

Vad kan en elev som kan prata engelska? Didaktisk transposition av muntlig färdighet i lärares matriser för bedömning av det nationella provet¹

ORIGINALARTIKEL

Liliann Byman Frisé, Erica Sandlund & Pia Sundqvist

ABSTRACT

Swedish students' oral proficiency in English as a foreign/second language (L2) is tested annually in a part of the English national test entitled *Speaking* (test construct: *oral production and interaction*), which students' own teachers both administer and assess. Despite access to extensive assessment guidelines from the Swedish National Agency for Education, many teachers choose to construct their own scoring rubrics for assessment. In this study, we examine teachers' professional knowledge about assessment of oral proficiency in L2 English, as it emerges from policy transformation in these teacher-generated rubrics. Data consists of 20 rubrics for assessment of *Speaking* in years 6 and 9. Rubrics were collected through professional networks and two English-teacher groups on Facebook. Qualitative content analysis was used to analyze teachers' conceptualizations of the test construct, as well as similarities and differences between conceptualizations. Results showed consensus of *what* to assess, but not necessarily *how*, which was particularly salient for *interaction*. Analysis of policy transformation indicated that when teacher-generated scoring rubrics were used for assessment of *Speaking*, each criterion was considered separately, making them beneficial for formative feedback. The professional knowledge that emerged is therefore formed by teachers' dual roles – as examiner and teacher – in the test situation.

Keywords: Oral proficiency, scoring rubrics, EFL, LAL, transformation

LILIANN BYMAN FRISÉ

PhD-student in English Linguistics
Department of Language, Literature, and
Intercultural studies
Karlstad University
liliann.frisen@kau.se

ERICA SANDLUND

Professor of English Linguistics
Department of Language, Literature, and
Intercultural studies
Karlstad University
erica.sandlund@kau.se

PIA SUNDQVIST

Professor of English Language Education
Department of Teacher Education and
School Research
University of Oslo
pia.sundqvist@ils.uio.no

INLEDNING

Att kunna *tala* ett andra-/främmandespråk (L2²) är själva kärnan i att lära sig det nya språket, men muntlig färdighet är den språkliga förmåga som är svårast att bedöma på ett tillförlitligt sätt (Alderson & Bachman, 2004). Svårigheten beror i mångt och mycket på att ett stort antal aspekter måste beaktas samtidigt i bedömningen (Bachman, 1990) samt att bedömningen sker i en social situation. För att storskaliga språkprov ska anses tillförlitliga ur ett test- och bedömningsperspektiv bör förfarandet kring dem standardiseras (Sundqvist m.fl., 2018), men av flera skäl är muntliga provsituationer särskilt svåra att standardisera både vad gäller genomförande och bedömning, där en viktig anledning är provens situationsspecifika karaktär. I situationer där testtagare genomför provet i par- eller gruppsamtal tillkommer dessutom att den produkt som ska bedömas – muntlig förmåga i ett samtal – inte är en individuell prestation, utan en samproduktion av alla test(del)tagare (se May, 2009; Young & He, 1998). Vissa forskare menar att det är omöjligt att nå full samstämmighet mellan bedömare eftersom de ofrånkomligen kommer att bedöma samma produkt olika (Youn & Chen, 2021).

I Sverige genomför alla elever i årskurs 6 och 9 det nationella provet i engelska, vars syfte är att stödja en likvärdig och rättvis bedömning och betygsättning (Skolverket, 2022). I vår studie fokuserar vi på delprov A (*Speaking*) med testkonstruktionen *muntlig produktion och interaktion*. Många välkända storskaliga prov i muntlig färdighet i engelska som L2, såsom *Test of English as a Foreign Language* (TOEFL) och *International English Language Testing System* (IELTS), använder sig av bedömare som särskilt tränats i att bedöma det aktuella provet. I Sverige är det däremot vanligen elevernas egna lärare som både genomför och bedömer delprovet. Under provet delas eleverna in i par (tidigare även i mindre grupper) för att diskutera ämnen/påståenden eller frågor på ämneskort i provmaterialet. Lärarna får ingen särskild bedömarträning för att utföra uppgiften, men har tillgång till utförliga bedömningsanvisningar från Skolverket samt bedömda exempelsamtal med tillhörande kommentarer. Trots detta väljer många lärare att skapa sina egna matriser som de använder i sitt bedömningsarbete, matriser som kan inkludera egna formuleringar och/eller bedömningskriterier. En närmare undersökning av dessa egenproducerade matriser kan belysa dels hur lärare tar sig an bedömningsuppgiften, dels hur de tolkar vad muntlig färdighet i engelska innebär när bedömningen avser unga elever med varierande färdighetsnivåer.

Tidigare studier kring bedömning av muntlig färdighet i L2 har visat att bedömare gör olika tolkningar av bedömningskriterier samt hur dessa kan eller bör appliceras på autentiska tal och samtal (se Borger, 2019; Ducasse & Brown, 2009; May, 2011; Sandlund & Sundqvist, 2019, 2021). I linje med dessa observationer har ett flertal forskare uttryckt att det behövs ytterligare studier av bedömningsprocesserna i sig (Ducasse & Brown, 2009; Sandlund & Sundqvist, 2019; Youn & Chen, 2021). I denna studie undersöker vi bedömande lärares tolkningar av testkonstruktionen genom att granska vilka bedömningskriterier de väljer att inkludera i sina egenskapade matriser samt hur dessa kriterier formuleras och begreppsliggörs. Studien fyller således en forskningslucka vad gäller lärares professionskunskap om bedömning av muntlig färdighet, där professionskunskap kommer till uttryck i den transformation (Chevallard, 2007) som bedömningspolicyn genomgår för att kunna operationaliseras i en bedömningspraktik. I vår studie kan *professionskunskap* liknas vid begreppet Language Assessment Literacy (LAL), som definieras som en kombination och användning av kunskaper, förmågor och principer för att kunna genomföra bedömning av språk i olika undervisningssammanhang (Giraldo, 2021). *Kraftfull kunskap* (eng. *powerful knowledge*) myntades

Liliann Byman Frisé, Erica Sandlund & Pia Sundqvist

DOI: <https://doi.org/10.58714/ul.v17i2.15844>

av Young (2009) för att beteckna kunskap som är ämnesspecifik och skild från kunskap som elever kan tillägna sig utanför skolan. Enligt Young (2009) bör alla elever få tillgång till denna kunskap genom (skol)undervisning för att kunna utveckla intellektuell kraft (eng. *power*). *Kraftfull kunskap* relaterar därmed till undervisningens *vad*-fråga. I ett försök att vidga begreppet till att även omfatta *hur*-frågan, alltså *hur* elever kan utveckla kraftfull kunskap, föreslår Gericke m.fl. (2018) att inkludera begreppet *transformation* eftersom det kan förklara hur disciplinär kunskap kan omvandlas till något som är undervisningsbart och relevant för elever. Enligt artikelförfattarna har få empiriska studier undersökt vad kraftfull kunskap i olika ämnen skulle kunna vara. I föreliggande studie undersöker vi om, och i så fall hur, policy för bedömning transformeras för att kunna genomföras. Vi undersöker även lärares tolkningar av vad som står i fokus för provet, eller med andra ord – vad lärare anser att en elev som kan tala engelska *kan*, och därigenom vad lärare menar är kraftfull kunskap i relation till muntlig färdighet i engelska.

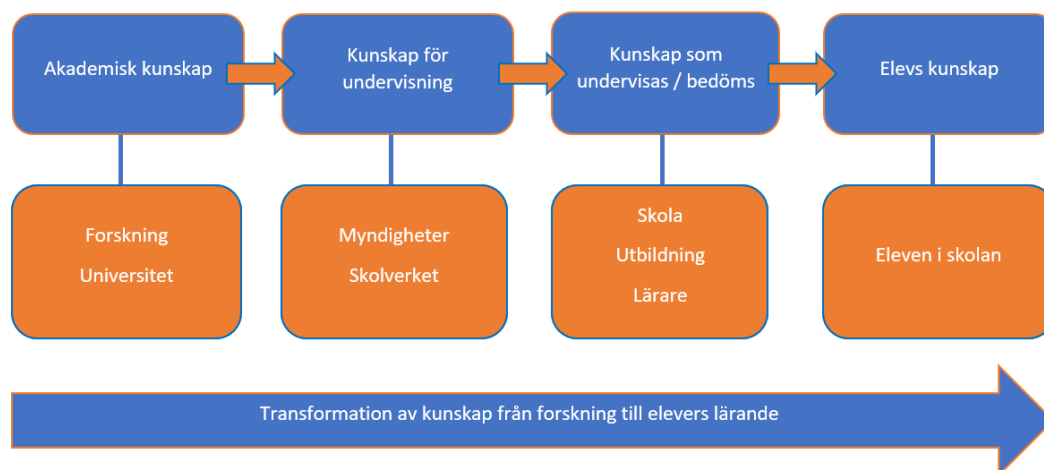
Syftet med studien är att belysa tolkningar av *muntlig färdighet* såsom de framträder i lärares egengjorda matriser för bedömning av den muntliga delen av det nationella provet i engelska i årskurs 6 och årskurs 9 (*Speaking*) för att få ny kunskap om, och i så fall hur, transformation (Chevallard, 2007) av policy sker när matriser används i bedömning.

Studiens forskningsfrågor lyder:

1. Vilka aspekter av muntlig färdighet fokuseras för bedömning i matriserna och hur är dessa formulerade?
2. Hur organiseras innehållet i matriserna?
3. På vilka sätt liknar matriserna varandra och hur skiljer de sig åt?

Teoretiskt ramverk

Eftersom syftet med studien är att belysa eventuella transformationer av testkonstruktionen *muntlig produktion och interaktion* när lärare operationaliserar policy för bedömning används det teoretiska ramverket *the Anthropological Theory of the Didactic* (Chevallard, 2007), vilket vi valt att översätta till *antropologisk didaktikteori* (ATD). Enligt detta ramverk finns ett dialektiskt förhållande mellan institutioner och människor verksamma inom dessa, där innehåll samproduceras på olika hierarkiska nivåer (Achiam & Marandino, 2014). Centralt i teorin om didaktiska transpositioner (eng. *the theory of the didactic transposition*, Chevallard, 2007, s. 132) är föreställningen att kunskap är föränderlig eftersom den påverkas av omständigheter inom den institution där den kommer till uttryck. På så sätt anpassas kunskap till den institution den existerar i, vilket i ramverket ATD kallas *the institutional relativity of knowledge* (Bosch & Gascón, 2014), av oss översatt till *kunskapens institutionella relativitet*.



Figur 1. Antropologisk didaktikteori (efter Chevallard, 2007)

Figur 1 visar hur kunskap, enligt ATD, genomgår en transformationsprocess för att kunna operationaliseras inom olika institutioner. Kunskap genererad i akademien behöver transformeras av personer verksamma inom olika organisationer och myndigheter (till exempel Skolverket) för att kunna bli policy (till exempel bedömningsanvisningar för nationella prov). Den kunskap som framträder i policydokument behöver i sin tur transformeras för att kunna omsättas till handling (till exempel i undervisning- och bedömningspraktiker). Slutligen formas varje elevs kunskap av den undervisning och bedömning hen får ta del av. Vi använder ATD i vår studie för att analysera eventuella transformationer av testkonstruktet *mundlig produktion och interaktion* när det operationaliseras i en bedömningspraktik.

I ATD finns en idé om *praxeologier*, enheter som består av *praxis* och *logos*. För att kunna förklara ”sann” kunskap (eng. *the fate of "true" knowledge*, Chevallard, 2007, s. 133) behöver båda dessa tas i beaktande. *Praxis* består av en uppgift (eng. *task*) och den teknik (eng. *technique*) med vilken uppgiften utförs. *Logos* består av teknologi (eng. *technology*) som är diskursen bakom tekniken (exempelvis varför en viss teknik lämpar sig för att utföra en uppgift), samt en teori som stödjer teknologin. I studien ses matriser skapade av lärare som *tekniken* som stödjer *uppgiften* att bedöma muntlig färdighet i engelska. Bedömande lärares bakomliggande anledningar till att skapa matriser, liksom teorin som stödjer matrisers användning i bedömningssituationen, ses som *logos*. Studiens empiri möjliggör inte analys av vilken *logos* som ligger till grund för användningen av matriser för bedömning av muntlig färdighet, men genom att studera tekniken och tolkningar av testkonstruktet som framkommer genom dessa, kan bedömande lärares kunskap om testkonstruktet belysas. Denna kunskap, genererad på ”skolnivån” i Figur 1, jämförs och kontrasteras mot kunskap om muntlig färdighet i engelska såsom den genererats på ”myndighetsnivån”, vilket reflekteras i bedömningsanvisningarna för provet. Genom att jämföra kunskap genererad på dessa två nivåer i ATD kan transformationer av testkonstruktet vid matrisanvändning framkomma.

TIDIGARE STUDIER

Bedömning av muntlig färdighet

Fri skriftlig eller muntlig produktion kallas ofta för komplexa prestationer i bedömningsssammanhang (se Wang, 2010), vilket delvis förklaras genom att uppgifternas ofta unika och oförutsägbara karaktär gör det svårt att skapa bedömningskriterier som bedömare tolkar på samma sätt. Forskning om interbedömarreliabilitet och bedömarträning har exempelvis visat att bedömare mycket väl kan landa i samma faktiska bedömning i form av ett betyg eller en poängsumma, men slutsatserna kan ha dragits på olika grunder (se till exempel Papajohn, 2002). Det är väl dokumenterat att brist på samstämmighet mellan bedömare när det gäller komplexa prestationer är förväntad eftersom bedömningen bygger på tolkningar av både bedömningskriterier och prestationer (se till exempel Brown, 2017; Davis, 2019; Meadows och Billington, 2005; Youn & Chen, 2021) Vad gäller muntlig fri produktion i samtalsform tillkommer en ytterligare komplexitet, nämligen den inneboende kollaborativa och samproducerade karaktären i ett mellanmänniskt samtal som också komplicerar individuell prestationsbedömning (se exempelvis Davis, 2009; May, 2009; Sandlund & Sundqvist, 2019; Sert, 2019). Bedömare måste således särskilja individuella prestationer hos de elever som bygger samtalet gemensamt, trots att en enskild testtagares bidrag också påverkas av och anknyter till samtalspartners handlingar.

Traditionellt sett har muntlig färdighet betraktats som bestående av några olika delar, som exempelvis CAF-modellen som omfattar de tre aspekterna *komplexitet*, *korrekthet* och *flyt* (eng. *complexity*, *accuracy* och *fluency*, se Skehan & Foster, 1999; Housen m.fl., 2012). Andra bedömningsmodeller delar in muntlig färdighet i fler analytiska kriterier, såsom flyt, anpassning, uttal, grammatik och vokabulär (till exempel McNamara, 2000). Huruvida bedömningen sker av varje delkomponent separat (analytisk bedömning) eller som en sammantagen helhet (holistisk bedömning) varierar dock för olika muntliga prov (se McNamara 2000, s. 43, om holistisk bedömning; se även Bilaga 1 för Skolverkets bedömningsanvisningar om analytisk/holistisk bedömning till bedömande lärare av delprov A). Forskningen är inte heller samstämmig vad gäller vilka komponenter som bör vägas in i bedömningen av muntlig färdighet i ett L2, eller vilka komponenter som bör väga tyngre för helheten. Så kallad kommunikativ språkundervisning och testning fick stort genomslag under 1980-talet (Canale och Swain, 1980; Bachman, 1990), vilket förde med sig ett ökat fokus på språkanvändning för just kommunikativa syften. På senare tid har dessutom forskning om interaktionell kompetens fått ett ökat inflytande (till exempel Kramsch, 1986; Kasper & Ross, 2013; Salaberry och Burch, 2021; Salaberry & Kunitz, 2019; Young & He, 1998). Många muntliga språkprov har idag bedömningskriterier för interaktionella förmågor som en del av muntlig färdighet. Ducasse och Brown (2009) understryker vikten av forskning om bedömares uppfattningar om muntliga elevprestationer eftersom deras uppfattningar och tolkningar återspeglas i elevernas betyg (se även Borger, 2019).

Studier av bedömares förståelse av muntlig färdighet i språk

Forskning om bedömares förståelse av bedömningskriterier för muntlig färdighet i ett L2 understryker komplexiteten i bedömningsuppdraget. Exempelvis rapporterar Ang-Aw och Goh (2011) att bedömare hade olika syn på den relativa vikten av olika kriterier i bedömningsskalan, och även att sådant som låg lite utanför kriterierna (exempelvis en tydlig ansträngning att lyckas) vägdes in i beslutet. Andra studier (till exempel Sato, 2012) har pekat på att bedömare fokuserar på olika

aspekter beroende på vilken språkfärdighetsnivå testtagarna befinner sig på. I en studie kring hur bedömare uppfattar interaktionen mellan testtagare använde May (2011) en kombination av skriftliga bedömarprotokoll, analys av samtalen, bedömandiskussioner och det som på engelska kallas för *stimulated recall*. May (2011) visar att många av de aspekter som bedömarna uppmärksammande och behandlade som viktiga föll under sådant som de facto samkonstrueras av deltagarna, exempelvis att visa förståelse för och svara på samtalspartnerns bidrag. Såväl May (2011) som Sandlund och Sundqvist (2016) rapporterar också att bedömare ibland har en tendens att jämföra testtagarna med varandra snarare än gentemot bedömningskriterier.

I en svensk kontext har Borger (2019) undersökt hur bedömare av *Speaking* uppfattade interaktionell kompetens genom att studera bedömningar och bedömningsanteckningar. Studien visade att tre interaktionella drag föreföll särskilt märkbara: handlingar för att utveckla samtalsämnet, turtagning samt strategier för att lyssna interaktivt i samtalet. Även Borger observerade att bedömarna ibland vägde in aspekter som låg utanför bedömningsanvisningarna och förordar mer detaljerade instruktioner och riktlinjer för bedömning.

I de flesta studier inom forskningsområdet har bedömare haft tillgång till bedömningskriterier eller matriser som bestämts på förhand. Det finns dock exempel på studier där ingen gemensam bedömningsskala använts, som Bøhn (2015). Hans resultat visar att bedömarna hade liknande uppfattningar om vilka aspekter som skulle ingå i bedömningen av muntlig färdighet, men det fanns skillnader beträffande hur viktiga de olika aspekterna ansågs vara i relation till varandra. Frisch (2015) visar dock att lärare faktiskt hade olika syn på muntlig färdighet, vilket kunde härledas till att lärarna hänvisade till olika kriterier i det nationella provets bedömningsinstruktioner.

Sammanfattningsvis pekar forskningen på att bedömare av muntlig färdighet i ett L2 har en komplicerad uppgift som förtjänar ytterligare vetenskaplig uppmärksamhet. Vår studie bidrar med ny kunskap om hur lärare – i egenskap av bedömare av det muntliga delprovet *Speaking* i engelska i årskurs 6 och 9 – hanterar denna svåra uppgift genom att begreppsliggöra de kategorier som används i bedömningen i sina egenutvecklade matriser för bedömningsarbetet. Tsagari (2021) menar att det finns mycket kvar att förstå vad gäller hur elever och lärare iscensätter bedömningspolicy i sina dagliga praktiker, och det är vår förhoppning att denna studie bidrar med ytterligare en pusselbit i sammanhanget.

Matriser

Internationellt har matriser använts som bedömningsverktyg flitigt både av provutvecklare och av lärare för formativ och summativ bedömning på alla nivåer i skolsystemen (Panadero och Jönsson, 2013). Matriser består vanligen av två delar: kriterier för bedömning och text där kriterier beskrivs på olika kvalitativa nivåer (Brookhart, 2018; Brown, 2012; Sadler, 2009). I forskningslitteraturen kategoriseras matriser ofta som *analytiska* eller *holistiska*, där analytiska avser matriser där varje kriterium bedöms för sig medan holistiska betecknar matriser där alla kriterier bedöms samtidigt (Brookhart, 2018; Brown, 2012). Innehållet i en analytisk och en holistisk matris kan vara identiskt; det som särskiljer dem är i första hand formatet, något som kan påverka både hur användbar en vald matris är samt syftet med att använda den. Val av matris beror därför på bedömningsituationen (Davis, 2018) eftersom matrisen representerar vad som anses viktigt i den prestation som ska bedömas (Khabbzbashi & Galaczi, 2020). Båda typer har sina för- och nackdelar. Holistiska matriser används ofta i storskaliga standardiserade tester (Brown, 2012; Jönsson & Svingby, 2007).

Deras främsta fördel är att de är praktiska eftersom endast ett bedömningsbeslut krävs (Brookhart, 2018; Brown, 2012; Davis, 2018), vilket gör dem mindre kognitivt påfrestande för bedömaren (Xi, 2007). Analytiska matriser anses vara särskilt lämpade för formativa ändamål eftersom de listar kriterier för en mångfacetterad prestation (Jönsson & Svingby, 2007). Många olika kriterier kan synliggöra elevers olika styrkor och utvecklingsområden (Davis, 2018; Khabbzbashi & Galaczi, 2020). Bedömningsprocessen blir även mer systematisk, vilket i sin tur kan förbättra tillförlitligheten (Khabbzbashi & Galaczi, 2020). Både Lee m.fl. (2010) och Xi (2007) fann dock en tydlig korrelation mellan testresultat baserade på analytisk respektive holistisk bedömning, vilket ledde till slutsatsen att analytisk bedömning är överflödigt. Då analytisk bedömning dessutom är mer kognitivt krävande, föreslår Xi (2007) att rigorös bedömarträning bör föregå bedömningsituationen för att bedömare på ett tillförlitligt sätt ska kunna särskilja mellan kriterier.

I bedömningslitteraturen används begreppen *rubrics*, *rating scales* och *checklists* för att beskriva olika typer av bedömningsdokument. Begreppen används ofta felaktigt som vore de utbytbara (Brookhart, 2018), men till skillnad från *rubric* betecknar *rating scale* och *checklist* bedömningsdokument som innehåller mycket kortfattad (eller ingen) text för att beskriva olika kvalitativa nivåer för de kriterier som anges. I denna artikel använder vi *matris* som samlingsbegrepp för alla bedömningsdokument som utgör data, men kategoriserar dokumenten enligt Brookharts (2018) definitioner av *rubric*, *rating scale* och *checklist*. Hon särskiljer de tre olika bedömningsdokumenten på följande vis:

A rubric articulates expectations for student work by listing criteria for the work and performance level descriptions across a continuum of quality [...]. Checklists ask for dichotomous decisions (typically has/doesn't have or yes/no) for each criterion. Rating scales ask for decisions across a scale that does not describe the performance.
(Brookhart 2018, s. 1)

I Sverige används bedömningsmatriser flitigt i skolor för formativa syften men det finns begränsat med forskning om vad matriserna innehåller och hur de skapas och används av lärare (Hirsh & Lindberg, 2015). Föreliggande studie bidrar framförallt med empiri beträffande matrisers *innehåll*. Eftersom innehåll representerar vad som anses vara av vikt i bedömningsituationen (Khabbzbashi & Galaczi, 2020) och eftersom själva organisationen av innehållet kan påverka det betyg som sätts (Brown, 2012; Davis, 2018; Khabbzbashi & Galaczi, 2020), kan vår studie belysa både *vad* som bedöms och *hur* det bedöms.

KONTEXT, DATA OCH METOD

Kontexten – nationella prov i engelska i Sverige

De bedömningsanvisningar som lärare får från Skolverket inför bedömning av det nationella provet i engelska omfattar anvisningar om hur provet ska organiseras (till exempel att dela in elever i par), administreras (till exempel att läraren i möjligaste mån ska hålla sig i bakgrunden) och förberedas (till exempel att lyssna på bedömda exempelsamtal, eller att spela in samtalen för en rättssäkrare bedömning). Skolverkets expertbedömare har bedömt varje exempelelevs muntliga färdighet i engelska. Bedömningen beskrivs i en sammanhängande text tillsammans med ett angivet betyg, vilka fungerar som referenspunkter för lärarens bedömning av de egna eleverna. Utöver detta bedömningsstöd får lärare dels ett noteringsunderlag som de kan använda vid bedömningen (Göteborgs Universitet, 2023, *Noteringsunderlag till delprov A*), och dels ett bedömningsdokument

på en sida som läraren ska utgå ifrån vid provsituationen (se Bilaga 1). Det sistnämnda dokumentet har reviderats i enlighet med de ändringar som skett i samband med den nya läroplanen 2022. Bilaga 1 är därför en översikt över hur bedömningsdokumentet var formulerat när vår studie genomfördes, vilket förklarar de skillnader som finns mellan Bilaga 1 och det bedömningsdokument som är gällande vid publikation av denna artikel (se Göteborgs Universitet, 2023, *Översikt: Bedömning delprov A*). Vid tiden för genomförandet av vår studie skulle enligt dokumentet läraren vid bedömningen utgå från att eleven ville och kunde utveckla och förmedla ett innehåll, både på egen hand och i samspel med andra. Läraren skulle göra en holistisk bedömning av elevernas muntliga produktion och interaktion utifrån språksynen som anges i kursplanen och betygskriterierna, vilka fanns definierade för betygen E, C och A. Utöver betygskriterierna angavs några *bedömningsfaktorer* som beskrevs som olika aspekter av kvalitet i muntlig språkfärdighet. Dessa inkluderade språkliga aspekter (till exempel *grammatisk struktur* och *vokabulär*), förmåga att producera innehåll (till exempel *att bidra med exempel och olika perspektiv*) och förmåga att interagera (till exempel *anpassning till syfte, situation och mottagare*) (se Bilaga 1). I dokumentet fanns inga explicita anvisningar för vilka grammatiska kunskaper eller vilken vokabulär eleverna förväntades använda sig av, men lärare kunde konsultera de bedömda elevexemplen där sådana aspekter tagits i beaktande vid bedömningen.

Data

Data utgörs av lärares egna matriser för bedömningen av *Speaking* i årskurs 6 och 9. För att inkluderas i studien skulle matrisen a) vara skapad av enskilda lärare, b) användas för bedömning av muntlig färdighet i engelska i årskurs 6 och/eller 9 och c) inte vara identisk med andra matriser i studien. Vi fick tillgång till matriserna via våra professionella nätverk med lärare samt två Facebook-grupper för lärare: ”Engelska för lärare i årskurs 4–6” (vid insamlingstillfället 4406 medlemmar) och ”Engelska för lärare i årskurs 6–9” (4394 medlemmar). Vår förfrågan resulterade i totalt 28 matriser (6 via nätverk, 22 via Facebook), där 20 kvalificerades för studien (8 för årskurs 6, 12 för årskurs 9). Dessa matriser var tillräckligt detaljerade för att kunna utgöra empiri för forskningsfrågorna. Varje matris från årskurs 6 kodades som ”Y6” (eng. *Year 6*) och från årskurs 9 som ”Y9” (kod ”Y912” = matris årskurs 9, nummer 12).

Analys

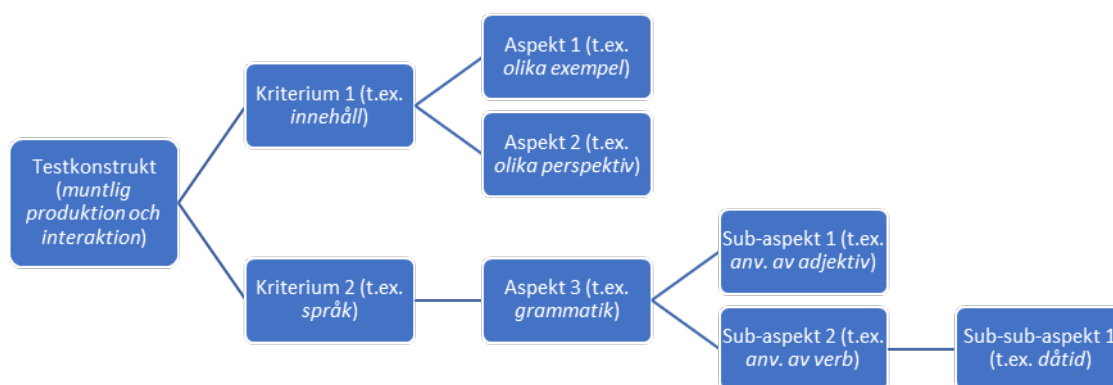
Vi analyserade data genom kvalitativ innehållsanalys som kompletterades med kvantitativa inslag (*summative content analysis*, Hsieh och Shannon, 2005). En sådan analys tar utgångspunkt i att kvantitativt identifiera och räkna förekomst av specifika ord och uttryck i textdokumenten, för att i nästa steg kvalitativt analysera underliggande betydelse av orden eller innehållet. Till hjälp för analysen användes begreppen *konstrukt*, *kriterium* och *sub-kriterium*, i enlighet med Bøhn (2015). I Bøhns studie av norska engelsklärares tolkningar av muntlig färdighet kategoriserades faktorer som lärarna valde att inkludera i en hierarkisk struktur. Termen *konstrukt* tillämpades för att beskriva makronivåer i bedömning, så i Bøhns studie representerade *konstrukt* alltså generella och övergripande faktorer i lärares utsagor av vad som bedömdes. För att representera olika mikronivåer (mer detaljerade faktorer för bedömning) använde Bøhn termerna *kriterium* och *sub-kriterium*.

Vi har valt att använda en liknande hierarkisk struktur för att kategorisera innehållet i bedömningsmatriserna, men till skillnad från Bøhn använder vi (vissa av hans) termer lite annorlunda. Detta beror dels på att vi har önskat hålla vår terminologi relativt nära den som används

av Skolverket, dels nära den terminologi som är något mer vanligt förekommande inom bedömningslitteraturen. Vi använder till exempel KONSTRUKT för att beskriva den språkliga förmåga som Skolverket explicit anger står i fokus för provet (med andra ord testkonstrukten *mundlig produktion och interaktion*). Begreppen KRITERIUM och ASPEKT använder vi för övergripande faktorer (ungefär jämförligt med Bøhns ”konstrukt” och ”kriterium”), medan vi för detaljerade faktorer på mikronivå har valt att använda SUB-ASPEKT, SUB-SUB-ASPEKT, och så vidare (se Figur 2). Att använda begreppet KRITERIUM för de faktorer som står i förgrunden för lärarnas bedömningar är i linje med hur det används i vetenskaplig litteratur (Brookhart, 2018; Brown, 2012; Sadler, 2009). Innehållsanalysen genomfördes i fyra steg. Nedan ges en översikt över hur analysen genomfördes i varje steg (för en mer detaljerad redogörelse av analysen samt hur intra- och interbedömarreliabilitet utvärderades, se Byman Frisé m.fl., 2021).

Översikt över analysens fyra steg

Steg 1 var att kategorisera de språkliga faktorer av muntlig färdighet som angavs i matriserna utifrån begreppen KRITERIUM, ASPEKT, SUB-ASPEKT, SUB-SUB-ASPEKT et cetera. Denna kategoriseringsprocess resulterade i ett träd-diagram (dendrogram) för varje matris där dess hierarkiska struktur tillika antal KRITERIER, ASPEKTER OCH SUB-ASPEKTER et cetera framträdde (se Figur 2).



Figur 2. Dendrogram för kategorier kriterium, aspekt, sub-aspekt och sub-sub-aspekt

I det fortsatta analysarbetet var det av vikt att se antalet KRITERIUM, ASPEKT, SUB-ASPEKT (et cetera) som framkom för varje matris endast som ett sätt att sortera data (i linje med Hsieh & Shannon, 2005).

Utvärderingen av intra- och interbedömarreliabilitet visade att det inte fanns en distinkt skiljelinje mellan de olika kategorierna KRITERIUM, ASPEKT, SUB-ASPEKT, et cetera. Vi bestämde därför att i det fortsatta analysarbetet betrakta data som bestående av två huvudkategorier: en med mer övergripande förmågor eller kunskaper där muntlig färdighet beskrivs i generella och allmänna ordalag på makronivå (till exempel *språk, innehåll*), och en med mer utförliga beskrivningar och/eller uttryck på mikronivå (till exempel *ställer frågor, håller med sin samtalspartner*). Kategorierna KRITERIUM och ASPEKT tillhör den första huvudkategorin (makronivå). Innehållsanalysen fortskred genom att fokusera på makronivån, det vill säga, KRITERIUM och ASPEKT för Y6-matriser, samt KRITERIUM och ASPEKT för Y9-matriser.

Steg 2 i analysarbetet bestod av innehållsanalys av makronivån (kategorierna KRITERIUM och ASPEKT), för vilket mjukvaran NVivo12 (QSR International Pty Ltd, 2018) användes. Vi kunde här urskilja nyckelbegrepp och mönster (Hsieh & Shannon, 2005), med vars hjälp teman kunde genereras (se Resultat). Baserat på frekvens valdes sex teman ut för ytterligare innehållsanalys (se Resultat för en utförlig beskrivning av urvalsgrunderna för våra teman).

I Steg 3 genomfördes innehållsanalys av mikronivån, vilket utgjorde den andra huvudkategorin (SUB-ASPEKT/SUB-SUB-ASPEKT, et cetera) med syfte att analysera hur de sex teman som genererats i Steg 2 exemplifierades och tolkades. I detta steg jämfördes även likheter och skillnader mellan hur varje tema beskrevs och tolkades på mikronivån i matriserna.

I Steg 4 kategoriserade vi matrisernas design genom att använda begreppen *rubric*, *rating scale* och *checklist*, i enlighet med Brookharts (2018) definitioner. Matriser kategoriserade som *rubric* skulle lista ett antal kriterier i fokus för bedömning, men även ange beskrivningar/förklaringar av kriterierna på olika färdighetsnivåer. En *rubric* ser således ut som ett rutsystem där kriterier listas i vänsterkolumnen. I resterande kolumner listas beskrivningar av varje kriterium, där varje kolumn representerar en viss färdighetsnivå. Även matriser kategoriserade som *rating scale* skulle lista ett antal kriterier i fokus för bedömning, men för var och en av dessa skulle det anges en skala där en separat bedömning av varje kriterium skulle anges. För att kategoriseras som *checklist* skulle matriser lista ett antal kriterier, men sakna text som beskriver eller förklarar vad som kännetecknar kvaliteten i förhållande till varje kriterium och på så sätt hur dessa kommer till uttryck i elevens produktion. En *checklist* ser därför ut som en lista med kriterier. En *checklist* och en *rating scale* liknar därför varandra, men där en *checklist* lämnar tomma ytor med plats för bedömarens kommentarer, har en *rating scale* en skala som uppmanar bedömare att ange en siffra, bokstav eller markering som representerar en viss färdighetsnivå.

RESULTAT

Resultaten av analysen redovisas genom att relatera till var och en av studiens tre forskningsfrågor.

Forskningsfråga 1: Vilka aspekter av muntlig färdighet har valts ut för bedömning i matriserna och hur är dessa formulerade?

Genom innehållsanalysen genererades nio teman för Y6-matriser och tretton för Y9-matriser (se Tabell 1 och 2 som även inkluderade frekvens av ord/begrepp).

Tabell 1. Teman i matriser årskurs 6 – antal omnämmanden i kategorierna KRITERIUM och ASPEKT.

Tema	N (procent)
Anpassning till syfte, mottagare och situation	14 (17,9)
Begriplighet och tydlighet	15 (19,2)
Strategier/ kommunikativa strategier	10 (12,8)
Fyllighet och variation	6 (7,7)
Sammanhang och struktur	10 (12,8)
Grammatiska strukturer	7 (9,0)
Uttal och intonation	6 (7,7)
Språk/Språk och uttrycksförmåga	5 (6,4)
Innehåll	5 (6,4)
Totalt	78 (100)

Not. Alla temaformuleringar är identiska med formuleringar i Skolverkets anvisningar.

Tabell 2. Teman i matriser årskurs 9 – antal omnämmanden i kategorierna KRITERIUM och ASPEKT.

Tema	N (procent)
Anpassning till syfte, mottagare och situation*	17 (11,3)
Begriplighet och tydlighet*	15 (9,9)
Strategier / Kommunikativa strategier*	16 (10,6)
Fyllighet och variation*	21 (13,9)
Sammanhang och struktur*	11 (7,3)
Grammatik	8 (5,3)
Uttal och intonation*	10 (6,6)
Språk / Språk och uttrycksförmåga*	9 (6,0)
Innehåll*	12 (7,9)
Engagemang / Initiativ	8 (5,3)
Flyt*	13 (8,6)
Ordförråd*	7 (4,6)
Interaktion / Interagera	4 (2,6)
Totalt	151 (100)

Not. Asterisken indikerar formuleringar som är identiska med formuleringar i Skolverkets anvisningar till lärare som fungerar som bedömare av det muntliga nationella provet i engelska.

Några begrepp/fraser användes mer än en gång per matris, till exempel fanns *anpassning till syfte, mottagare och situation* två gånger i flera matriser. En trolig förklaring till att lärarna har med just denna fras två gånger torde vara att den anges två gånger i Skolverkets *bedömningsfaktorer* (som en ASPEKT av KRITERIUM *innehåll* och som en ASPEKT av KRITERIUM *språk och uttrycksförmåga*). Utöver detta exempel är det tydligt att ord, uttryck och fraser från Skolverkets *bedömningsfaktorer* har modifierats i lärarmatriserna.

Som tidigare nämnts valdes sex teman för en fördjupad analys, för att avgöra om dessa tolkades på ett likartat sätt i matriserna.

1. Anpassning till syfte, mottagare och situation
2. Tydlighet och begriplighet
3. Strategier/Kommunikativa strategier
4. Fyllighet och variation
5. Engagemang/initiativ
6. Interaktion

Vi kan nämna att tema *Engagemang/initiativ* samt tema *Interaktion* inkluderades trots att de ej nämndes alls i kriterium eller aspekt i Y6-matriserna och endast ett fåtal gånger i Y9-matriserna (*engagemang* och/eller *initiativ* åtta gånger, *interaktion* och/eller *interagera* fyra gånger). De var dock framträdande på mikronivån när dendrogrammen upprättades i analysens Steg 1 och inkluderas av den anledningen.

Ett svar på forskningsfråga 1 är därför att de kriterier som står i fokus för bedömning i de lärarproducerade matriserna är starkt influerade av de analytiska bedömningsfaktorerna från Skolverket. Utöver att använda identiska ord och fraser, så som *anpassning till syfte, mottagare och situation*, visade våra data att de egna ord, uttryck och begrepp lärare förde in i sina matriser liknade Skolverkets uttryck i bedömningsanvisningarna. Vi fann många exempel på modifieringar som användes som kriterium och/eller aspekt, vilket gjordes på följande sätt:

- genom att dela upp långa uttryck till kortare, till exempel ändrades *kommunikativa strategier för att utveckla och föra samtal vidare* till *kommunikativa strategier och föra samtal vidare*,
- genom att fokusera på en mindre del av ett långt uttryck, till exempel ändrades *omfång, variation, tydlighet och säkerhet* till *variation och omfång*,
- genom att omformulera, till exempel ändrades *Innehåll – fyllighet och variation – olika exempel och perspektiv* till *Behandling av ämnet – med djup/ytligt*,
- genom att skapa nya kriterier baserade på delar av Skolverkets formuleringar, till exempel ändrades *Flyt och ledighet* samt *Uttal och intonation* till *Uttal, intonation och ledighet*.

Även om merparten av de kriterier som återfanns i matriserna påminde mycket om Skolverkets *bedömningsfaktorer* fanns även exempel på kriterier som inspirerats av *kunskapskraven*. Till exempel hade matris Y605 fyra kriterier (*muntlig presentation, muntlig interaktion, strategier och anpassning till syfte, situation och mottagare*), varav tre återfinns i kunskapskraven. Det förefaller alltså som om en ”etikett” sattes på tre av kunskapskraven och därefter användes som kriterier i matrisen.

Även om det var ovanligt var vissa kriterier inte i linje med provets bedömningsanvisningar. Ett exempel är *argumentera*, som fanns som ett KRITE RIUM i Y607. Förmågan att kunna argumentera på ett andraspråk är inget som nämns varken i bedömningsanvisningarna eller kursplanen i engelska för årskurs 6. Kursplanen nämner inte heller att elever på denna nivå ska läsa och/eller lyssna till argumenterande texter. Det är därför anmärkningsvärt att *argumentera* anges som ett kriterium för bedömning i årskurs 6. En möjlig förklaring kan vara att elever i provmaterialet uppmanas att ta ställning för eller emot påståenden, och att läraren således tolkar detta som att argumentation ingår i provet.

Forskningsfråga 2: Hur organiseras innehållet i matriserna?

För att analysera hur olika aspekter av muntlig färdighet framställdes i matriserna, studerade vi matrisernas design och deras textinnehåll.

Design av matriser

Som nämnts använde vi oss av Brookharts (2018) definitioner av *checklist, rating scale* och *rubric* vid kategoriseringen av matrisernas design (se Tabell 3).

Tabell 3. Kategorisering av matriser för årskurs 6 och 9.

Nivå	Checklist	Rating scale	Rubric
Årskurs 6: Antal kategoriserade matriser	4 (Y601, Y602, Y603, Y604)	0	4 (Y605, Y606, Y607, Y608)
Årskurs 9: Antal kategoriserade matriser	4 (Y901, Y903, Y910, Y911)	0	8 (Y902, Y904, Y905, Y906, Y907, Y908, Y912, Y913)

Tabell 3 visar att en majoritet av matriserna (12/20) kunde kategoriseras som *rubric* eftersom de innehöll kvalitativa beskrivningar av kriterier. Resterande åtta matriser var *checklist* då de hade KRITERIUM/ASPEKTER/SUB-ASPEKTER som bedömdes, men de innehöll *inte* beskrivningar av vad läraren skulle tolka som tecken på uppfyllelse av dessa. Dessa matriser innehöll i stället en tom yta där läraren kunde skriva in sina egna kommentarer, något som liknar det noteringsunderlag som Skolverket tillhandahåller (Noteringsunderlag till delprov A, Göteborgs Universitet, 2023). Ett exempel på en *checklist* är Y911 som bestod av tre kolumner. I vänsterkolumnen listades KRITERIER (på engelska). Övriga två kolumner var tomma, men mittkolumnen benämndes *comments*, och högerkolumnen benämndes *tick*, vilket förmodligen gjorde det möjligt för bedömande lärare att sätta en bock för varje KRITERIUM som uppfylldes av eleven i fråga.

De tolv matriser som kategoriserades som *rubric* skiljde sig åt med avseende på antal kriterier för bedömning samt detaljnivån i beskrivningarna. Gemensamt var formen av ett ruttmönster med fyra eller fem kolumner. Kriterier för bedömning listades rad för rad i vänsterkolumnen, medan resterande kolumner innehöll text för att nivågradera varje kriterium. I nio av de tolv matriserna benämndes dessa kolumner med "E", "C" och "A" i enlighet med betygssystemet. Två matriser hade i tillägg till dessa tre E/C/A-kolumner en fjärde kolumn kallad "F", och i en av matriserna saknade kolumnerna benämningar.

Vid analysen av de matriser vi kategoriserade som *rubric* kunde vi konstatera att de flesta matris-cellerna innehöll beskrivande text, även om det i vissa fall förfaller ha varit svårt att gradera separata kriterier i olika nivåer. Att designa en matris i form av ett ruttmönster verkar resultera i att varje cell fylls med text. Det faktum att alla utom en av *rubric*-matriserna hade benämnt kolumnerna enligt betygsskalan tyder på att läraren kan ha skapat sin egen matris för att underlätta vid betygsättningen av provet. Eftersom det dessutom finns lite plats för anteckningar antyder *rubric*-matriserna att när de används i provsituationen sker noteringar om elevens prestationer genom understrykningar/markeringar i redan befintlig text, snarare än att det skrivs nya noteringar/kommentarer.

Baserat på Brookharts (2018) definitioner av analytiska och holistiska matriser drar vi slutsatsen att matriserna i vår studie är analytiska, eftersom alla listar kriterier som står i fokus vid bedömningen. Även om så många som sju av tjugo matriser endast listade två kriterier (*innehåll* och *språk/språk och uttrycksförmåga*) exemplifierades dessa två med ett antal ASPEKTER och SUB-ASPEKTER. Sammantaget pekar vår innehållsanalys på att det sker en analytisk bedömning av muntlig färdighet

när de undersökta egenproducerade lärarmatriserna används i bedömningssituationen, och det gäller både för årskurs 6 och 9.

Textinnehåll i matriserna

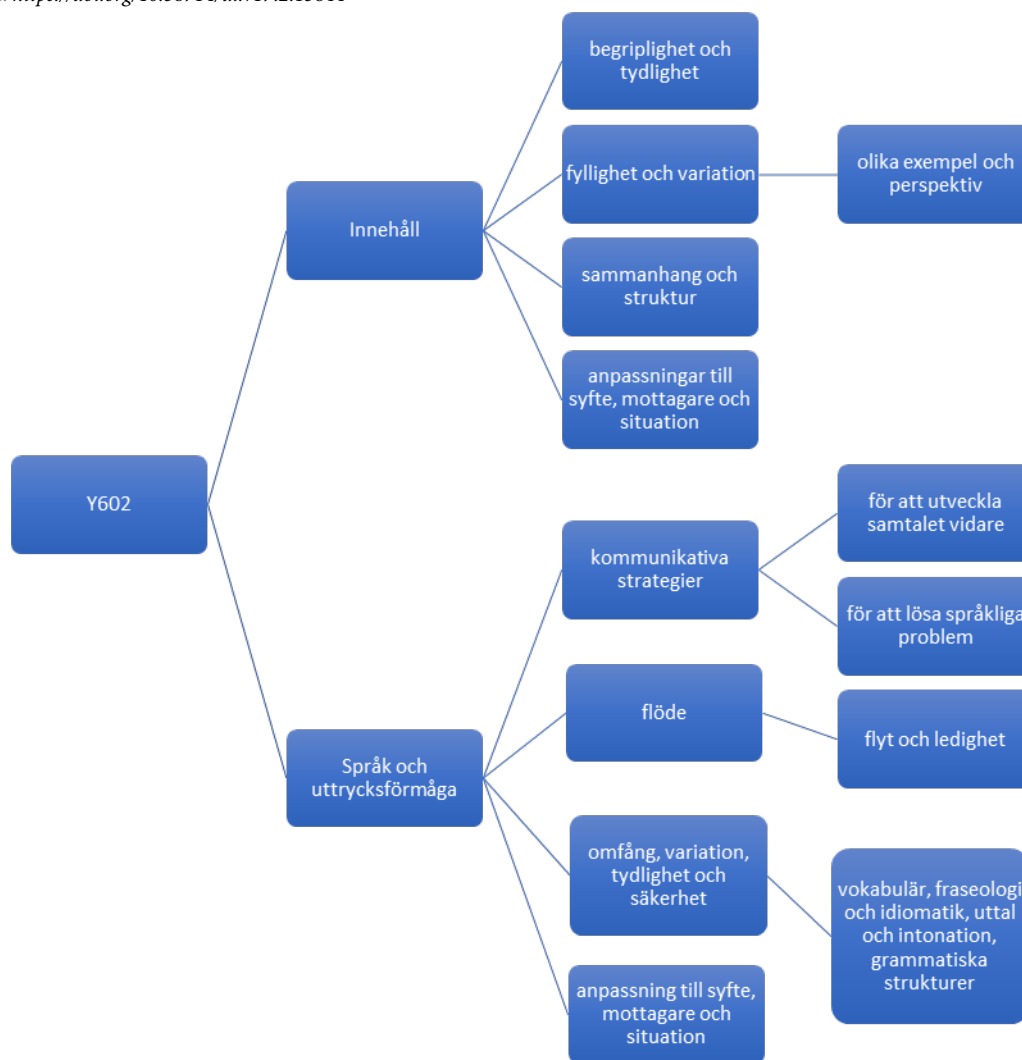
Trots att de kriterier som angavs i matriserna liknade varandra, identifierade vi sammanlagt 18 olika kriterier i Y6-matriserna, och 38 olika i Y9-matriserna. Något som angavs som ett KRITERIUM i en matris kunde vara en ASPEKT i en annan. Ett exempel var *strategier för språkliga problem* som angavs som ett KRITERIUM i Y901, medan *uses strategies to solve communication problems* angavs som en ASPEKT av KRITERIUM *Språk* i matrisen Y911. Antalet kriterier som angavs i matriserna varierade minimalt mellan årskurserna (2–10 i Y6-matriser; 2–11 i Y9-matriser) och samtliga matriser (oavsett årskurs) som endast innehöll två KRITERIER angav *Innehåll* och *Språk* (alternativt *Språk och uttrycksförmåga*). Det är värt att notera att i Skolverkets *bedömningsfaktorer* är det just dessa två som nämns, vilket tyder på att lärare har väglett av bedömningsfaktorerna också med avseende på hur innehållet i matriserna strukturerats.

Även om aspekter av muntlig färdighet för det mesta uttrycktes på ett generellt sätt i matriserna (till exempel *anpassning till syfte, mottagare och situation*), var det flera som beskrev elevers produktion på ett mycket mer detaljerat sätt. Ett exempel är Y905 som sträckte sig över tre sidor. Föga förvånande innehöll ordrika matriser många SUB-ASPEKTER med ett flertal exempel (även om de inte alltid presenterades som sådana) på vad elever gör eller säger för att uppvisa muntlig färdighet i relation till ett specifikt kriterium. I matris Y902 fanns konkreta ord och fraser som elever skulle kunna tänkas använda på tre olika nivåer. För att kunna bedöma om elevers innehåll är *fylligt och varierat* skulle elever framföra olika perspektiv, med fraserna ”If...I would” och ”I don’t, but...” som konkreta exempel.

De tolv matriser som kategoriserades som *rubrics* innehöll beskrivningar av olika kvalitativa nivåer för varje KRITERIUM. Vid innehållsanalysen av dessa beskrivningar fann vi att ord och fraser liknade de *värdeord* (till exempel *relativt tydligt, i någon mån*) som särskiljer de kvalitativa nivåer som uttrycks i Skolverkets *kunskapskrav* (till exempel i Y906: *uttrycker sig relativt obehindrat*). Vi fann även att beskrivningar av kvalitet på lägre nivåer ofta uttrycktes genom att ange brister i elevers tal och samtal, eller avsaknad av förmågor. Exempel på detta brist-perspektiv var formuleringar om antalet gånger en elev väljer att rätta sig själv. En elev på E-nivå beskrivs sällan göra korrigeringar i syfte att förtydliga sitt tal, medan en elev på A-nivå beskrivs göra det ofta. Andra exempel på utgångspunkt från ett brist-perspektiv var *använder svenska ord* (i kolumn för betyget E i Y605, utan motsvarighet i C- och A-kolumnerna) och *brister i intonation* (för betyget E i Y908).

Forskningsfråga 3. På vilka sätt liknar matriserna varandra och hur skiljer de sig åt?

För att förstå lärares tolkningar av vad som ingår i bedömning av testkonstruktionen skapade vi dendrogram (se Figurer 3–5).

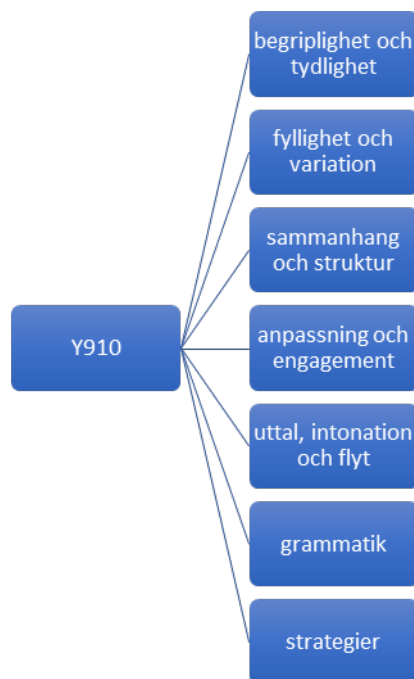


Figur 3. Matris Y602

VAD KAN EN ELEV SOM KAN TALA ENGELSKA? TOLKNINGAR AV MUNTIG FÄRDIGHET I
LÄRARES EGENPRODUCERADE MATRISER FÖR BEDÖMNING

Liliann Byman Frisé, Erica Sandlund & Pia Sundqvist

DOI: <https://doi.org/10.58714/ul.v17i2.15844>

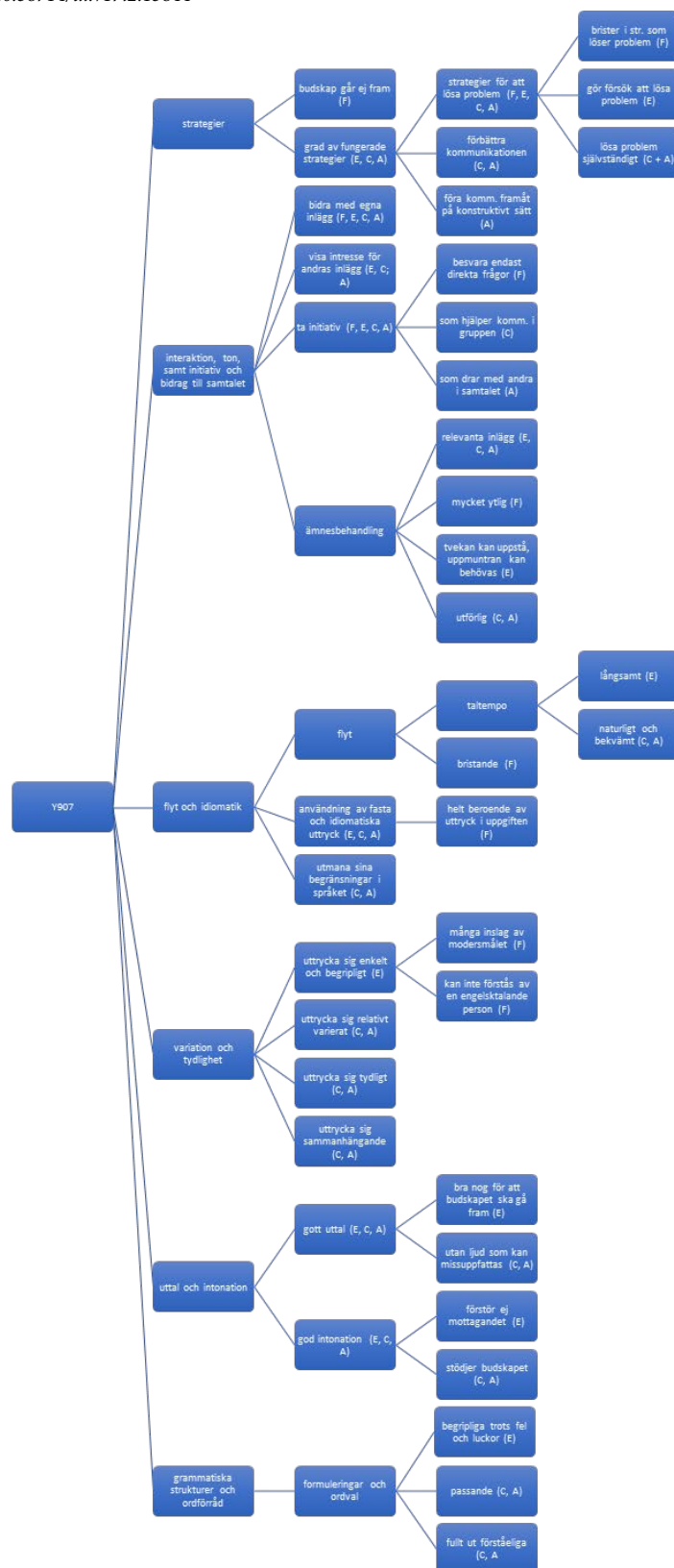


Figur 4. Matris Y910

VAD KAN EN ELEV SOM KAN PRATA ENGELSKA? DIDAKTISK TRANSPOSITION AV MUNTLLIG FÄRDIGHET I LÄRARES MATRISER FÖR BEDÖMNING AV DET NATIONELLA PROVET

Liliann Byman Frisé, Erica Sandlund & Pia Sundqvist

DOI: <https://doi.org/10.58714/ul.v17i2.15844>



Figur 5. Matris Y907

Det första dendrogrammet för en matris kategoriserades som *checklist* (se Y602, Figur 3). Här fanns 2 KRITERIUM (*Innehåll* och *Språk och uttrycksförmåga*), vardera exemplifierad med 4 ASPEKTER (till exempel var *Innehåll* exemplifierad med *begriplighet och tydlighet*, *fyllighet och variation*, *sammanhang och struktur* och *anpassningar till syfte, mottagare och situation*). Några ASPEKTER beskrevs ytterligare med SUB-ASPEKTER (till exempel *olika exempel och perspektiv*) och sammantaget innehöll denna matris 2 KRITERIUM, 8 ASPEKTER och 5 SUB-ASPEKTER. I analysen stod det klart att KRITERIUM, ASPEKTER och SUB-ASPEKTER nästan var identiska med Skolverkets *bedömningsfaktorer*. Förmodligen hade dessa kopierats in i lärarens egendesignade matris i syfte att bedöma testkonstrukten.

Matrisen Y910 (Figur 4) angav 7 KRITERIUM. Inga av dessa beskrevs eller exemplifierades, vilket innebar att denna matris innehöll lägst antal ASPEKTER, SUB-ASPEKTER et cetera.

Figur 5 (Y907) var en detaljerad matris som innehöll beskrivningar på vad elever gör eller inte gör på olika kvalitativa nivåer för att visa sin muntliga färdighet. Den innehöll 6 KRITERIUM, 16 ASPEKTER, 22 SUB-ASPEKTER, och 5 SUB²-ASPEKTER (den upphöjda siffran indikerar antal gånger ordet SUB repeteras, där SUB² betyder SUB-SUB-ASPEKTER). Dendrogrammet för den mest detaljerade matrisen i studien (Y904), som är alltför detaljerad för att kunna återges, omfattade 7 KRITERIUM, 11 ASPEKTER, 10 SUB-ASPEKTER, 19 SUB²-ASPEKTER, 30 SUB³-ASPEKTER, 14 SUB⁴-ASPEKTER, OCH 3 SUB⁵-ASPEKTER.

Genom att studera mikronivån (flera nivåer av SUB-ASPEKTER) framkom lärares tolkningar av de sex teman vi hade identifierat. Innehållsanalysen visade på både skillnader och likheter i lärarnas tolkningar. Ett tydligt resultat är att några teman genererade få SUB-ASPEKTER (exempelvis temat *anpassning till syfte, mottagare och situation*), medan andra genererade många fler (exempelvis *kommunikativa strategier*). En konsekvens av detta faktum är att mängden data som vi analyserat i relation till olika teman varierat stort. På grund av detta och av utrymmesskäl presenterar vi därför enbart några utvalda exempel nedan.

För två teman, *Begriplighet och tydlighet* och *Fyllighet och variation*, fanns stor samstämmighet mellan lärares tolkningar. *Begriplighet och tydlighet* tolkades som ett KRITERIUM med vilket elevs uttal och intonation kunde bedömas, även om det i några fall även inkluderade bedömning av i hur hög utsträckning en elev använder sig av sitt modersmål (L1). I två matriser, Y606 och Y907, relaterades *begriplighet och tydlighet* till om elevens muntliga produktion kunde förstås av en infödd talare. *Fyllighet och variation* tolkades å ena sidan som bedömning av elevs vokabulär, å andra sidan som bedömning av elevs förmåga att använda exempel och olika perspektiv om ett samtalsämne i provet. Temat *Anpassning till syfte, mottagare och situation* återfanns i så många som 16 matriser. Det exemplifierades dock endast i sju av dessa, vilket skulle kunna indikera att temat antingen anses självförklarande, eller så övergripande att det är svårt att definiera i konkreta exempel. I de matriser där temat exemplifierades knöts det tydligt till interaktiva färdigheter och till förmågan att spela sin roll i testsituationen, såsom ”visa stort intresse till att delta i samtalet” (betyget A, Y605) (jämför *testwiseness*, Bachman, 1990).

Temana *Engagemang/initiativ* och *Interaktion* var så sammanlänkade i matriserna att det var svårt att avgöra om lärare tolkade dem som olika eller identiska. Det fanns däremot en klar skillnad mellan Y6- och Y9-matriser. I Y6-matriser nämndes ”interaktion” endast 1 gång och ordet ”engagemang” nämndes inte överhuvudtaget (även om KRITERIUM *vilja att i tal och samtal...* fanns i 2 matriser). I Y607 var elevs förmåga att ”aktivt delta i samtalet” samt ”göra andra delaktiga i samtalet” angivna

som grund för bedömning. I Y9-matriser nämndes orden *engagemang/initiativ* och/eller *interaktion/interagera* i 10 av 12 matriser, och i 5 av dessa utgjorde de KRITERIUM (se Tabell 4).

Tabell 4. Förekomst av Engagemang/Initiativ och/eller Interaktion/Interagera som kriterium i matriserna för årskurs 9.

I vilken matris?	Titel på kriteriet i matrisen
Y901	<i>Interaktion/Initiativ</i>
Y902	<i>Interaktion/Strategier</i>
Y903	<i>Vilja och förmåga att i tal och samtal...</i>
Y907	<i>Interaktion, ton samt initiativ och bidrag till samtalet</i>
Y910	<i>Anpassning och engagemang</i>

Tabell 4 visar hur *engagemang*, *interaktion*, (*kommunikativa*) *strategier* och *anpassning till syfte, mottagare och situation* tolkades som utbytbara begrepp. Att *interaktion* bedömdes i ett prov med testkonstruktet *mundlig produktion och interaktion* var inte förvånande, men det verkade främst vara Y9-lärare som tolkade att det ingår i bedömningen. Som tidigare nämnts fanns det många likheter mellan ord och fraser som användes i matriserna och de som finns i Skolverkets bedömningsanvisningar (och då främst *bedömningsfaktorer*), men detta gällde dock inte *interaktion*, och endast i mycket liten utsträckning *engagemang*.

DISKUSSION

Syftet med studien var att belysa tolkningar av *mundlig färdighet* i lärares egenproducerade matriser för bedömning av den muntliga delen (*Speaking*) av det nationella provet i engelska i årskurs 6 och årskurs 9 samt om, och i så fall hur, transformation (Chevallard, 2007) av policy sker när matriser används i bedömning. Genom att använda oss av ATD-ramverket framträdde *kunskapens institutionella relativitet*, med andra ord den institutionsspecifika kunskap lärare har vad gäller bedömning av muntlig färdighet i engelska. Detta kan ses som professionskunskap. Den *teknik* (matriserna) lärare skapat för att utföra *uppgiften* (bedömningen) visade hur *praxis* tolkades av lärarna. Vi kunde tydligt se hur policy (här bedömningsanvisningarna för provet) påverkade vad lärare fokuserade på i sin bedömning av muntlig färdighet. Den transformation av policy som synliggjordes i lärarnas egenskapade matriser för bedömning var att Skolverkets *bedömningsfaktorer* varit av stor vikt. Khabbazbashi och Galaczi (2020) menar att innehållet i matriser reflekterar vad som görs viktigt i bedömning och i vår studie fick Skolverkets *bedömningsfaktorer* en särskild tyngd i det innehåll som lärarna valde. Eftersom *bedömningsfaktorerna* fick en så betydande roll för innehållet fanns det generellt stor enighet lärare emellan om vad som bör stå i fokus för bedömningen, något som är i linje med resultat från tidigare studier (Borger 2014, 2019; Bøhn, 2015; Frisch, 2015). Dessa *bedömningsfaktorer* introduceras för lärare i samband med det nationella provet, vilket kan förklara varför de fångat lärares uppmärksamhet på vad som kan och bör bedömas.

Det var inte möjligt för oss att applicera ATD-ramverket på våra data för att analysera *logos*, det vill säga bakomliggande anledningar till varför lärare skapar sina egna matriser samt vilken *teori* som stödjer matrisers användande. Våra resultat ger dock en indikation om varför denna *teknik* kunde underlätta utförandet av *uppgiften*. Trots att holistiska matriser anses vara mindre kognitivt krävande (och därmed mer effektiva) för bedömning av standardiserade prov än vad analytiska matriser är (Brookhart, 2018; Brown, 2012; Davis, 2018; Xi, 2007), var alla lärargenererade matriser i vår studie analytiska. En tänkbar förklaring är att analytiska matriser är mer användbara för

formativ återkoppling (Jönsson och Svingby, 2007), vilket är något som lärare är vana vid från sin klassrumsundervisning. Kanske är de därför mer benägna att skapa analytiska matriser när de väl skapar egna verktyg för bedömningsarbetet. Skolverkets *bedömningsfaktorer* passar dessutom väl in i en analytisk matris. Även om Skolverket poängterar vikten av att göra en holistisk bedömning av elevers produktion kan det vara så att införandet av *bedömningsfaktorer* skickar en signal till lärare att bedömningen snarare ska göras analytiskt. Valet av en analytisk matris i stället för en holistisk skulle också kunna förklaras med att elevers "taggiga profiler" (Davis, 2018) synliggörs lättare, vilket ger bättre möjligheter för formativ återkoppling (Jönsson och Svingby, 2007) efter den summativa bedömningssituationen. Ytterligare en förklaring till varför egenproducerade matriser skapas skulle kunna vara att en mer systematisk bedömningsprocess kan ge större insikt i vilka kriterier, aspekter och faktorer som ligger till grund för det holistiska betyget (Khabbazbashi och Galaczi, 2020). Detta kan underlätta jämförelser av samma elevs (muntliga) produktion gjorda av två eller fler bedömare, och på så sätt öka reliabiliteten i ett prov som är erkänt svårt att standardisera (jämför Sundqvist m.fl., 2018). En sista förklaring skulle kunna vara att analytiska matriser synliggör den systematiska bedömningsprocessen och därmed möjliggör för läraren att visa sin bedömarkompetens i händelse att den skulle ifrågasättas.

Som tidigare nämnts hade Skolverkets *bedömningsfaktorer* stort inflytande på vilka kvaliteter av muntlig färdighet som angavs som KRITERIUM, ASPEKT och SUB-ASPEKT i lärares matriser, och därmed vilka teman som genererades genom vår analys. Det fanns dock ett undantag och det var temat *Interaktion*. *Interaktion/interaktiva färdigheter* nämndes i matriserna men det verkade vara särskilt svårt att definiera vad det är, samt hur det kommer till uttryck i elevers tal och samtal. Det fanns också en tydlig skillnad mellan hur temat framställdes i Y6- och Y9-matriser. Här kan det finnas flera möjliga förklaringar. En är att bedömare tar elevernas nivå i engelska i beaktande (jämför Sato, 2012) och att lärare därför anser att Y9-elever bör ha mer utvecklade interaktiva förmågor än vad Y6-elever har. En annan är att Y6- och Y9-lärare uppmärksammar olika delar av bedömningsanvisningarna (jämför Frisch, 2015), och en tredje skulle kunna vara att våra data består av fler Y9-matriser än Y6-matriser. Oavsett anledning visar vår empiri att lärarna anser att Y9-elevers förmåga att interagera är en avgörande del av bedömning av testkonstruktionen – även om lärarnas tolkningar av vad det är och hur det tar sig uttryck varierar (i linje med Borger, 2019, och May, 2009). Denna variation kan hänga samman med att *interaktion* inte nämns i Skolverkets *bedömningsfaktorer* (utan enbart i beskrivningen av konstruktionen) och därmed inte heller definieras/exemplifieras där. Både Borger (2019) och May (2009) menar att bedömningsanvisningar för bedömning av *interaktion* bör utvecklas, med exempel på hur denna förmåga kommer till uttryck i elevers samtal/vid provsituationen samt rekommendationer kring hur individuella betyg kan ges för en produkt som samproduceras. Resultaten av föreliggande studie stödjer denna syn.

Våra resultat visar att lärarnas bedömning baseras på en kommunikativ språksyn, vilket är helt i linje med Skolverkets bedömningsanvisningar där lärare uppmanas att göra en helhetsbedömning utifrån denna språksyn. Det är därför inte förvånande att *engagemang* samt elevers *vilja* till att delta i provsituationen dök upp i matriserna som bedömningsbara faktorer, även om dessa saknas i de analytiska *bedömningsfaktorerna* som i övrigt visade sig påverka vilka kriterier lärarna valde att ha med i sina egenproducerade matriser. På samma sätt som med *interaktion* visar våra resultat att tolkningar av elevers *engagemang* varierade mellan lärarna, och därför föreslår vi tydligare riktlinjer för på vilket sätt elevers vilja (eng. *effort*, Ang-Aw & Goh, 2011) ska utgöra en del av bedömningen, och hur den tar sig uttryck i elevers muntliga produktion och interaktion.

Transformation av policy synliggjordes också i matrisernas design, eftersom de flesta av dem var utformade för att kunna differentiera mellan olika betygsnivåer, något som möjligen kan förväntas i en summativ bedömningssituation. Emellertid innehöll många matriser relativt utförliga beskrivningar av vad kvalitet på olika nivåer innebär och då förefaller dessa matriser snarare vara mer tillämpbara för formativ än summativ bedömning. Våra resultat visade också att ju mer detaljerade kriteriebeskrivningarna blev i matriserna, desto mer skilde sig lärarnas tolkningar åt. Därmed inte sagt att en lärare som använder en mer detaljerad matris gör en annan bedömning än en lärare som använder sig av en mer avskalad variant, då den senare läraren kan ha tydliga (mentala) uppfattningar istället. Vi menar att de mer detaljerade matriserna i vår studie kan ses som verbaliseringar av lärares tysta kunskap om bedömning. Det är känt att det sätt på vilket innehåll organiseras i en matris kan ha stor påverkan på det betyg som sätts (Brown, 2012; Davis, 2018; Khabbazbashi och Galaczi, 2020), där olika typer av matriser har visat sig rendera olika betyg (eng. *score*) för samma prestation (Khabbazbashi och Galaczi, 2020). Ytterligare studier skulle därför kunna ge nya värdefulla insikter om till exempel hur betygssättning påverkas när lärarproducerade matriser används i bedömningssituationer.

Bedömare av ett prov i muntlig färdighet i engelska, där den sociala situationen är ett persamtal mellan elever, möter minst tre utmaningar. För det första måste bedömaren beakta många olika aspekter av kvalitet i tal och samtal samtidigt. För det andra är varje interaktion unik, situationsspecifik och samproducerad av eleverna, men trots det ska individuella bedömningar göras. För det tredje måste bedömaren förhålla sig till tillförlitligheten i bedömningen, både vad gäller att vara konsekvent i sin egen bedömning (intra-reliabilitet) och att bedöma så likvärdigt som möjligt i jämförelse med kollegor (inter-reliabilitet). Dessa tre utmaningar reflekteras i våra data. Det är möjligt att egenskapade matriser hjälper lärare beakta alla de aspekter som anses stå i fokus för bedömningen för att sedan kunna integrera dessa i det holistiska betyget. Matriserna kan även öka både intra- och interreliabilitet genom att de har potential att öka möjligheten att vara konsekvent i sitt bedömningsuppdrag. Dessutom kan kriterier som fokuserar på elevers *interaktion* och *engagemang* vara till hjälp när individuella betyg ska ges för den samproducerade produkten.

Enligt ATD (Chevallard, 2007) genomgår bedömningsanvisningar en transformationsprocess när de operationaliseras i en skolkontext och därigenom framträder den institutionella kunskapen om bedömning. Transformationsprocesserna som vår studie belyser visar att lärares professionskunskap om bedömning av muntlig färdighet i engelska grundas i den dubbla roll lärare har när de agerar bedömare av det nationella provet. Eftersom en elevs muntliga färdighet i engelska är tätt sammanlänkad med *kontext* kan en bedömaren som känner eleven väl ta kontextuella faktorer i beaktande vid bedömningen. Dessutom kan skapandet av en egen analytisk matris möjliggöra formativ återkoppling till de egna eleverna, trots det summativa syftet och det uttryckta holistiska målet med bedömningen från myndighetshåll. Lärares huvuduppdrag är trots allt att hjälpa elever lära sig och utveckla kunskaper och förmågor. Genom att vara integrerade i bedömningsprocessen får lärare mycket god insikt och förståelse för hur den disciplinära kunskapen kan omvandlas för att bli bedömningsbar, och i förlängningen – undervisningsbar. Det ger lärare god insikt om vad en elev som kan tala engelska *kan*, eller med andra ord – vad *kraftfull kunskap* skulle kunna vara i denna kontext.

Bilaga 1

Författarnas sammanställning av Skolverkets bedömningsanvisningar till bedömande lärare i årskurs 6 och årskurs 9 så som de var formulerade vid studiens genomförande 2018.

Bedömningsanvisningar för delprov A för årskurs 6:

ÖVERSIKT: BEDÖMNING DELPROV A

Ämnets syfte

Detta delprov relaterar framför allt till fyra av de långsiktiga målen i kursplanens syftestext: Eleverna ska ges förutsättningar att utveckla sin förmåga att

- förstå och tolka innehållet i talad engelska ...
- formulera sig och kommunicera i tal ...
- använda språkliga strategier för att förstå och göra sig förstådda
- anpassa språket efter olika syften, mottagare och sammanhang

Bedömning av muntlig produktion och interaktion

Bedömningen av muntlig språkfärdighet utgår från att eleven, med utgångspunkt i den givna uppgiften, baserad på kursplanen, vill och kan uttrycka och utveckla ett innehåll, på egen hand och i samspel med andra.

Bedömningsfaktorerna till höger bygger på den kommunikativa och handlingsorienterade språksyn som ligger till grund för de svenska kurs- och ämnesplanerna i engelska och moderna språk. Faktorerna är avsedda att vara ett stöd för analysen vid en helhetsbedömning och ska ses som olika aspekter av kvaliteter i talat språk.

Bedömningsfaktorer

Innehåll

- begriplighet och tydlighet
- fyllighet och variation
 - olika exempel och perspektiv
- sammanhang och struktur
- anpassning till syfte, mottagare och situation

Språk och uttrycksförmåga

- kommunikativa strategier
 - för att utveckla och föra samtal vidare
 - för att lösa språkliga problem genom t.ex. omformuleringar, förklaringar och förtydliganden
- flöde (flyt och ledighet)
- omfång, variation, tydlighet och säkerhet
 - vokabulär, fraseologi och idiomatik
 - uttal och intonation
 - grammatiska strukturer
- anpassning till syfte, mottagare och situation

Nedan följer de delar av kunskapskraven i engelska för år 6, som framför allt betonar muntlig produktion och interaktion. Helhetsbedömningen bör göras i relation till den språksyn och nivå som beskrivs i kursplanen.

Kunskapskrav

Betyget E

I muntliga ... framställningar av olika slag kan eleven formulera sig **enkelt och begripligt med fraser och meningar**.

För att förtydliga och variera sin kommunikation kan eleven ... göra **enstaka enkla** förbättringar av egna framställningar.*

I muntlig ...interaktion kan eleven uttrycka sig **enkelt och begripligt med ord, fraser och meningar**.

Dessutom kan eleven välja och använda sig av **någon strategi** som löser problem i och förbättrar interaktionen.*

Betyget C

I muntliga ... framställningar av olika slag kan eleven formulera sig **enkelt, relativt tydligt och till viss del sammanhängande**.

För att förtydliga och variera sin kommunikation kan eleven ... göra **enkla** förbättringar av egna framställningar.*

I muntlig ...interaktion kan eleven uttrycka sig **enkelt och relativt tydligt med ord, fraser och meningar**.

Dessutom kan eleven välja och använda sig av **några olika strategier** som löser problem i och förbättrar interaktionen.*

Betyget A

I muntliga ... framställningar av olika slag kan eleven formulera sig enkelt, **relativt tydligt och relativt sammanhängande**.

För att förtydliga och variera sin kommunikation kan eleven ... göra **enkla** förbättringar av egna framställningar.*

I muntlig ...interaktion kan eleven uttrycka sig enkelt och **tydligt med ord, fraser och meningar samt i någon mån anpassat till syfte, mottagare och situation**.

Dessutom kan eleven välja och använda sig av **flera olika strategier** som löser problem i och förbättrar interaktionen.*

Betyget D

Kunskapskraven för betyget E och till övervägande del för C är uppfyllda.

Betyget B

Kunskapskraven för betyget C och till övervägande del för A är uppfyllda.

* Fokuseras inte specifikt, men delprovet ger möjlighet till bedömning av denna förmåga.

Bedömningsanvisningar för delprov A för årskurs 9:

ÖVERSIKT: BEDÖMNING DELPROV A

Ämnets syfte

Detta delprov relaterar framför allt till fyra av de långsiktiga målen i kursplanens syftestext: Eleverna ska ges förutsättningar att utveckla sin förmåga att

- förstå och tolka innehållet i talad engelska ...
- formulera sig och kommunicera i tal ...

- använda språkliga strategier för att förstå och göra sig förstådda
- anpassa språket efter olika syften, mottagare och sammanhang

Bedömning av muntlig produktion och interaktion

Bedömningen av muntlig språkfärdighet utgår från att eleven, med utgångspunkt i den givna uppgiften, baserad på kursplanen, vill och kan uttrycka och utveckla ett innehåll, på egen hand och i samspel med andra.

Vidstående bedömningsfaktorer bygger på den kommunikativa och handlingsorienterade språksyn som ligger till grund för de svenska kurs- och ämnesplanerna i engelska och moderna språk. Faktorerna är avsedda att vara ett stöd för analysen vid en helhetsbedömning och ska ses som olika aspekter av kvaliteter i talat språk.

Bedömningsfaktorer

Innehåll

- begriplighet och tydlighet
- fyllighet och variation
 - olika exempel och perspektiv
- sammanhang och struktur
- anpassning till syfte, mottagare och situation

Språk och uttrycksförmåga

- kommunikativa strategier
 - för att utveckla och föra samtal vidare
 - för att lösa språkliga problem genom t.ex. omformuleringar, förklaringar och förtydliganden
- flyt och ledighet
- omfång, variation, tydlighet och säkerhet
 - vokabulär, fraseologi och idiomatik
 - uttal och intonation
 - grammatiska strukturer
- anpassning till syfte, mottagare och situation

Vid betygsättningen av delprov A relateras framför allt till följande delar av kunskapskraven, som särskilt betonar muntlig produktion och interaktion.

Kunskapskrav

Betyget E

I muntliga ... framställningar i olika genrer kan eleven formulera sig **enkelt, begripligt och relativt sammanhängande**.

För att förtydliga och variera sin kommunikation kan eleven...göra **enkla** förbättringar av egna framställningar.*

I muntlig ...interaktion i olika sammanhang kan eleven uttrycka sig **enkelt och begripligt** samt **i någon mån anpassat** till syfte, mottagare och situation.

Dessutom kan eleven välja och använda sig av **i huvudsak fungerande** strategier som **i viss mån löser problem** i och förbättrar interaktionen.*

Betyget C

I muntliga ... framställningar i olika genrer kan eleven formulera sig **relativt varierat, relativt tydligt och relativt sammanhängande**. Eleven **formulerar sig även med visst flyt och i någon mån anpassat till syfte, mottagare och situation**.

För att förtydliga och variera sin kommunikation kan eleven bearbeta och göra **välgrundade** förbättringar av egna framställningar.*

I muntlig ...interaktion i olika sammanhang kan eleven uttrycka sig **tydligt och med visst flyt** samt **med viss anpassning** till syfte, mottagare och situation.

Dessutom kan eleven välja och använda sig av **fungerande** strategier som **löser problem** i och förbättrar interaktionen.*

Betyget A

I muntliga ... framställningar i olika genrer kan eleven formulera sig **relativt varierat, tydligt och sammanhängande**. Eleven **formulerar sig även med flyt och viss anpassning till syfte, mottagare och situation**.

För att förtydliga och variera sin kommunikation kan eleven bearbeta och göra **välgrundade** förbättringar av egna framställningar.*

I muntlig ...interaktion i olika sammanhang kan eleven uttrycka sig **tydligt och med flyt** samt **med viss anpassning** till syfte, mottagare och situation.

Dessutom kan eleven välja och använda sig av **väl fungerande** strategier som **löser problem** i och förbättrar interaktionen **och för den framåt på ett konstruktivt sätt**.*

Betyget D

Kunskapskraven för betyget E och till övervägande del för C är uppfyllda.

Betyget B

Kunskapskraven för betyget C och till övervägande del för A är uppfyllda.

* Fokuseras inte specifikt, men delprovet ger möjlighet till bedömning av denna förmåga.

REFERENSER

- Achiam, M., & Marandino, M. (2014). A framework for understanding the conditions of science representation and dissemination in museums. *Museum management and curatorship*, 29(1), 66–82.
- Alderson, J. C., & Bachman, L. F. (2004). Series editors' preface to *Assessing Speaking*. In S. Luoma, *Assessing speaking* (ss. ix–xi). Cambridge University Press.
- Ang-Aw, H. T., & Goh, C. C. M. (2011). Understanding discrepancies in rater judgement on national-level oral examination tasks. *RELC Journal* 42(1), 31–52.
- Bachman, L. (1990). *Fundamental Considerations in Language Testing*. Oxford University Press.
- Borger, L. (2014). *Looking beyond scores. A study of rater orientations and ratings of speaking*. Licentiatuppsats. Göteborgs universitet.
- Borger, L. (2019). Assessing interactional skills in a paired speaking test: Raters' interpretation of the construct. *Apples – Journal of Applied Language Studies*, 13(1), 151–174.
- Bosch, M., & Gascón, J. (2014). Introduction to the anthropological theory of the didactic (ATD). I A. Bikner-Ahsbals & S. Prediger (Red.), *Networking of Theories as a Research Practice in Mathematics Education* (ss. 67–83). Springer.
- Brookhart, S. M. (2018). Appropriate criteria: Key to effective rubrics. *Frontiers in Education*, 3, 1–12. <https://doi.org/10.3389/educ.2018.00022/full>
- Brown, J. D. (2012). *Developing, using, and analyzing rubrics in language assessment with case studies in Asian and Pacific languages*. National Foreign Language Resource Center, University of Hawaii.
- Brown, A. (2017). Interlocutor and rater training. I Fulcher, G. & Davidson, F. (Red.), *The Routledge handbook of language testing* (ss. 413–425). Routledge.
- Byman Frisé, L., Sundqvist, P., & Sandlund, E. (2021). Policy in practice: Teachers' conceptualizations of L2 English oral proficiency as operationalized in high-stakes test assessment. *Languages* 6(4), 1–23. <https://doi.org/10.3390/languages6040204>
- Bøhn, H. (2015). Assessing spoken EFL without a common rating scale. *SAGE Open* 5(4), 1–12. <https://doi.org/10.1177/2158244015621956>
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied linguistics* 1(1), 1–47. <https://doi.org/10.1093/applin/I.1.1>
- Chevallard, Y. (2007). Readjusting didactics to a changing epistemology. *European Educational Research Journal*, 6(2), 131–134.
- Davis, L. (2009). The influence of interlocutor proficiency in a paired oral assessment. *Language Testing*, 26(3), 367–396. doi:10.1177/0265532209104667
- Davis, L. (2018). Analytic, holistic, and primary trait marking scales. I J.I. Hoboken (Red.), *The TESOL Encyclopaedia of English Language Teaching* (ss. 1–6). Wiley.
- Davis, L. (2019). Rater training in speaking assessment: Impact on more- and less-proficient raters. I S. Papageorgiou & K. M. Bailey, (Red.), *Global perspectives on language assessment* (ss. 18–31). Routledge.
- Ducasse, A. M., & Brown, A. (2009). Assessing paired orals: Raters' orientation to interaction. *Language Testing*, 26(3), 423–443.
- Frisch, M. (2015). *Teachers' understanding and assessment of oral proficiency. A qualitative analysis of results from interviews with language teachers in Swedish lower secondary schools*. Licentiatuppsats. Göteborgs universitet.
- Gericke, N., Hudson, B., Olin-Scheller, C. and Stolare, M. (2018). Powerful knowledge, transformations and the need for empirical studies across school subjects. *London Review of Education*, 16(3), 428–444. <https://doi.org/10.18546/LRE.16.3.06>

Liliann Byman Frisé, Erica Sandlund & Pia Sundqvist

DOI: <https://doi.org/10.58714/ul.v17i2.15844>

- Giraldo, F. (2021). Language assessment literacy and teachers' professional development: A review of the literature. *Profile: Issues in Teachers' Professional Development*, 23(2), 265–279. <https://doi.org/10.15446/profile.v23n2.90533>
- Göteborgs Universitet (1 mars 2023). *Nationellt prov i engelska i årskurs 9*. <https://www.gu.se/nationella-prov-frammande-sprak/prov-och-bedomningsstod-i-engelska/engelska-arskurs-7-9/nationellt-prov-i-engelska-for-arskurs-9>
- Hirsh, Å., & Lindberg, V. (2015). *Formativ bedömning på 2000-talet: En översikt av svensk och internationell forskning* (delrapport från Skolforsk-projektet). Vetenskapsrådet. https://www.vr.se/download/18.2412c5311624176023d25aca/1555423122350/Formativ-bedoemning-svensk-o-internationell-forskning_VR_2015.pdf
- Housen, A., Kuiken, F., & Vedder, I. (2012). Complexity, accuracy and fluency. Definitions, measurement and research. I A. Housen, F. Kuiken & I. Vedder (Red.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (ss. 1–20). John Benjamins.
- Hsieh, H. F., & Shannon, S.E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research* 15(9), 1277–1288.
- Jönsson, A., & Svingby, G. (2007). The use of scoring rubrics: reliability, validity and educational consequences. *Educational Research Review* 2, 130–144. doi:10.1016/j.edurev.2007.05.002
- Kasper, G., & Ross, S. (2013). Assessing second language pragmatics: An overview and introductions. I S. Ross & G. Kasper (Red.), *Assessing second language pragmatics* (ss. 1–40). Palgrave Macmillan.
- Khabbzbashi, N., & Galaczi, E. D. (2020). A comparison of holistic, analytic, and part marking models in speaking assessment. *Language Testing* 37(3), 333–360.
- Kramsch, C. (1986). From language proficiency to interactional competence. *The Modern Language Journal* 70(4), 366–372.
- Lee, Y.-W., Gentile, C. & Kantor, R. (2010). Toward automated multi-trait scoring of essays: Investigating links among holistic, analytic, and text feature scores. *Applied Linguistics* 31(3), 391–417.
- May, L. (2009). Co-constructed interaction in a paired speaking test: The rater's perspective. *Language Testing*, 26(3), 397–421. doi:10.1177/0265532209104668
- May, L. (2011). Interactional competence in a paired speaking test: Features salient to raters. *Language Assessment Quarterly*, 8(2), 127–145.
- McNamara, T. (2000). *Language Testing*. Oxford University Press.
- Meadows, M., & Billington, L. (2005). *A review of the literature on marking reliability*. National Assessment Agency.
- Panadero, E., & Jönsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review* 9, 129–144. doi:10.1016/j.edurev.2013.01.002
- Papajohn, D. (2002). Concept Mapping for Rater Training. *TESOL Quarterly* 36(2), 219–233. doi:10.2307/3588333
- QSR International Pty Ltd. (2018). NVivo (Version 12). <https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home>
- Sadler, D. R. (2009). Indeterminacy in the use of preset criteria for assessment and grading. *Assessment and Evaluation in Higher Education* 34(2), 159–179.
- Salaberry, M. R., & Kunitz, S. (Red.). (2019). *Teaching and testing L2 interactional competence: bridging theory and practice*. Routledge.
- Salaberry, M. R., & Burch, A. R. (Red.). (2021). *Assessing speaking in context. Expanding the construct and its applications*. Multilingual Matters.
- Sandlund, E., & Sundqvist, P. (2016). Equity in L2 English oral assessment: Criterion-based facts or works of fiction? *Nordic Journal of English studies* 15(2), 113–131.

- Sandlund, E., & Sundqvist, P. (2019). Doing versus assessing interactional competence. I R. Salaberry & S. Kunitz (Red.), *Teaching and testing L2 interactional competence: Bridging theory and practice* (ss. 357–396). Routledge.
- Sandlund, E., & Sundqvist, P. (2021). Rating and reflecting: Displaying rater identities in collegial L2 English oral assessment. I M. R. Salaberry & A. R. Burch (Red.), *Assessing speaking in context. Expanding the construct and its applications* (132–162). Multilingual Matters.
- Sato, T. (2012). The contribution of test-takers' speech content to scores on an English oral proficiency test. *Language Testing* 29(2), 223–241. doi:10.1177/0265532211421162
- Sert, O. (2019). The interplay between collaborative turn sequences and active listenership: Implications for the development of L2 interactional competence. I M. R. Salaberry & S. Kunitz (Red.), *Teaching and testing L2 interactional competence: bridging theory and practice* (ss. 142–166). Routledge.
- Skehan, P., & Foster, P. (1999). The influence of task structure and processing conditions on narrative retellings. *Language Learning* 49(1), 93–120.
- Skolverket (2022). *Nationella prov*. <https://www.skolverket.se/innehall-a-o/landningssidor-a-o/nationella-prov>
- Sundqvist, P., Wikström, P., Sandlund, E., & Nyroos, L. (2018). The teacher as examiner of L2 oral tests: A challenge to standardization. *Language Testing*, 35(2), 217–238. doi:10.1177/0265532217690782
- Tsagari, D. (2021). Language assessment literacy: Concepts, challenges, and prospects. I S. Hidri (Red.), *Perspectives on Language Assessment Literacy: Challenges for Improved Student Learning* (ss. 13–32). Routledge.
- Wang, B. (2010). On rater agreement and rater training. *English Language Teaching* 3(1), 108–112.
- Xi, X. (2007). Evaluating analytic scoring for the TOEFL® academic speaking test for operational use. *Language Testing* 24(2), 251–286. <https://doi.org/10.1177/0265532207076365>
- Youn, S. J., & Chen, S. (2021). Investigating raters' scoring processes and strategies in paired speaking assessment. I M. R. Salaberry & A. R. Burch (Red.), *Assessing speaking in context. Expanding the construct and its applications* (ss. 107–131). Multilingual Matters.
- Young, R. F., & He, A.W. (Red.). (1998). *Talking and testing: Discourse approaches to the assessment of oral proficiency*. John Benjamins.
- Young, M. (2009). What are schools for? I H. Daniels, H. Lauder, J. Porter & S. Hartshorn (Red.), *Knowledge, Values and Education Policy: A Critical Perspective* (ss. 10–18). Routledge.

¹ Föreliggande text bygger på en tidigare publicerad engelskspråkig artikel (Byman Frisé, L., Sundqvist, P., & Sandlund, E. (2021). Policy in practice: Teachers' conceptualizations of L2 English oral proficiency as operationalized in high-stakes test assessment. *Languages*, 6(4), 1–23. <https://doi.org/10.3390/languages6040204>). Innehållet har översatts och omarbetats i syfte att nå en svensk läsekrets.

² I denna artikel används begreppet L2 för både andra- och främmandespråk.

Vol 17, nr 2 2023

Tema: Komparativ ämnesdidaktik: Transformationer som utvecklar kraftfull kunskap

Tema: Komparativ ämnesdidaktik: Transformationer som utvecklar kraftfull kunskap

Martin Stolare, Christina Olin-Scheller & Yvonne Liljekvist

Vad kan en elev som kan prata engelska? Didaktisk transposition av muntlig färdighet i lärares matriser för bedömning av det nationella provet

Liliann Byman Frisén, Erica Sandlund & Pia Sundqvist

Att introducera språklig mångfald och migration som tema på mellanstadiet

Anna Lindholm, Lise Iversen Kulbrandstad & Birgitta Ljung Egeland

Semantiska vågor i undervisningen: Likheter och skillnader i skolämnena matematik och samhällskunskap

Martin Jakobsson, Jorryt van Bommel, Ann-Christin Randahl & Niclas Modig

Fotosyntesundervisning 2.0: Kraftfull kunskap och en vidgad syn på fotosyntesundervisning

Anders Eriksson, Niklas Gericke, & Daniel Olsson

Föreställningar om kunskap och lärande i lärares planeringssamtal i matematik och historia

Ann-Christin Randahl, Yvonne Liljekvist, Martin Jakobsson & Kenneth Nordgren