuerligt av Skolverket (se t.ex. Naeslund, 2004) och resultatrapporter och andra publikationer nås via NAFS-projektets hemsida (<http://www.nafs.gu.se/publikationer/>). Sett ur verksamma engelsklärares perspektiv, så visar ett flertal utvärderingar genom åren att provet är uppskattat (Erickson, 2009b) (för en närmare beskrivning av provet hänvisar vi till Sandlund & Sundqvist, 2012).

Att betyget på det nationella provet emellanåt avviker från slutbetyget i ämnet torde inte vara förvånande, men det finns nu rapporter från Skolinspektionen som påtalar att avvikelserna är allvarliga och frekvent förekommande (Skolinspektionen, 2010, 2011, 2012a, 2012b, 2013). På uppdrag av regeringen har Skolinspektionen under tre läsår samlat in och omrättat nationella ämnesprov, däribland det engelska nationella provet i årskurs 9. Totalt har cirka 1800 skolor sänt in provunderlag motsvarande 94000 prov (Skolinspektionen, 2012a). En generell slutsats är att avvikelserna varit stora under samtliga tre år och att det för engelskans del framförallt handlar om uppsatsdelen, där man förklarar avvikelserna med att det finns ett för stort tolkningsutrymme för bedömningen. Avvikelse mellan ursprungs rättarens bedömning och Skolinspektionens bedömning går oftast i negativ riktning, vilket innebär att Skolinspektionens anlitade bedömare (erfarna engelsklärare) bedömer att elevprestationen motsvarar ett lägre betyg än vad ursprungsrättaren (vanligtvis elevens egen engelsklärare) tillskrivit prestationen (Skolinspektionen, 2012a, s. 5).

Från vårt perspektiv är det intressant att konstatera hur Delprov A, det muntliga provet, över huvud taget inte ingår i Skolinspektionens granskningar. Som vi påpekat tidigare (Sandlund & Sundqvist, 2012) beror detta sannolikt på att det är omständligt och tidskrävande att hantera inspelat material samt att förhållandevis få lärare de facto gör inspelningar, trots rekommendation om detta från provkonstruktören. Enligt Erickson (2009b, s. 7) har andelen engelsklärare som gör inspelningar legat stabilt mellan 20 och 25 % under ett flertal år. Vi tycker det är rimligt att anta att tolkningsutrymmet vid bedömning av muntlig färdighet torde vara minst lika stort som vid bedömning av uppsatsskrivning, om inte större.

Inom NAFS-projektet har man självklart undersökt graden av samstämmighet mellan bedömningar av de olika delproven väldigt noga, såsom i rapporten *Nationella prov i*

engelska – en studie av bedömersamstämmighet (Erickson, 2009b). Generellt är bedömersamstämmigheten mycket god vad gäller läs- och hörförståelseuppgifter (även om det finns undantag för enstaka uppgifter) och på uppsatsdelen ligger korrelationskoefficienterna mellan 0,86 och 0,93, vilken således påvisar hög interreliabilitet. Med detta som bakgrund framstår Skolinspektionens avvikelserapporter som något märkliga. Förklaringen till skillnaderna i utfall mellan grad av samstämmighet i bedömningar hos Skolinspektionen jämfört med de som gjorts inom NAFS-projektet torde vara att man i NAFS egen granskning tillämpar en 10-gradig skala ("rutorna" i Fig. 1), medan Skolinspektionen tillämpar den 4-gradiga betygsskalan: Ej uppnått målen (EUM), Godkänt (G), Väl godkänt (VG) och Mycket väl godkänt (MVG). (Notera att EUM inte finns explicit uttryckt i Figur 1, men ska tolkas som att det ligger långt till vänster i bedömningsformuläret.) En konsekvens kan då bli att graden av samstämmighet blir lägre hos Skolinspektionen, eftersom skalan innehåller färre och därmed bredare steg. Vi noterar dock att den muntliga delen inte heller ingår i NAFS egen undersökning av bedömersamstämmighet. Detta är något överraskande med tanke på att det sammantagna nationella prov-betyget baseras på en modell där resultaten på samtliga tre delprov vägs samman och lika vikt ges åt de fyra färdigheterna tala/samtala, lyssna, läsa och skriva (Erickson, 2009b, s. 4). Vid utprovning av lämpliga uppgifter som ska ingå i det muntliga provet samlas inspelningar in och exempelinspelningar tas fram tillsammans med kommentarsmaterial avseende bedömning och betyg (ingår som CD-material till skolorna), men några inspelningar från Delprov A i skarpt läge samlas alltså inte in för rutinmässiga utvärderingar eftersom *krav* på inspelning saknas i instruktionerna till skolorna. Detta omöjliggör givetvis en slumpmässig insamling av just Delprov A, men medger detsamma för övriga två delprov (Erickson, 2009b, s. 7). Vi konstaterar att det muntliga provet i engelska i årskurs 9 framstår som något styvmoderligt behandlat i utvärderingar och att forskning på ämnet behövs.



Figur 1. "Bedömningsformulär" där läraren markerar elevens betyg; det finns ett formulär för varje delprov samt ett för det sammanlagda betyget på hela provet.

Data och metod

Det empiriska materialet som ligger till grund för föreliggande analys samlades in under vårterminen 2013 som en del av förberedelserna för en stor datainsamling kring det muntliga nationella provet i engelska planerad till höstterminen samma år. Materialet består av två delar: enkätsvar (kvantitativ analys) och intervjusvar (kvalitativ analys). Syftet med både enkäten och intervjuerna är att undersöka verksamma engelsklärares

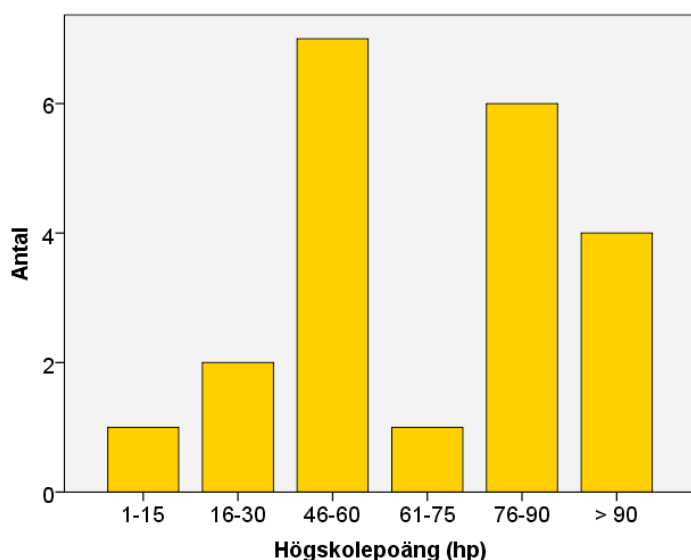
uppfattningar om provet. Den kommande stora enkäten kommer att omfatta ett slumpmässigt representativt urval av engelsklärare från hela Sverige, medan de kommande intervjuerna kommer att genomföras med lärare på de skolor som deltar i projektet. Nedan presenteras först resultaten av lärarenkäten med korta förklarande kommentarer, följt av en redovisning av intervjuerna. Syftet med såväl enkät som intervjuer är att få en bild av hur lärare utför, förbereder och tänker kring bedömning av det muntliga nationella provet i engelska.

Resultat lärarenkät

Enkäten var webbaserad och bestod av 23 frågor. Den inleddes med ett antal bakgrundsfrågor för att sedan fokusera helt på det muntliga nationella provet; enkäten innehöll dessutom specifika enkätutvärderingsfrågor på slutet. Själva distributionen av enkäten gick till så att vi e-postade en enkätlink till vårt kontaktnät av lärare samt delgav deltagarna vid en fortbildningsdag på Karlstads Universitet densamma. I instruktionen framgick att de endast skulle besvara enkäten om de hade erfarenhet av att genomföra muntliga nationella prov i engelska. Länken fanns öppen att besvara under en vecka i slutet av vårterminen. Målsättningen var att få in minst 20 svar, vilket uppfylldes. En utvärderingsfråga visade att det tog ungefär fem minuter att besvara enkäten, vilket var den förväntade tidsåtgången.

Respondenter

Enkäten besvarades av 21 engelsklärare (21 kvinnor, 1 man). Medelåldern var 46 år; den yngste var 34 år och den äldste 61. Majoriteten arbetade för närvarande i någon värmäländsk kommun (66,7 %), medan övriga (33,3 %) arbetade utanför länet. En av respondenterna arbetade på en friskola medan samtliga övriga tillhörde kommunala skolor. Samma siffror gällde för övrigt beträffande lärarexamen: alla utom en hade lärarexamen. På en fråga om de hade lärarlegitimation för att undervisa i engelska på lägst högstadienivå svarade 15 (71 %) jakande, medan 6 (29 %) saknade en sådan. Beträffande akademiska studier i engelska så hade drygt hälften (52 %) 61 högskolepoäng (hp) eller fler, det vill säga minst två terminers studier. Fyra stycken (19 %) hade studerat mer än tre terminer, se Figur 2 för ytterligare information.



Figur 2. Antal högskolepoäng i engelska bland respondenterna.

Våra respondenter representerar en relativt erfaren skara engelsklärare. Det var bara två (9 %) som jobbat i färre än fem år. Sju (33 %) hade jobbat i 6–10 år, sju (33 %) i 11–20 år och fem (24 %) i fler än 20 år. I vårt urval arbetade tre på mellanstadiet, tolv på högstadiet, två på gymnasiet och fyra inom den kommunala vuxenutbildningen.

Lärares åsikter om provet

Vi undrade vad lärarna tyckte om provet, så en fråga löd: ”Hur bra tycker du att det muntliga nationella provet fungerar för att mäta muntlig språkfärdighet?” Här svarade fyra lärare (19,0 %) ”mycket bra”, femton (71,4 %) ”bra” och två (9,5 %) ”ganska dåligt” (ingen valde alternativet ”dåligt”). Med andra ord, nästan alla (90,5 %) tyckte att provet mäter muntlig färdighet på ett bra sätt.

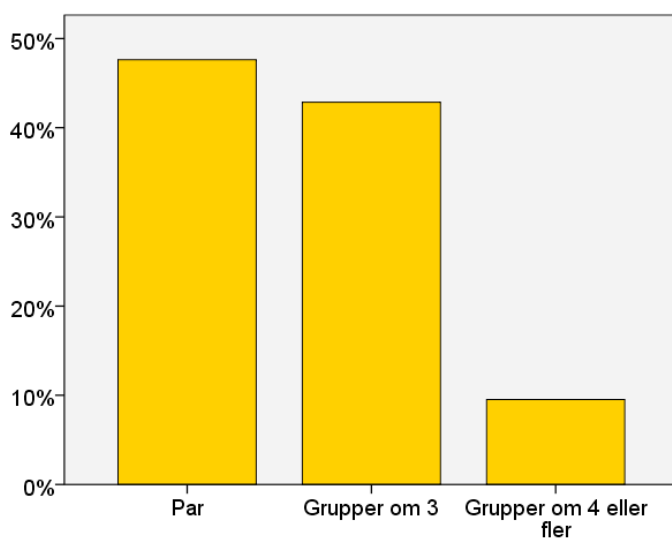
Provförberedelse

I Skolverkets/provkonstruktörens instruktioner till skolorna inför det nationella provet står det beskrivet hur man som lärare bör förbereda både sig själv och sina elever inför de olika provdelarna. På NAFS-projektets hemsida tillhandahåller man också exempeluppgifter från gamla prov, som det är vanligt att lärare låter elever ta del av och öva på. I vårt aktuella urval var det mycket vanligt att man gjorde just så och det gällde samtliga provdelar, inte bara den muntliga. Vi har i tidigare forskning funnit att när det uppstår problem under det muntliga provet kan dessa relativt ofta knytas till olika förståelser av själva provuppgiften, *task-related trouble* (Sandlund & Sundqvist, 2011), och vi har därför betonat vikten av att lärare förbereder eleverna noggrant för att minimera risken för att just den typen av problem ska behöva uppstå (Sandlund & Sundqvist, 2013).

Provgenomförande

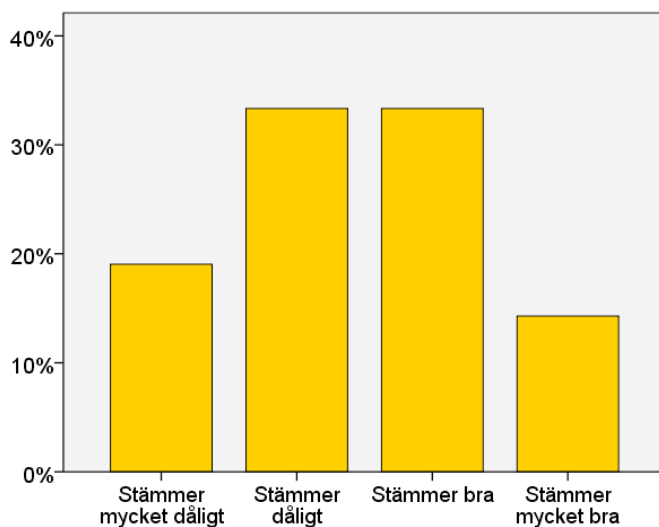
Trots utförliga lärarinstruktioner är det inte ovanligt att själva genomförandet av provet ser mycket olika ut på olika skolor. I Skolinspektionens granskning av skolor med stora bedömningsavvikelser blir detta tydligt – förutsättningarna lokalt skiljer sig markant åt (se t.ex. Skolinspektionen, 2013). I vår enkät ställde vi ett antal frågor där svaren mycket väl kan vara avhängiga sådana lokala premisser men även lärarens egna preferenser.

I lärarinformationen till det muntliga provet har det under en rad år stått skrivet att läraren kan låta eleverna genomföra provet i par eller grupper om tre eller fyra elever samt att eleverna helst inte ska befinna sig på mycket olika färdighetsnivåer. Med avseende på antalet elever har skrivningen i lärarinformationen från vårterminen 2013 (Skolverket, 2013) blivit något striktare än tidigare: ”Två eller ev. tre elever i taget genomför uppgiften tillsammans” (s. 17). Tyvärr lämnas dock läraren med oklara instruktioner om antalet elever, eftersom en annan skrivning faktiskt används i samma dokument: ”Provet genomförs som ett samtal mellan två eller fler elever” (s. 9) – det vill säga, här anges ingen övre gräns vid tre elever. Resultatet av vår enkätfråga om hur läraren vanligtvis grupperar sina elever speglar sannolikt både skiftande lokala förutsättningar och den oklarhet som finns i instruktionerna. Som Figur 3 visar så varierar det mycket hur engelsklärarna väljer att gruppera sina elever. De flesta har elev-par eller grupper om tre, men även grupper om fyra eller fler elever förekommer. En Chi-2 analys av resultatet ger nästan stöd till att spridningen kan generaliseras (χ^2 5,429, $df = 2$, $p = 0,066$). Med andra ord, det är ganska troligt att denna variation i grupperingar finns bland lärarkåren som helhet; vår kommande stora studie bör ge ett tillförligt svar på frågan.



Figur 3. Svar på frågan ”Hur brukar du vanligtvis gruppera dina elever när de gör det muntliga provet?”

Det varierar också mycket bland lärarna beträffande var de väljer att sätta sig under själva provet (se Fig. 4). I lärarinformationen sägs det att lärarens roll vid genomförandet är viktig, men att huvudprincipen är att läraren ”bör hålla sig i bakgrunden så mycket som möjligt och inte leda eller ta över samtalet” (Skolverket, 2013, s. 17). Det är möjligt att nästan hälften av lärarna tolkar instruktionen om att hålla sig i bakgrunden mer eller mindre bokstavligen och sätter sig därför en bit bort från eleverna under provtillfället (se staplarna ”stämmer bra” och ”stämmer mycket bra”, Fig. 4), medan övriga således sätter sig med/nära sina elever. En Chi-2 analys avslöjar dock att resultatet av denna fråga är svårtolkat (χ^2 2,429, $df = 3$, $p = 0,488$).



Figur 4. Grad av instämmande i påståendet ”Jag brukar sätta mig en bit från eleverna då de genomför provsamtalen.”

Det är möjligt att valet av var man sätter sig hänger samman med hur vanligt det är att läraren går in i provsamtalen. I Tabell 1 redovisar vi fördelningen på svar till fem påståenden som relaterar till själva provsituationen. Där ses till exempel att åtta lärare (38,1 %) tycker det ”stämmer bra” att det är vanligt att de behöver gå in och hjälpa eleverna under samtalet. Vi noterar också att samtliga gör någon form av anteckningar under provsamtalen och att majoriteten försöker vara tysta under elevernas samtal. Det är för övrigt förhållandevis få (4, 19 %) som upplever det som ”jobbigt” när det uppstår en tystnad under provet.

Lärarnas grad av instämmande i fyra påståenden som är knutna till bedömning redovisas i Tabell 2. Här är det tydligt att de allra flesta finner ett gott stöd i CD:n från Skolverket/provkonstruktören och att de ger individuell feedback till sina elever. Däremot är det relativt många (6 lärare, nästan 30 %) som tycker det är svårt att sätta betyg på det muntliga provet. Huruvida man sambedömer eller ej på skolorna varierar kraftigt (se Tabell 2), vilket är i linje med Skolinspektionens rapporter.

Tabell 1. Lärarnas grad av instämmande i fem påståenden knutna till provsituationen.

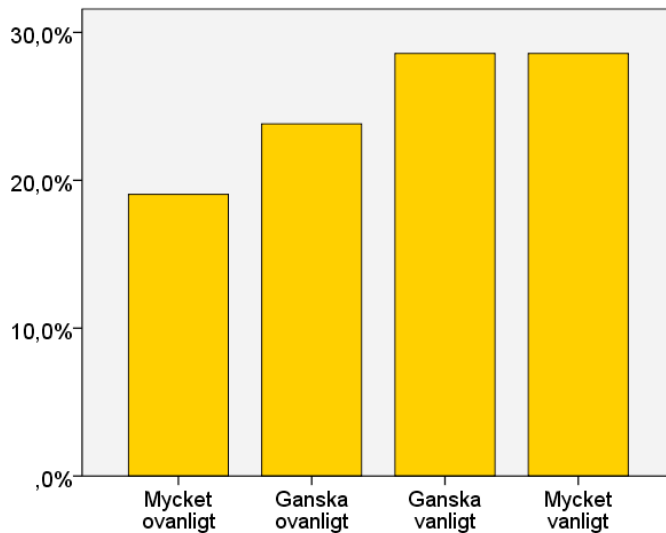
Påstående	N=21	Stämmer mycket dåligt	Stämmer dåligt	Stämmer bra	Stämmer mycket bra
Jag gör anteckningar om elevernas prestationer under provsamtalet.	<i>n</i> %	0 0	0 0	9 42,9	12 57,1
Jag brukar vara tyst under själva provet, så att eleverna får hantera provet på egen hand.	<i>n</i> %	0 0	4 19	13 61,9	4 19
Det är vanligt att jag behöver gå in och hjälpa eleverna under provsamtalet.	<i>n</i> %	3 14,3	10 47,6	8 38,1	0 0
Det är vanligt att eleverna växlar till svenska (eller annat språk) under provsamtalet.	<i>n</i> %	9 42,9	8 38,1	4 19	0 0
Jag tycker att det känns jobbigt om det uppstår tystnad under provsamtalet.	<i>n</i> %	3 14,3	14 66,7	4 19	0 0

Tabell 2. Lärarnas grad av instämmande i fyra påståenden knutna till bedömning.

Påstående	N=21	Stämmer mycket dåligt	Stämmer dåligt	Stämmer bra	Stämmer mycket bra
Det är vanligt att vi sambedömer det muntliga provet på min skola.	<i>n</i> %	3 14,3	8 38,1	8 38,1	2 9,5
Jag tycker att det är svårt att sätta betyg på det muntliga provet.	<i>n</i> %	0 0	15 71,4	6 28,6	0 0
Jag har gott stöd av Skolverkets bedömningsanvisningar och CD i betygsättningen.	<i>n</i> %	0 0	2 9,5	16 76,2	3 14,3
Jag ger mina elever individuell feedback på det muntliga provet.	<i>n</i> %	1 4,8	0 0	11 52,4	9 42,9

I lärarinformationen förordas att provet ska spelas in: ”Utifrån likvärdighetsaspekter och krav på dokumentation av elevprestationer är inspelning att starkt rekommendera” (Skolverket, 2013, s. 17). Men, som nämnts ovan, detta är ovanligt. I vårt urval var det visserligen ganska eller mycket vanligt bland nästan 60 % av lärarna att använda inspelning, men andelen lärare som finner det mycket eller ganska ovanligt att spela in är fortfarande stor, drygt 40 % i vårt urval (se Fig. 5). Möjligheten att lämna en kommentar på denna fråga gav ytterligare intressant information. Ett antal lärare nämner att de fått egna iPads, vilket medfört att de börjat spela in sina elever. Någon nämner att en privat iPad har använts. De som inte spelar in anger exempelvis brist på tid (”det finns absolut ingen tid att lyssna en gång till”) eller teknisk utrustning som skäl. En Chi-2-analys antyder

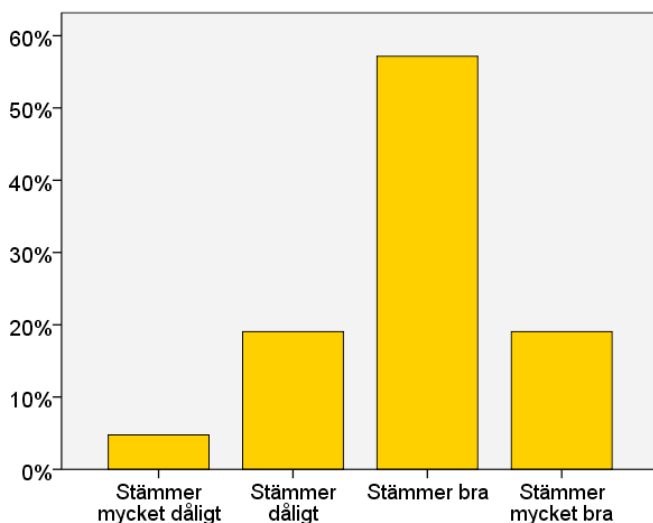
att resultatet att lärarna i hög utsträckning gör på olika sätt (en jämn spridning mellan svarsalternativen) sannolikt kan generaliseras till lärarkåren som helhet (χ^2 0,524, $df = 3$, $p = 0,914$).



Figur 5. Svar på frågan "Hur vanligt är det att du spelar in eleverna under det muntliga provet?"

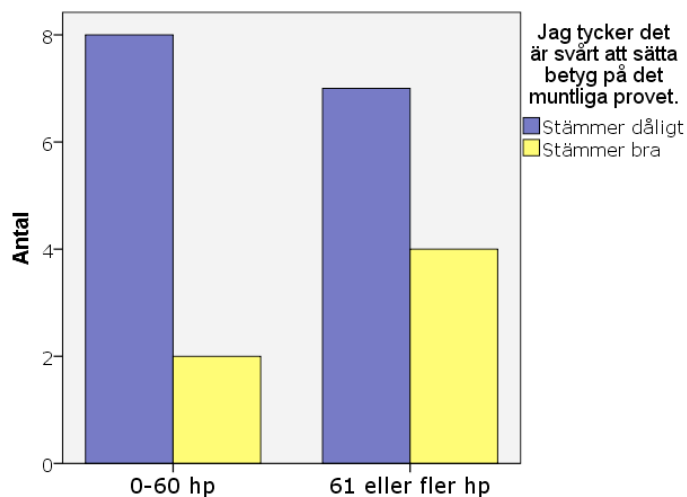
Bedömningspraktiker

I denna sektion berör vi de enkätfrågor som kan kopplas till lärarnas bedömningspraktiker. Vi frågade exempelvis lärarna i hur stor grad de instämde i påståendet "Jag sätter ett betyg på det muntliga provet direkt i anslutning till genomförandet" (se Fig. 6). En Chi-2-analys antyder att resultatet att de flesta lärare sätter betyget i direkt anslutning till genomförandet av provet kan generaliseras till lärarkåren som helhet (χ^2 12,714, $df = 3$, $p = 0,005$).



Figur 6. Grad av instämmande i påståendet "Jag sätter ett betyg på det muntliga provet direkt i anslutning till genomförandet."

Vi frågade också hur svårt engelsklärarna tycker det är att sätta betyg på det muntliga provet. Här delade vi in respondenterna i två grupper utifrån akademiska studier i engelska, med en skärningspunkt vid 60 hp (två terminers studier; tidigare behörighetsnivå för undervisning på högstadiet). En analys av svaren visade att det inte fanns någon signifikant skillnad mellan dessa två grupper (Fisher exact; $p = 0,635$; $\Phi = 0,181$). I båda grupperna finns en majoritet av lärare som inte tycker att det är svårt att sätta betyg på det muntliga provet, vilket illustreras tydligt i Figur 7; det är dock relativt sett fler lärare i gruppen med längre studier (> 60 hp) som tycker det är svårt att sätta betyg jämfört med hur det ser ut i gruppen med kortare studier.



Figur 7. Grad av instämmande i påståendet "Jag tycker det är svårt att sätta betyg på det muntliga provet" för lärare med 0-60 respektive 61 eller fler högskolepoäng (hp) i engelska.

Resultat intervjuer

Respondenter

I vår pilotstudie har vi intervjuat tre engelsklärare som vi fick kontakt med tack vare personliga kontakter: Anna, Bitte och Conny (fingerade namn). Nämnas kan att ingen av dem hade besvarat den skriftliga enkäten. Intervjuerna var semi-strukturerade och varje intervju tog ungefär 45 minuter. Anteckningar fördes och vid intervjuerna av Anna och Conny användes dessutom inspelning. Alla tre har lärarexamen. Anna har ämneskombinationen engelska-franska och har arbetat i 36 år som lärare på högstadiet. Bitte har kombinationen svenska-engelska-svenska som andraspråk och har arbetat i nio

år, även hon på högstadiet. Conny är utbildad gymnasielärare i historia och samhällskunskap och har läst till engelska i efterhand. Han har jobbat som lärare i knappt tio år på högstadiet. Lärarna är från olika skolor och delar av landet.

Åsikter om Skolverkets material om det nationella provet

Överlag uttrycker de tre lärarna en positiv syn på det nationella provet i engelska som helhet och de är nöjda med det muntliga provet. Anna verkade ha åsikten att om eleverna gillar det muntliga provet, så är det ett bra prov. Något år hade det varit sämre, till exempel en gång då eleverna förväntades prata om vykort/fotografier, vilket förutsatte god fantasi. Anna uttryckte också uppskattning över att få en stund med en liten grupp elever, vilket inträffade alltför sällan. Detsamma uttryckte Bitte, som berättade att det inte är så ofta eleverna sitter och lyssnar så på varandra: ”Det kan bli ganska intimt”. Conny ansåg att provet fungerar bra och mäter det som det ska mäta.

Provförberedelse och -genomförande

För att få en bild av vilka förutsättningar som råder på olika skolor frågade vi lärarna om förberedelser samt rutiner kring själva genomförandet av provet.

En viktig del av förberedelserna inför provet är arbetet med gamla prov. Alla tre lärarna uppger att de använder sådana som övningsmaterial så att eleverna ska få en möjlighet att bekanta sig med formen för provet. En lärare börjar arbeta med dessa gamla prov så tidigt som ett år innan det riktiga provet ska utföras, medan de andra två lärarna börjar jobba med de gamla proven ungefär ett halvår innan. Alla är eniga om att de gamla proven är viktiga i förberedelserna.

Vad gäller lärarnas egna förberedelser med hjälp av tillgängligt material så uppger de att de gör på lite olika sätt. CD-skivan som ingår i förberedelsematerialet från Skolverket lyfts fram av Anna och Conny, vilka båda lyssnar på den i bilen till och från arbetet. Anna säger sig lyssna väldigt många gånger på skivan för att få en bild av hur bedömningen har gjorts, medan Conny lyssnar uppskattningsvis fyra gånger. Conny lyssnar först på elevens prestation och gör sedan en egen bedömning innan han lyssnar på själva bedömningen. Ofta diskuterar han hur provkonstruktörens bedömning gjorts med sina kollegor. Bitte anger att hon inte lyssnar på skivan varje år, men oftast försöker att lyssna igenom denna.

En annan del i de mer omedelbara förberedelserna inför proven är gruppindelning av eleverna. Här resonerar de tre lärarna ganska lika. Alla påpekar att det allra viktigaste är att skapa goda förutsättningar för ett samtal. Conny understryker vidare det ansvar läraren har för att skapa goda förutsättningar och säger i intervjun att ”det gäller för läraren att skapa

en situation där eleverna kan nå sin topprestation”. Bitte betonar att det är viktigt att eleverna ska känna sig trygga. På grundval av detta delar alla de tre lärarna in eleverna i grupper med hänsyn till kunskapsnivå och förmåga att föra samtal. De siktar alla på en relativt jämn nivå hos eleverna i varje grupp, så att ingen elev ska känna sig sämre genom att någon i gruppen är väldigt mycket bättre. De tror också att detta främjar talförheten, vilken de påpekar är det viktigaste i sammanhanget. Andra saker det tas hänsyn till är sociala aspekter, såsom eventuell osämja elever emellan, i den mån läraren är medveten om dessa.

Vad gäller elevgruppernas storlek skiljer det sig dock åt mellan skolorna. Anna delar in i grupper om tre elever på grund av att hon måste ta hänsyn till hur många grupper och således lektioner som får gå åt till proven. Bitte och Conny behöver inte ta hänsyn till att ett visst antal lektioner ska gå åt, och Bitte delar vanligen in i grupper om fyra så att det varken är för få eller för många, och Conny delar in i grupper om mellan två och fyra elever. Conny har en del elever med diagnoser som till exempel Aspergers syndrom och ADHD, och för dessa är det extra viktigt att de blir grupperade tillsammans med någon som de kan prata med, säger han i intervjun.

Här ser vi att yttre omständigheter kan påverka gruppernas storlek. Anna uppger att om det skulle vara två elever i varje grupp skulle alltför många lektioner gå åt till detta, eftersom hon avsätter en grupp per lektionstillfälle. Det här leder till att provet är utspritt under ett antal veckor, ofta upp till tre månader, innan alla elever har klarat av provet. På Connys skola däremot, som är relativt stor, anordnas speciella dagar då alla muntliga prov i alla ämnen som har detta samlas ihop. Även Bitte avsätter en dag för proven, men på hennes skola är detta inte en självklarhet – hon uppger att hon kräver en vikarie som kan sysselsätta övriga elever som för tillfället inte gör provet.

Själva provet genomförs i alla fallen i olika grupprum, aldrig i ett klassrum. Bitte uppger att hon ”tar en ledig skrub”, medan Anna alltid är i ett och samma grupprum som ligger i anslutning till hennes eget klassrum, där hon indirekt har uppsikt över övriga elever genom att hon hör om de inte gör vad de ska. På Connys skola finns det många mindre grupprum, vilka används för proven. För Connys elever med exempelvis diagnosen Aspergers syndrom, vilka kräver väldigt speciella förutsättningar för att fungera i provsituationen, krävs emellanåt lite mer förberedelser. Conny uppger att han till exempel tänker på möblemanget och praktiskt förbereder rummet så att eleven ska känna sig bekväm.

För att återvända till att provet ibland sprids ut över tid och ibland inte, menar Bitte att det är en fördel att alla proven kan klaras av samtidigt ”så att det är gjort”, men nackdelen är

att det är väldigt tröttsamt för henne som ska göra bedömningen. Conny uppger samma för- och nackdelar, medan Anna inte tycks fundera så mycket över huruvida upplägget på hennes skola är bra eller dåligt då hennes upplägg tycks vara på det vis man alltid gjort på den skolan.

Lärarna har således olika yttre omständigheter att ta hänsyn till vad gäller att lägga upp själva genomförandet av provet. På vissa skolor sker en samordning vilken leder till att alla elever – på den skolan – får samma praktiska förutsättningar för provet. En konsekvens av detta är att eleverna skolorna emellan får väldigt olika förutsättningar. Exakt när provet ges under vårterminen kan innebära stora skillnader, eftersom det är upp till varje skola/lärare lokalt att besluta om provdatum inom ett stort tillåtet tidsspänn (se ovan).

Bedömningspraktiker

Gällande bedömningen av de muntliga nationella proven så framkommer det dels att de olika lärarna har olika rutiner, dels att de olika skolorna arbetar på olika vis.

Vad gäller rutiner för inspelning skiljer det sig åt mellan skolorna. Anna spelar aldrig in provtillfällena på grund av att hon inte har tid att gå tillbaka till dessa i efterhand. Bitte har gjort det någon gång och då mest för att någon annan lärare gjorde det, men hon tycker inte att det är någon större hjälp. Conny däremot spelar alltid in, och för det mesta på video. På hans skola spelar de flesta lärare in.

Under själva provtillfället uppger alla tre lärare att det är fokuserade på vad de ska lyssna efter. Anna och Conny menar att instruktionerna är glasklara: det är elevernas ”kommunikativa förmåga” som ska bedömas, och Anna specificerar att det är ”förståelse, kommunikation och intonation” som hon lyssnar efter. Bitte har mer specifik vägledning än så, hon använder ett protokoll som hon inte minns vem som utarbetat, men som används av flera lärare på skolan. Protokollet är utformat så att det har kolumner med olika rubriker såsom *idiomatiska uttryck, uttal, grammatik, samspel med kompis, försök att utveckla samtal* etc, samt en rad per elev. I protokollet sätter Bitte kryss samt antecknar. Anna gör anteckningar, medan Conny gör färre anteckningar på grund av att han också spelar in proven och kommer att återvända till inspelningarna senare. De här svaren visar att det finns olika formaliserade rutiner för hur lärarna gör. De utgår från de övergripande riktlinjerna, men har hittat egna rutiner för att ”söka” efter hur den kommunikativa förmågan ska bedömas i praktiken.

Vidare ger Anna och Bitte uttryck för att de redan under provtillfället får en klar bild av vilket betyg som ska sättas. Bitte menar att även om provet ska bedömas summativt så är

man som lärare färgad av till exempel vetenskapen om någons goda uttal och fokuserar vid provtillfället mindre på just detta. Anna uppger att hon tidigare år ofta berättade för eleverna vilket betyg de skulle få direkt efter provet, eftersom hon tyckte att provtillfället utgjorde tillräcklig grund för betygssättning. Numer gör hon det däremot inte, och det mest av rättviseskäl så att de som fortfarande inte har gjort provet inte ska känna sig stressade över att de har provet framför sig. Conny däremot ser provtillfället som ett tillfälle att fokusera på det som sker i stunden och använder sedan inspelningen för att gå tillbaka och utvärdera och revidera elevernas prestationer. Här ser vi att rutinerna för inspelning får konsekvenser. Eftersom själva provtillfället samt anteckningar utgör den enda grunden för betygssättning för Anna och Bitte fokuserar de på bedömning och betyg redan här på ett sätt som inte Conny gör, eftersom han kan gå igenom provet i efterhand med hjälp av inspelningen. Conny menar vidare att inspelningen ofta ger upphov till en mer nyanserad bild av det intryck han fick av elevens prestation under själva provet, då han till exempel upptäcker sådant som han inte noterade under provet.

Rutiner för inspelning har också betydelse för möjligheter till sambedömning. På Connys skola bedömer man sina egna elever men ”skickar ofta över filer till andra” för att få ytterligare en bedömning. Det finns inga formaliserade rutiner för sambedömning av det muntliga provet på Connys skola, men han ger uttryck för att detta sker frekvent. Så ser det däremot inte ut på de andra skolorna där ingen sambedömning sker av de muntliga proven, troligtvis till stor del på grund av att det inte är möjligt eftersom inga inspelningar görs.

En specifik fråga i relation till bedömning är den som berör hur kodväxling betraktas och bedöms av lärarna. Anna är tydlig med att minsta form av kodväxling, det vill säga ett enda ord på svenska, leder till utesluten möjlighet att få högsta betyg. Anna menar att instruktionerna är väldigt tydliga, eleverna ska bara prata engelska, och hon har vidare som måttstock att en person med engelska som modersmål som inte behärskar svenska ska förstå allt eleverna säger. Conny är också inne på samma linje, att instruktionerna är tydliga och att svenska drar ner helhetsintrycket. Dock menar han att det snarare är så att de med högsta betyg inte kodväxlar, än att kodväxlingen skulle vara anledningen till att vissa inte får högsta betyg. Han är mer nyanserad i sin syn på kodväxling än Anna och menar att man som lärare trots allt måste ta hänsyn till om eleven är väldigt duktig i övrigt och då inte kan straffa denne för några enstaka ord på svenska. Bitte har en betydligt mer liberal syn på kodväxling än de andra två lärarna och menar att i den mån kodväxling sker, beror den nog oftast på att eleverna är nervösa och slänger in småord som *typ* och *liksom* på svenska. Huvudsaken är dock att de pratar, resonerar Bitte, och reducerar inte betyg efter antal ord på svenska.

Synen på kodväxling är ett exempel på en specifik detalj gällande bedömningen av proven, vilken visar sig skilja sig åt mellan de tre lärarna. Detta är ett konkret exempel på hur lärarna dels tolkar instruktionen om att eleverna ska prata engelska på olika sätt, dels på hur elevers prestationer bedöms olika. Här visar det sig att diskrepansen mellan bedömningen av helhetsintryck kontra detaljer kan få konsekvenser. I projektet Testing Talk pågår därför särskilda delstudier av kodväxling under de inspelade proven, med fokus på de sociala sammanhang där eleverna verkar kodväxla frekvent och vad man kan utläsa av detta i relation till kodväxlingens funktioner (Nyroos, Sandlund, & Sundqvist, kommande).

Diskussion och utvärdering av pilotstudien

Som mätinstrument tjänade både enkäten och intervjuerna sitt syfte och vi kommer efter vissa justeringar använda dem i vår stora studie. Sammanfattningsvis belyser såväl enkäten som intervjuerna, sitt blygsamma antal till trots, att det råder stor variation vad gäller till exempel förutsättningar vid genomförandet av det muntliga nationella provet. I vissa avseenden är det praktiska skäl som leder till att omständigheterna för eleverna skiljer sig åt, i andra är det lärarens erfarenhet och åsikter som gör det. Att exakt samma förutsättningar inte kan råda för alla elever över hela landet är det kanske ingen som förvånas över, men resultaten av pilotundersökningen belyser några viktiga skillnader som råder mellan skolor och lärare, och pekar vidare på att konsekvenserna av dessa bör undersökas vidare. Vi ser till exempel antalet elever vid gruppindelningen som en sådan viktig skillnad. Resultaten av pilotstudien belyser också att det inte bara är elevernas förutsättningar som skiljer sig åt utan även lärarnas, beroende på vilket skola de tillhör. I intervjumaterialet framkom att Anna, som jobbar på en mindre skola, uppger att man ”gör som man alltid har gjort”, vilket fungerar väl enligt henne eftersom de flesta har arbetat på skolan under en lång tid. Conny däremot arbetar på en större skola där det förefaller förekomma mycket samarbete och diskussion lärarna emellan. När det kommer till sambedömning, eller ens möjligheterna för en sådan vilken förutsätter att det finns kollegor med kompetens inom samma ämne, ser det också väldigt olika ut lokalt på skolorna. Vidare är sambedömning inte ens möjlig om proven sällan spelas in (jfr Erickson, 2009b). Eftersom det åtminstone hitintills endast finns en stark rekommendation om inspelning från Skolverket/provkonstruktören men inte någon kravformulering, lär andelen lärare som spelar in knappast öka. Visserligen kan en ökad tillgång på bra teknisk utrustning (jfr synpunkter om iPads) möjligen bidra till att fler lärare börjar spela in, men för att möjliggöra en systematisk sambedömning lokalt och en central granskning av provet i skarpt läge i en storskalig studie förefaller det nödvändigt att göra inspelning till ett obligatorium för skolorna. Avslutningsvis, det faktum att lärare skapar egna bedömningsmatriser (jfr Bitte) innebär att lärare har olika formaliserade sätt att

bedöma en elevs muntliga förmåga, vilket i sin tur medför att lärare lyssnar efter olika aspekter i den muntliga elevproduktionen. Detta är något som onekligen inverkar negativt med avseende på den likvärdiga bedömning som egentligen eftersträvas.

Avslutande kommentar

Avslutningsvis kan vi konstatera att engelsklärare är nöjda med det muntliga nationella provet så som det ser ut idag. Det är glädjande. Däremot har pilotstudien också visat på allvarliga brister, inte beträffande provet i sig utan snarare beträffande omständigheter omkring provet i praktiken. Till exempel, utifrån perspektivet standardiserade prov, där det nationella provet i årskurs 9 förväntas mäta elevens kunskapsnivå i slutet av grundskolan, blir det problematiskt när vissa elever testas så tidigt som i januari på sin muntliga färdighet i engelska, medan andra testas i maj. Det är också otillfredsställande att förutsättningarna varierar så markant för provgenomförandet och att inspelningar görs i så liten utsträckning. Vi ser att just det muntliga nationella provet – trots att det är värt lika mycket som de andra provdelarna för betyget – behandlas styvmoderligt i utvärderingar: indirekt tillmäts därmed inte muntlig färdighet samma vikt som övriga språkliga färdigheter. Det behövs tvivelsutan mer forskning på det muntliga nationella provet i engelska.

Referenser

- Antaki, C. (1999). Assessing quality of life of persons with a learning disability: how setting lower standards may inflate well-being scores. *Qualitative Health Research*, 9(4), 437-454.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford: Oxford University Press.
- Brown, H. D., & Abeywickrama, P. (2010). *Language assessment. Principles and classroom practices* (2 ed.). White Plains, NY: Pearson Education.
- de Jong, J. H. A. L., & van Ginkel, L. W. (1992). Dimensions in oral foreign language proficiency. I L. Verhoeven & J. H. A. L. de Jong (Red.), *The construct of language proficiency: Applications of psychological models to language assessment* (s. 187-205). Amsterdam: John Benjamins.
- Erickson, G. (2009a). "Att bäras åt". Om den goda bedömningens flerfaldighet och ömsesidighet. I U. Törnberg, A. Malmqvist & I. Valfridsson (Red.), *Språkdiraktiska perspektiv. Om undervisning och lärande i främmande språk* (s. 159-173). Stockholm: Liber.
- Erickson, G. (2009b). *Nationella prov i engelska - en studie av bedömersamstämmighet*. Göteborg: Göteborgs Universitet, Institutionen för pedagogik och didaktik, Enheten för språk och litteratur.
- Erickson, G., & Börjesson, L. (2001). Bedömning av språkfärdighet i nationella prov och bedömningsmaterial. I R. Ferm & P. Malmberg (Red.), *Språkboken* (s. 255-269). Stockholm: Myndigheten för skolutveckling.
- Hasselgren, A. (2000). The assessment of the English ability of young learners in Norwegian schools: an innovative approach. *Language Testing*, 17(2), 261-277.
- Hasselgren, A. (2005). Assessing the language of young learners. *Language Testing*, 22(3), 337-354.
- Houtkoop-Steenstra, H., & Antaki, C. A. (1997). Creating happy people by asking Yes-No questions. *Research on Language and Social Interaction*, 30(4), 285-313.
- Lindblad, T. (1992). Oral tests in Swedish schools: a five-year experiment. *System*, 20(3), 279-292.
- Marlaire, C. L., & Maynard, D. W. (1990). Standardized testing as an interactional phenomenon. *Sociology of Education*, 63(2), 83-101.
- McKay, P. (2006). *Assessing young language learners*. Cambridge: Cambridge University Press.
- Mehnert, U. (1998). The effects of different lengths of time for planning on second language performance. *Studies in second language acquisition*, 20(1), 83-107.
- Naeslund, L. (2004). *Prövostenar i praktiken. Grundskolans nationella provsystem i ljuset av användares synpunkter*. Stockholm: Skolverket.

- Nyroos, L., & Sandlund, E. (under utgivning). From paper to practice: Asking and responding to a standardized question item in performance appraisal interviews. *Pragmatics & Society*.
- Nyroos, L., Sandlund, E., & Sundqvist, P. (kommande). The systematics of code-switched particles: The Swedish conjunction 'eller' in EFL testing interaction.
- Porter, P. A. (1986). How learners talk to each other: Input and interaction in task-centered discussions. I R. R. Day (Red.), *Talking to learn: Conversation in second language acquisition* (s. 200-222). Rowley, MA: Newbury House Publishers.
- Potter, J., & Hepburn, A. (2005). Qualitative interviews in psychology: problems and possibilities. *Qualitative Research in Psychology*, 2(4), 281–307.
- Reed, D. J. (1992). The relationship between criterion-based levels of oral proficiency and norm-referenced scores of general proficiency in English as a second language. *System*, 20(3), 329-345.
- Riggenbach, H. (1998). Evaluating learner interaction skills: Conversation at the micro level. I R. Young & A. W. He (Red.), *Talking and testing: Discourse approaches to the assessment of oral proficiency* (s. 53-67). Amsterdam: John Benjamins.
- Sandlund, E., & Sundqvist, P. (2011). Managing task-related trouble in L2 oral proficiency tests: Contrasting interaction data and rater assessment. *Novitas-ROYAL*, 5(1), 91-120.
- Sandlund, E., & Sundqvist, P. (2012). "But what can we learn from history" – om konsten att diskutera "rätt" under det muntliga nationella provet i engelska. *KAPET*, 8(1), 45-66.
- Sandlund, E., & Sundqvist, P. (2013). Diverging task orientations in L2 oral proficiency tests – a conversation analytic approach to participant understandings of pre-set discussion tasks. *Nordic Journal of Modern Language Methodology*, 2(1), 1-21.
- Seedhouse, P., & Egbert, M. (2006). The interactional organisation of the IELTS speaking test. *IELTS Research Reports*, http://www.ielts.org/pdf/Vol6_Report6.pdf (Vol. 6, s. 161-206).
- Skehan, P., & Foster, P. (1999). The influence of task structure and processing conditions on narrative retellings. *Language Learning*, 49(1), 93-120.
- Skolinspektionen. (2010). *Kontrollrättning av nationella prov i grundskolan och gymnasieskolan*. Stockholm: Skolinspektionen.
- Skolinspektionen. (2011). *Lika eller olika? Omrättning av nationella prov i grundskolan och gymnasieskolan. Redovisning av regeringsuppdrag*. Stockholm: Skolinspektionen.
- Skolinspektionen. (2012a). *Lika för alla? Omrättning av nationella prov i grundskolan och gymnasieskolan under tre år*. Stockholm: Skolinspektionen.
- Skolinspektionen. (2012b). *Riktad tillsyn av bedömning och betygssättning hos skolor med stora avvikelser vid omrättning av nationella prov*. Stockholm: Skolinspektionen.

- Skolinspektionen. (2013). *Tillsyn av bedömning och betygssättning. Med fokus på skolor med stora avvikelser i förhållande till Skolinspektionens omvärdering och i relationen mellan nationella prov och slutbetyg*. Stockholm: Skolinspektionen.
- Skolverket. (2013). *Ämnesprov, läsår 2012/2013. English. Lärarinformation. I Lärarinformation finns bedömningsanvisningar för Part A. Årskurs 9*. Stockholm: Skolverket.
- Suchman, L., & Jordan, B. (1990). Interactional troubles in face-to-face survey interviews. *Journal of the American Statistical Association*, 85(409), 232–241.
- Velling Pedersen, D. (2013). Ämnesprovet i engelska. I K. Hector-Stahre (Red.), *Ämnesproven 2012 i grundskolans årskurs 9 och specialskolans årskurs 10* (s. 8-18). Stockholm: Skolverket.
- Young, R. F., & Milanovic, M. (1992). Discourse variation in oral proficiency interviews. *Studies in second language acquisition*, 14(4), 403-424.