

Bedömning och återkoppling – med fokus på muntlighet

Gudrun Erickson

Göteborgs universitet

gudrun.erickson@ped.gu.se

Abstract

Att kunna tala och samtala är en central och självklar del av en kommunikativ och handlingsorienterad språkfärdighet, som behöver läras, undervisas och bedömas. Med utgångspunkt i min keynote vid konferensen TAL! vid Köpenhamns universitet i oktober 2021 är bedömning och återkoppling, främst av muntlighet, fokus i denna artikel. Ett antal fundamentala kvalitetsfrågor används för att skapa struktur i texten, nämligen, *Varför, Vad, Hur, Vem* samt *Och...?* I detta ligger följaktligen reflektioner kring bedömningars syfte, kunskapsobjekt eller konstrukt, metoder, agens samt användning och konsekvens, vilka sammantaget kan knytas till teorier om validitet, eller kvalitet i vid bemärkelse. Resonemang kring detta förs från såväl konceptuella och kontextuella utgångspunkter som på basis av tidigare forskning och empiriska exempel på muntlig bedömning. Exempel på konkreta frågeställningar som behandlas är formativa och summativa syften, bedömning av produktion och interaktion, analytisk och holistisk bedömning, subjekt och objekt i bedömning och retorik kring bedömning, samt frågor om att bli och vara överens kring bedömning av muntlig språkfärdighet. I detta diskuteras även implikationer av ovanstående på tre nivåer, nämligen den individuella, pedagogiska och strukturella. Slutligen betonas vikten av ett systematiskt kvalitetsarbete, enskilt och i samverkan, kring frågor om bedömning, utvärdering och betygssättning i språk.

Nyckelord

Lärande, undervisning och bedömning, feedback, muntlig språkfärdighet, systematiskt kvalitetsarbete, validitet

1. Introduktion

Denna artikel bygger på en keynote jag inbjöds att hålla vid *'Tal! – Den fællesskandinaviske konference om mundtlighet/mundtlig kommunikation i fremmedsprogsundervisningen'* vid Köpenhamns universitet den 28 oktober 2021. Texten följer i stort sett min muntliga presentation, med fokus på frågor om syfte, kunskapsobjekt eller konstrukt, metoder, agens och användning/konsekvens och innehåller även vissa vidareutvecklingar av dessa grundfrågor. Målet är att belysa bedömning, i allmänhet och av språk och muntlighet i synnerhet, från ett brett utbildningsperspektiv baserat på såväl konceptuella som kontextuella aspekter. Vidare ges konkreta, empiriska exempel från olika praktiker och studier, företrädesvis i de länder som medverkade i konferensen TAL!, nämligen Danmark, Norge och Sverige. Detta breda upplägg innebär att fokus ligger på den inklusiva definition av språkutbildning som präglar mycket av dagens språkpedagogiska och didaktiska diskussion, bland annat manifesterad i *The Common European framework of reference for languages: Learning, teaching, assessment* (CEFR) samt *Companion volume* (Council of Europe, 2001; 2020), där lärande och undervisning liksom bedömning ingår, givetvis med distinkta särdrag, men på ett övergripande plan som en helhet med såväl individuella, som pedagogiska och strukturella aspekter och implikationer. De flesta empiriska exempel som ges är hämtade från grundläggande och gymnasial utbildning engelska i de medverkande länderna, medan de mera principiella och konceptuella delarna är avsedda att vara mera generella.

2. Begrepp och definitioner

Frågor om kunskapsbedömning kan på ett plan betraktas som en självklar del av all utbildning – av såväl lärande som undervisning – men samtidigt finns en icke ovanlig skepsis, eller till och med misstänksamhet och negativitet, till företeelsen. De allra flesta människor har gått i skolan och har därför ofta bestämda åsikter om utbildningsrelaterade frågor, baserade på sin egen erfarenhet (Lortie, 1975). I detta ligger att de stundtals också bär med sig starka minnen av olika bedömnings- och/eller examinationsformer och tillfällen. Dessa kan givetvis vara både positiva och negativa, men de senare tycks ofta dominera eller i alla fall artikuleras betydligt tydligare. Orsakerna till detta torde vara komplexa och därmed knappast möjliga att beskriva och analysera på ett enkelt sätt. Något som ofta är tydligt är dock att kritik mot bedömning och betyg mycket sällan handlar om företeelsen som sådan, utan snarare om användningen. Att kunskaper och kunskapsutveckling på något sätt bör följas, värderas och följas upp ifrågasätts sällan, men en vanlig förhoppning är, för att uttrycka det metaforiskt och relaterat till den

vardagliga situationen att gå till tandläkaren, att det inte ska hända för ofta, att det ska gå fort och inte göra ont (Erickson & Gustafsson, 2020).

För att bättre nalkas den mångfacetterade, utbildningsrelaterade företeelsen bedömning, kan begreppet med fördel betraktas från ett antal olika perspektiv, nämligen semantiskt, strukturellt/politiskt, pedagogiskt och psykologiskt. Detta har också relation till det vidare begreppet validitet, som vi återkommer till i den fortsatta texten.

Semantiskt är det svenska ordet 'bedömning' problematiskt, eftersom förstaveln lätt kan tappas bort och ordet därmed komma att handla om att döma, dvs. något helt annat med betydligt mörkare konnotationer, som lätt kan föra tankarna till disciplinering och makt; om att hålla människor på plats snarare än att hjälpa dem att växa (jfr. Foucault, 1977; Gipps, 1999). Denna (ord)problematik uppmärksammades även av det svenska Skolverket i antologin *Att bedöma eller döma* (2002), i samband med införandet av ett mål- och kriterierelat betygssystem, där en godkändgräns introducerades efter flera decenniers användning av ett normrelaterat/relativt system. Synonymer till 'bedömning' är dock inte lätta att finna; ordet 'värdering' är i detta sammanhang inte nämnvärt bättre och dessutom använt i ordet 'utvärdering', som har en mera systeminriktad betydelse; 'Prov', 'test', och 'examination' representerar mera begränsade och tidspreciserade företeelser. Alltså torde ordet bedömning leva vidare och utmaningen blir snarare att förklara både vad det betyder och vad det verkligen inte betyder.

Strukturellt och politiskt fyller kunskapsbedömning flera viktiga funktioner, relaterade både till innehåll och nivå samt till lärande och likvärdighet. I länder som de nordiska, med nationella styrdokument i form av läroplaner med övergripande och ämnesrelaterade mål, är bedömning ett sätt att avgöra om, hur och hur väl de definierade kunskapsmålen nåtts av hela grupper och undergrupper av elever. I Sverige innebär detta att statistik om bedömning i form av resultat på nationella prov publiceras årligen på Skolverkets hemsida, liksom upplysningar på riksnivå, kommunal- och skolnivå om lärares betygssättning. Likaså tillhandahålls information uppdelad på elevers kön och språkliga bakgrund samt vårdnadshavares högsta utbildningsnivå. Allt detta kan ses som ett sätt att granska likvärdigheten i skolsystemet, men det medför också problem i form av grova, stundtals ytliga och felvisande jämförelser mellan skolor. I kombination med en situation med både kommunala och enskilda skolhuvudmän och dessutom en princip om fritt skolval gör detta att frågor om ökande skillnader mellan kunskapsresultat i landet, det vill säga påtagligt minskande likvärdighet, kommit att fokuseras alltmer (Skolverket, 2018), och i detta också effekter av det som ofta refereras till som marknadisering av skolan (Kornhall m.fl., 2022).

Pedagogiskt fyller bedömning av kunskap en uppenbar funktion på olika plan och för olika syften. En vanlig uppdelning brukar här göras mellan så kallad formativ och summativ bedömning (Scriven, 1967), med ett initialt tydligt systemfokus vid sidan av det individualpedagogiska, och även innefattande lärare, vars professionella kompetens kan gynnas påtagligt av den information som genereras av bedömning. Termerna formativ och summativ används idag ofta synonymt med 'bedömning för lärande' respektive 'av lärande', begrepp som fått stor spridning bland annat genom arbetet i och av den så kallade *Assessment Reform Group*, som verkade i Storbritannien mellan 1996 och 2010. Denna grupp hade medlemmar från olika forsknings- och utbildningsområden med det gemensamma målet att utveckla pedagogisk bedömning på alla nivåer i utbildningssystemet (Black & Wiliam, 1998; Gardner, 2012; Gipps, 1999). Väsentligt att påpeka är dock att inget motsatsförhållande råder mellan de olika bedömningsformerna (Harlen, 2012; Taras, 2005). Syften och metoder varierar men den grundläggande ambitionen är densamma, nämligen att på ett så ingående, brett och respektfullt sätt som möjligt ge så många individer som möjligt chansen att visa vad de faktiskt kan och kan göra med sina kunskaper. De grundläggande principerna kan därmed sägas vara desamma, nämligen att pedagogisk bedömning generellt ska vila på ett fundament av tydlighet/transparens, validitet, reliabilitet och respekt/etik (Erickson, 2018).

Psykologiskt har bedömningar betydelse vad gäller individers förståelse för vilken kunskap som efterfrågas, tillmäts betydelse och premieras. Vidare bidrar bedömningar på olika sätt till deras insikt om och syn på den egna och andras kompetens. Vi återkommer till denna användnings- och konsekvensrelaterade aspekt i den fortsatta texten.

Slutligen i detta avsnitt om begrepp och definitioner är det viktigt att säga att en naturlig aspekt av begreppet bedömning är återkoppling, eller *feedback*, det vill säga hur resultaten presenteras för den individ vars kompetens har granskats. Detta bör göras på ett sätt som både förtydligar och förklarar iakttagelserna som gjorts och samtidigt stödjer och förstärker det fortsatta lärandet, eller som det också kan uttryckas, som *feed forward* mot en fortsatt utveckling (Sadler, 1983). Återkoppling är en företeelse med specifika karakteristika, ofta kontextuella och individuella, men också ett område där samma fundamentala kvalitetskriterier som vid bedömning kan sägas råda, inte minst vad gäller transparens och respekt. I detta ligger givetvis även en välfungerande kommunikation mellan olika parter i processen (se t.ex. Hattie & Timperley, 2007). Även i detta fall kommenteras frågan vidare senare i texten.

3. Övergripande begrepp och frågor kring kvalitet

För att nalkas den helt centrala frågan om bedömningars kvalitet kan med fördel vissa övergripande begrepp och kvalitetsfrågor begrundas och diskuteras. Enkelt uttryckt kan dessa sammanfattas i frågorna *Varför?*, *Vad?*, *Hur?*, *Vem?* samt *Och...?*, det vill säga med fokus på syfte, kunskapsobjekt eller konstrukt, metod, agens samt användning och konsekvens (Erickson, 2011; Takala m.fl., 2016). I samtliga fall kan både generella och ämnesspecifika perspektiv anläggas. I det följande inkluderas det muntliga bedömningstemat för denna artikel i resonemangen kring samtliga frågor men betonas särskilt vad gäller diskussioner om kunskapsobjekt, metod och agens. Först berörs emellertid det grundläggande kvalitetsbegreppet validitet, eller giltighet, som i sig har en tydlig etisk dimension, men även frågan om tillförlitlighet, dvs. reliabilitet.

3.1 Validitet

De övergripande kvalitetsfrågorna som ställts ovan har en tydlig relation till det grundläggande begreppet validitet, eller giltighet som det stundtals benämns. Den traditionella definitionen, som de flesta torde ha stött på, om inte annat i sin lärarutbildning, är att validitet innebär att man faktiskt bedömer det som ska bedömas – att alltså inget annat kommer in och stör. Detta är givetvis fortfarande centralt, men begreppet har successivt kommit att vidgas mot ett mera enhetligt och användningsrelaterat håll, med en stark betoning av värderingar, etik och konsekvenser (Kane, 2013; Kunnan, 2004; Messick, 1989, 1996). I detta varnas bland annat just för störningsrisken, att något annat än det som definierats som det centrala, faktiskt bedöms (*construct irrelevant variance*), men även att för lite av det viktiga finns med (*construct underrepresentation*) (vad gäller begreppet *construct*, se vidare avsnitt 3.3). Exempel på det förra kan till exempel vara att innehållet i en muntlig uppgift är så specifikt att endast vissa personer känner att de överhuvudtaget har något att tala om men också att individer i en samtalssituation känner sig obekväma med varandra och påverkar varandras prestationer negativt. Det senare hotet mot validitet, så kallad underrepresentation, är tyvärr lätt att finna vad gäller muntlig språkfärdighet, som ofta inte tydliggörs på samma sätt som bedömning av exempelvis reception och skriftlig produktion, och som långt ifrån alltid finns med i storskaliga bedömningar av nationellt och internationellt slag. Vad gäller etiska aspekter av bedömning finns mycket skrivet inom området språkbedömningen, där inte minst forskare som Alan Davies (2017) och Elana Shohamy (2001) ska nämnas. Slutligen bör också ytterligare en fundamental egenskap vad gäller bedömning betonas, nämligen reliabilitet eller tillförlitlighet – att de resultat som

framkommer faktiskt går att lita på som giltiga indikatorer på kompetens. Detta är en förutsättning för validitet som traditionellt brukar definieras för sig själv. Reliabilitet har dock allt oftare kommit att ses som en aspekt av, och integrerad med validitet, alltså som en del av det övergripande kvalitetskrav som åvilar bedömning i alla dess olika former (O'Sullivan & Weir, 2011)

3.2 Syfte

Bedömningar kan ha olika syften, allt från diagnos och kartläggning till examination och betygssättning (t.ex. Lindblad, 1985). Avsikten kan även vara att med hjälp av bedömningarna försöka påverka definitionen av och synen på ämnet och därmed uppnå eftersträfvade effekter. I fallet språk kan detta till exempel handla om att handfast visa att muntlig språkförmåga tillmäts stor betydelse, lika stor som andra delar av språkfärdigheten. Denna typ av påverkan kan vara både omedveten och medveten men kan i någon mån också sägas vara ofrånkomlig. Medveten påverkan, såväl på det individuella som på det pedagogiska och strukturella planet, brukar refereras till som *washback* (t.ex. Messick, 1996) och utgör i sig ett intressant och viktigt forskningsområde, som dock inte behandlas närmare i denna text. En tydlig syftesdistinktion handlar vidare om formativitet respektive summativitet, som bör vila på samma grundprinciper men i regel genererar delvis olika metodval.

3.3 Kunskapsobjekt/Konstrukt

Föremålet för bedömningar, alltså svaret på frågan *Vad?*, refereras ibland till som kunskapsobjektet, men en allt vanligare term är också ett direkt lån från engelska, nämligen termen *construct*, på svenska med den anpassade stavningen 'konstrukt.' I skolsammanhang brukar definitioner och exemplifieringar av dessa konstrukt finnas i läro- och kursplaner och de påverkar därmed på ett högst påtagligt sätt lärares praktik både vad gäller undervisning och bedömning. Språksynen i dessa dokument är i regel starkt präglad av begreppen kommunikativ eller funktionell kompetens, som har sin bas i teorier och resonemang formulerade av i synnerhet Hymes (1972) och Canale & Swain (1980) och som vidareutvecklats i den europeiska referensramen för språk och dess *Companion volume* (Council of Europe, 2001 och 2020). Detta medför att fokus, i vid mening, ligger på språkens användning i sammanhang med relevans för dagligt liv, allt från det personliga till det utbildnings- och yrkesrelaterade, och med en tydlig anknytning till kontextualitet och flerspråkighet. Muntlighet är i detta sammanhang av central betydelse och delas i regel upp i en produktiv och en interaktiv komponent med fokus

på att kunna *tala* och att kunna *samtala*. Det senare innebär följaktligen att minst två personer är inblandade och att såväl förmåga att förstå (reception) som att själv tala (produktion) krävs för att samtalet ska fungera. Detta har givetvis stora effekter på utformningen av såväl undervisning som bedömning, där begrepp som *interlocutor effects* och *co-construction*, alltså effekter av ett gemensamt skapande av ett fungerande innehåll, beskriver situationen (May, 2011; O’Sullivan, 2002). I bedömningssammanhang har detta givetvis betydande effekter både på utformning av uppgifter och på analys och värdering av medverkande personers gemensamma och individuella prestationer.

3.4 Metod

Frågan om metod vad gäller bedömning i allmänhet och av muntlig språkfärdighet i synnerhet är bred och djup och innefattar inte minst uppgifters utformning och formulering, sätt att bedöma samt metoder för analys och återkoppling. I detta ligger det uppenbara behovet av att, i enlighet med konstruktet, utveckla uppgifter som på bästa sätt ger de aktuella personerna möjlighet att visa sin språkliga kompetens. Detta görs med fördel i en iterativ process, där aktiv användning av synpunkter från huvudpersonerna – de tänkta provtagarna/eleverna – tillmäts stor betydelse vid sidan av andra typer av analyser, det vill säga där undersökning av performans kombinerad med perception utgör analysens tyngdpunkt (Erickson m.fl., 2022). En ytterligare central aspekt av metodfrågan är sättet att bedöma muntlig språkfärdighet, inte minst den interaktiva kompetensen. Detta görs ofta med hjälp av par- eller gruppmodeller, vilket ställer specifika krav på komposition och analys (Borger, 2018; Brookes, 2009; Galaczi, 2014), i synnerhet när enskilda prestationer ska kunna särskiljas trots att samtalet är samkonstruerat och deltagarna alltid, om än i olika omfattning, påverkar varandra.

En väsentlig aspekt av Hur-frågan är givetvis också det grundläggande sättet att utföra själva bedömningen, det vill säga om man väljer ett analytiskt eller holistiskt förfaringssätt. I det förra bedöms och betygssätts olika aspekter av den avsedda kompetensen var för sig, t.ex. funktionalitet, innehåll och språk, i det senare fallet finns dessa delaspekter av konstruktet med som stöd för en helhetsbedömning, men de betygsätts inte separat. Exempel på kvalitativa aspekter av muntlighet finns till exempel i Europarådets Companion volume (2020), Appendix 3, och definieras som *range*, *accuracy*, *fluency*, *interaction*, *coherence* och *pronunciation*. Indelningar av detta slag är i regel ett gott stöd för analysen, men det råder långt ifrån konsensus kring vilket bedömningssätt som är det mest gynnsamma. En ofta artikulerad synpunkt är dock att analytisk bedömning lämpar sig bättre för formativ bedömning och individuell *feedback*,

medan en väl underbyggd helhetsbedömning fungerar bäst för mera sammanfattande, betygsgrundande bedömningar med höga krav på intern konsistens och reliabilitet (Jönsson, 2021; Ounis, 2017; Torrance, 2007). Slutligen kan konstateras att definitionen av vad som är bra bedömning har många facetter, som har att göra både med lärande och likvärdighet. En viktig faktor i detta är naturligtvis tillförlitligheten eller reliabiliteten. Hur överens är bedömare, hur kan man bli mer överens – och fortsätta att vara det? Det finns många olika aspekter av detta, som jag i min *keynote* sammanfattade på följande sätt:

Vad behövs för att uppnå god överensstämmelse mellan bedömare och bedömningar?

- God kännedom om styrdokument/konstrukt
- Tydliga bedömningskriterier
- Tydliga uppgifter och bedömningsanvisningar
- Tydliga, kommenterade exempel (*benchmarks*)
- Tid för samverkan kring bedömning, inom och mellan skolor
- Envishet och gnetighet! Övning ger färdighet!

3.5 Agens

Frågan om agens handlar om vem eller vilka som är engagerade i bedömning i skola och utbildning. Traditionellt har bedömning och examination varit lärares exklusiva ansvar; de har haft rollen som subjekt i processen, medan eleverna varit objekt – givetvis inte i sitt lärande men väl i bedömningen av den individuella kunskapsutvecklingen och kunskapsnivån. Detta har successivt kommit att ändras, bland annat genom att modeller för självbedömning och kamratbedömning fått spridning och acceptans. I synnerhet självbedömning är en central företeelse i CEFR (Council of Europe, 2001) och dess operationalisering i form av den europeiska språkportfolion ELP (Little, 2009), med konsekvent användning av ett *can do*-perspektiv, det vill säga de lärandes egen syn på sin kompetens. Detta betyder att lärande individers unika kunskap om sin egen kunskapsutveckling tas tillvara, vilket innebär att värdefulla och stundtals helt nödvändiga perspektiv tillkommer för lärare både i deras undervisnings- och bedömningsarbete. Viktigt att framhålla är dock att det juridiska ansvaret för betygssättning naturligtvis fortfarande ligger i professionen.

En ytterligare kommentar behöver dock ges till frågan om subjeks- och objektsfunktioner och -definitioner. Det handlar här om föremålet för bedömningen, vad det är som bedöms. Både i formella sammanhang och i mera allmän retorik finns nämligen en mycket vanlig och mycket olycklig felformulering, som i värsta fall riskerar ta makt över tanken. Det talas alltför ofta om

att man som lärare eller examinator bedömer individer – barn, elever eller studenter. Detta är givetvis fel. Det är kunskapsutveckling och kunskaper som ska bedömas, inget annat, eller som Bachman (1990, s. 20) så riktigt uttrycker det: ”*Whatever attributes or abilities we measure, it is important to understand that it is these attributes or abilities and not the persons themselves that we are measuring*”. Trots den uppenbara självklarheten, behöver detta dock påpekas, och sorgligt kan nog även den gängse terminologin vad gäller en breddad syn på agens i bedömning ge uttryck för samma – säkerligen oavsiktliga – feltänkande, detta genom orden självbedömning och kamratbedömning.

3.6 Användning och konsekvens

Som tidigare nämnts, har definitionen av validitet successivt modifierats och expanderats och innefattar idag – oftast – ett enhetligt begrepp med fokus på användning och effekter/konsekvenser av bedömning (Messick, 1989) och med tydliga inslag av etiska överväganden (Kunnan, 2004). Vidare har betydelsen för en vidgad krets av aktörer kommit att inkluderas. Följaktligen är även lärare och nationella beslutsfattare inkluderade i den grupp som omfattas av den expanderade synen på validitet. Den högst konkreta kvalitetsfrågan kan därmed formuleras som ’hur tolkar vi resultat av bedömningar och hur använder vi dem i beslut och handling’? Konkretiserat kan detta sammanfattas i ett antal punkter som handlar om bedömningens påverkan på och konsekvenser för:

- individers lärande, syn på sig själva och andra
- individers kompetens och kompetensutveckling
- individers framtidsval och möjligheter
- undervisningens och utbildningens innehåll och metoder, *samt*
- samhället i stort.

Detta vidgade förhållningssätt gör att bedömning på ett tydligt sätt kommer att positioneras som en del av en pedagogisk och didaktisk helhet; det handlar tveklöst om lärande, undervisning och bedömning på ett individuellt plan, för elever och lärare, eller uttryckt i mera strukturella termer, om lärande och likvärdighet.

4. Exempel på studier av språklig bedömning med fokus på muntlighet

Synen på bedömning som en integrerad del av den pedagogiska och didaktiska helheten har vuxit sig starkare under senare decennier, vilket även medfört att antalet studier med funktionalitet i fokus har ökat. Detta gäller inte minst inom det språkliga området och därmed

även, mera specifikt, muntlighet. I det följande kommer fyra sådana studier med relevans för konstruktfrågan som behandlats i det föregående avsnittet att beskrivas kortfattat. Vidare nämns två något mera allmänna undersökningar med betydelse för bedömning av språklig kompetens i ett vidare perspektiv och med fokus på de metod och agensaspekter som kommenterats ovan.

Obligatoriska muntliga prov har ingått i de nationella proven i engelska i Sverige sedan mitten av 1990-talet och infördes även successivt i ett tidigt skede i de nationella bedömningsstöden i moderna språk (franska, spanska och tyska). Inkluderingen av en muntlig provdel föregicks av omfattande försöksverksamheter, rapporterad bland andra av Lindblad (1992). I grundskolan innebar reformen på nittioalet att ett nationellt prov, inklusive en muntlig del, också infördes i årskurs 5, alltså för elever i 10-11-årsåldern. Detta föranledde initialt viss oro bland lärare men kom att bli ett mycket uppskattat material med syfte att stödja bedömning och identifiera elever med behov av extra insatser, men även med ambitionen att konkretisera den funktionella och kommunikativa kunskaps- och ämnessyn som präglar de nationella styrdokumentet och därmed de nationella proven. För att ytterligare belysa dessa prov, publicerade Skolverket 2004 en antologi med den talande titeln ”Att visa vad man kan” – för övrigt en av de grundprinciper som gäller för de nationella proven i språk (se vidare [projektets hemsida](#)).

4.1 *It's better to talk!*

Ett av kapitlen i Skolverkets antologi (Nihlén, 2004) handlar om bedömning av muntlig engelska och har ett autentiskt elevcitrat i rubriken: ”*It's better to talk!* Om att samtala och tala på engelska”. I texten, som i sin helhet genomsyras av ett användarperspektiv med eleverna i fokus, beskriver författaren, med förankring i skolans styrdokument, bakgrunden till och rationalen för proven. Vidare berörs metodfrågan på ett brett plan, med fokus såväl på utveckling av uppgifter som på bedömning och genomförande på olika skolor, allt detta interfolierat med citat från elever och lärare om olika aspekter av proven. Teori och praktik knyts därmed samman på ett fruktbart sätt, något som blir extra tydligt i avsnittet ”Hur lyckas eleverna när de pratar?”, som innehåller resonemang kring och exemplifieringar av fundamentala aspekter av muntlig kompetens, exempelvis interaktionen i sig med sina receptiva och produktiva komponenter samt turtagning som en hörnsten i ett fungerande samtal. Vidare berörs frågor om strategier och språkliga fel, det senare även med hänvisning till en annan av [projektets grundprinciper](#), nämligen att skilja på fel som (eventuellt) stör och sådana som förstör, det vill säga mellan fel som i olika hög grad påverkar kommunikationen. Artikeln

avslutas med en sammanfattande kommentar med ett tydlig didaktiskt anslag – att och hur lärande, undervisning och bedömning hör ihop. Och, i linje med artikelns uppläggning, får en elev sista ordet: ”*Det var kul när man skulle fråga varandra frågor och det gick jättebra. Det var grymt skoj!*”.

4.2 Investigating and Validating Spoken Interactional Competence

En senare studie baserad på svenska nationella prov i engelska och med data från elever i slutet av gymnasial utbildning är Borgers (2018) doktorsavhandling med titeln ”*Investigating and Validating Spoken Interactional Competence: Rater Perspectives on a Swedish National Test of English*”. Data härrör i detta fall från en undersökning av 31 bedömningar av sex parsamtal på engelska mellan gymnasieelever, samt de kommentarer som bifogades bedömningarna. Informantgruppen bestod av 17 lärare från 11 svenska gymnasieskolor, som använde svenska ämnesplaner och kunskapskrav för sin bedömning, samt 14 bedömare i Finland och Spanien, som baserade sina bedömningar och kommentarer på CEFR. De svenska styrdokumentet i språk är textuellt jämförda med referensramens nivåer, vilket därmed gav en viss möjlighet att undersöka även detta förhållande. Överensstämmelse mellan bedömarna, liksom korrelationer och reliabilitet studerades med i huvudsak deskriptiva statistiska metoder, medan bedömarnas kommentarer analyserades med hjälp av kvalitativ innehållsanalys. I avhandlingen ingår även en redovisning av en attitydundersökning bland 267 svenska gymnasielärare i engelska, som besvarade en omfattande enkät om sina synpunkter på det muntliga provet. Vidare ger den så kallade kappan en grundlig genomgång av avhandlingens kontextuella och konceptuella förankring samt en redovisning av hur data analyserats och validerats i relation till ett socio-kognitivt ramverk (O’Sullivan & Weir, 2011). Studien genererade en mängd intressanta resultat, bland annat att samstämmigheten mellan bedömare var tillfredsställande men, föga förvånande, med en förbättringspotential. Vidare indikerade analyserna god överensstämmelse mellan de svenska nivåerna och motsvarande beskrivningar i CEFR. Den innehållsliga analysen gav vid handen att bedömning av individuell, interaktionell kompetens var en utmaning, samt att det fanns en uppenbar variabilitet i genomförandet av det muntliga provet. Allt detta är frågor som bör analyseras vidare och i slutänden tillåtas påverka nationella beslut.

4.3 What is to be assessed?

I en doktorsavhandling med titeln *What is to be assessed? Teachers’ understanding of constructs in an oral English examination in Norway* undersöker Bøhn (2016) lärares

uppfattningar om konstrukt i ett muntligt examensprov för gymnasial nivå. Avhandlingen består av tre delstudier, där den första innehåller semistrukturerade intervjuer med 24 lärare i engelska om deras allmänna uppfattning om vad som bedöms i det muntliga provet. Resultaten visade att två huvudsakliga kunskapsobjekt identifierades, nämligen innehåll och kommunikation, som båda innehöll ett antal delkonstrukt. Lärares förståelse uppvisade god, övergripande samstämmighet även i den konkreta bedömningen, dock med vissa skillnader både vad gällde begrepp, till exempel uttal, och viktning av olika delkonstrukt. I den andra delstudien, där 70 lärare i engelska medverkade, användes semistrukturerade intervjuer och enkäter med fokus på uttal. Resultaten gav en blandad bild, med överensstämmelse vad gällde vissa företeelser, till exempel vikten av begriplighet, men stor variation exempelvis i synen på en 'infödd' uttalsnorm. Den tredje studien använde en kombination av verbala protokoll och semistrukturerade intervjuer för att nalkas det identifierade konstruktet innehåll, som visade sig tolkas i en ämnesrelaterad och en förmågerelaterad dimension. Jämförelser gjordes också av relationen mellan lärares uppfattningar om innehåll i relation till skolans styrdokument. Resultaten visade bland annat att lärare tycktes fästa större vikt vid den förmågerelaterade än den ämnesrelaterade dimensionen och att *higher order thinking skills* ofta betonades som väsentliga karakteristika för de högst bedömda prestationerna. Vidare konstaterades att lärare på yrkesprogram uppvisade mindre stränghet i sin bedömningsinriktning än sina kolleger på mera allmänna och studieförberedande program. Sammanfattningsvis genererade studien intressanta inblickar i bedömningen av muntlighet i en nationell kontext, där bedömningsskalor saknades. Liksom i den tidigare beskrivna, svenska studien, gav den aktuella avhandlingen gott underlag för utveckling av den nationella beslutsnivån vad gäller bedömning av muntlig språkkompetens.

4.4 “*Hvis du vakler, er der ridser i lakken*”

En dansk doktorsavhandling (Isager, 2021) har ett generellt och högst relevant tema, nämligen den muntliga delen av den danska studentexamen, oavsett vilket ämne som är i fokus. Även denna avhandling är av sammanläggningstyp, med tre artiklar och en kappa. Fyra forskningsfrågor låg till grund för arbetet, med fokus på (1) hur muntliga examina framställs i styrdokumentet; (2) hur eleverna uppfattar vad som bedöms i proven; (3) vilka adressater eleverna uppfattar att de möter i examinationen och hur de undersöker vad som förväntas under prövningen; samt (4) vilka överväganden studierna ger upphov till för dansk tentamensforskning och för den muntliga gymnasieexamen. Den första artikeln har en

textanalytisk ansats och använder bland annat Kanes (2013) argumentationsbaserade förhållningssätt till validering. En slutsats av studien är att det inte framstår som tydligt vad den muntliga examen avser att bedöma. Vidare diskuteras hur bedömningsteori är kompatibel med dansk examens- och utbildningstradition. Den andra artikeln har en etnografiskt inspirerad ansats och fokuserar på studenters upplevelser av vad ett lämpligt examenssvar skulle kunna vara, detta med utgångspunkt i tre *case*-elevers deltagande och olika sätt att navigera i muntliga examina i historia och engelska. Även den tredje rapporterade studien är inspirerad av ett etnografiskt sätt att studera företeelser och uppfattningar och fokuserar på elevernas uppfattningar om dem som gör bedömningen i examenssituationen, det vill säga lärarna, och hur man kan ge dem vad man tror att de förväntar sig. Avhandlingen är av stort intresse, inte minst för att den lyfter blicken från ett enstaka ämne till en mera generell nivå, där såväl makro- som mikroperspektiv anläggs på frågan om muntlig bedömning.

Slutligen i denna empiriska del av den aktuella artikeln kommer två mera allmänna svenska studier av språkbedömning att nämnas. Den första har, liksom den ovan rapporterade danska avhandlingen, anknytning till ett elevperspektiv och rapporterar resultaten från en europeisk undersökning av elevers och lärares uppfattningar om bedömning av språklig kompetens. Den har därmed anknytning till den grundläggande agensfråga som tidigare diskuterats. Den andra studien har en klar metodkoppling och beskriver en undersökning av bedömersamstämmighet vad gäller elevers skriftliga produktion, som har betydande likheter med bedömning av muntlighet.

4.5 Some European Students' and Teachers' Views on Language Testing and Assessment

Som del av förarbetet inför etablerandet av EALTA, *European Association for Language Testing and Assessment*, gjordes i början av 2000-talet ett antal studier av språkbedömning i Europa. En av dessa hade fokus på elevers och lärares uppfattningar och baserades på enkäter till dessa båda grupper (Erickson & Gustafsson, 2005). Sammanlagt medverkade 1 373 elever (12 -23 år gamla med ett medelvärde på 15) och 62 lärare i 10 europeiska länder. Enkätfrågorna – Likertskalor samt öppna svar, som kunde avges på engelska eller modersmålet – koordinerades i viss utsträckning, vilket möjliggjorde analyser av relationen mellan de båda gruppernas synpunkter. Resultaten visade bland annat att frekvensen för explicit bedömning av muntlig språkfärdighet (produktion, interaktion och uttal) i regel rapporterades som lägre än för andra språkliga kompetenser av både lärare och elever. Vidare var det tydligt att muntlig bedömning var mera vanligt förekommande under lektioner än i formella prov. Elevernas

uppfattningar om bra respektive dålig bedömning sammanställdes och kategoriserades och visade bland annat på följande frekventa karakteristika för god bedömning: variation, fokus på kommunikation/”nytta”, möjlighet att lära sig något samt klarhet och tydlighet. Det följande elevcitatet (ordagrant återgivet) ger uttryck för en icke ovanlig positiv inställning till bedömning av muntlighet och även goda insikter om språklig kompetens:

I think that a good language test/assessment should get students chance to show what they know and devolope their skills. It should contain of listening and reading comprehations, as well as writing and grammar part. But the most important thing is to speak in that language, so I think there should be also oral exams. Speaking is the most important, because without it, we wouldn't communicate. Because of that students should also learn pronunciation. (17-årig elev)

4.6 Nationella prov i engelska: en studie av bedömersamstämmighet

Som ett led i kvalitetssäkringen av de svenska nationella proven, initierade Skolverket 2008 en studie av bedömersamstämmighet i de nationella proven i engelska, matematik och svenska i åk 9. Undersökningen omfattade samtliga delprov med undantag av de muntliga. Orsaken till det senare var inget krav föreligger att spela in de muntliga proven, varför ett krav på insändande av dokumentation skulle ha varit selekterat på ett sätt som inte ansågs acceptabelt. Undersökningarna i de tre ämnena följde en gemensam design med 100 slumpmässigt valda och bedömda prov, tillgång till undervisande lärares bedömningar samt därutöver tre oberoende bedömares bedömningar. Resultaten visade på god eller mycket god samstämmighet vad gällde proven i matematik och engelska, medan motsvarade resultat i svenska var väsentligt lägre. Ett exempel på det senare var att korrelationerna mellan bedömarna vad gällde uppsatsdelen i svenska befanns vara mellan .36 och .46, medan motsvarande värden för den engelska uppsatsen var .86 – .93 (Erickson, 2009; Skolverket, 2009). Medan de engelska resultaten stämde väl överens med kontinuerliga, interna studier, där även liknande studier av bedömning av muntlighet uppvisade motsvarande höga korrelationer, var de svenska förvånande låga, även för provutvecklingsgruppen i svenska. Liknande granskningar, med delvis annorlunda resultat, har sedan 2009 också genomförts på reguljär basis av Statens Skolinspektion. Vi återvänder till detta i den avslutande diskussionen.

5. Diskussion och konklusion

Konferensen *Tal!* hade temat muntlig språkfärdighet – om att kunna tala och samtala på de språk man har lärt och lär sig – och min *keynote* som denna artikel bygger på handlade om hur denna centrala kompetens kan bedömas. I texten har några grundläggande begrepp och principer för god bedömning och återkoppling behandlats och exempel på viktiga studier kring temat har givits. Jag kommer i detta avslutande avsnitt att diskutera några av de nämnda aspekterna på ett lite djupare och mera problematiserande plan och med anknytning till vissa av de grundläggande kvalitetsfrågor som definierats.

Först något om det grundläggande faktum att bedömning har en dubbel funktion, nämligen att i görligaste mån och på bästa sätt stödja såväl lärande som likvärdighet (Erickson, 2016). Här finns inga kompromisser: man måste som lärare och forskare inse att båda dessa funktioner finns och ska beaktas, även om man väljer delvis olika metoder för olika syften. De grundläggande principerna är desamma och handlar om transparens, validitet, reliabilitet och etik. Dock måste också sägas att allt förvisso inte är bedömning! Det finns anledning att varna för en situation där elever kan uppfatta att allt de gör registreras, granskas och värderas. Att som lärare till exempel säga till sina elever att de inte behöver vara nervösa, för läraren ger inga prov, utan hon/han bedömer i stället hela tiden, torde inte vara en särskilt lyckad strategi, även om man säkert menar väl. Lärande innebär en spännande resa, fylld av *ups and downs*, ifrågasättande, hypotesprövning och till synes 'dumma frågor', och under denna process ska man inte behöva känna att allt man gör registreras, vägs och mäts. Följande citat ur den europeiska referensramen (Council of Europe, 2001; s. 185) sammanfattar väl vad som måste undvikas och vad en alltför heltäckande och sömlös definition av bedömning i värsta fall kan innebära, om vi inte skapar rimlig tydlighet kring vad som dokumenteras, när och hur: *"Continuous assessment... can, if taken to an extreme, turn life into one long never-ending test for the learner and a bureaucratic nightmare for the teacher"*.

Ett karakteristikum för språk, som både är en tillgång och kan skapa komplikationer, är den dubbla funktionen som mål och medel; samtidigt som det lärs och utvecklas används det som medel för det mest centrala i lärandekonstruktet, nämligen kommunikation. Detta kan ibland skapa ett behov av att använda en enkel metafor för att förklara egenskaper och samband. Om målet ses som ett stort och välfungerande hus, (*en funktionell språklig kompetens*) måste det först och främst vila på en fast grund – *en kommunikativ och handlingsorienterad språksyn*. Vidare måste ett antal element för att bygga huset identifieras, strategiskt valda och anpassade till beställning, beställare och situation. Dessa kan beskrivas som byggstenar och bräder (*ord,*

fraser, uttryck); murbruk och spik (*grammatik, syntax*) och ytskikt (*stavning, uttal, prosodi*). Ingen kan rimligen ifrågasätta elementen; utan dem blir det inget hållbart hus, *men de är inte huset* (Erickson, 2016). Följaktligen måste vi akta oss för att göra det tämligen lätt mätbara – elementen – till det viktigaste utan hela tiden sträva efter att göra det viktigaste mät- och bedömbart, nämligen den kommunikativa språkliga kompetens som styrdokumentet beskriver.

En väsentlig aspekt av de metodfrågor som kräver svar vad gäller kunskapsbedömning utgörs av frågan om bedömning av produktiva och interaktiva kompetenser, såväl skriftliga som muntliga. Dessa torde vara tämligen självklara delar av den kommunikativa och handlingsorienterade språksyn som präglar allt fler länders skolsystem, men frågan om hur dessa kompetenser kan och ska bedömas med så stor samstämmighet och likvärdighet som möjligt vållar ofta debatt och kontrovers. Ett exempel på detta är den svenska Skolinspektionen under några år hävdande att skrivmomenten i de nationella proven i språk borde tas bort, eftersom granskningar visade att bedömaröverensstämmelsen kunde ifrågasättas. Förslaget möttes visserligen med betydande protester från såväl lärare som forskare och avslogs från den politiska nivån, men den väsentliga frågan om bedömarsamstämmighet kvarstår givetvis. I detta diskuteras ofta analytisk respektive holistisk bedömning, vilket kort berörs i avsnitt 3.4. Även i detta fall används ibland viss metaforik för att bättre ringa in och beskriva fenomenet om relationen mellan delar och helhet, stundtals med målet att påpeka att summan av separata delar inte är additiv i bemärkelsen identisk med helheten. Detta har under våren 2022 aktualiserats i den svenska skoldebatten, när den analytiska bedömning som görs i de nationella proven i svenska kritiserats från lärarhåll, bland annat med konstaterandet att ”kunskap är mer än summan av olika fragment” (Herlitz, 2022). Samma grundtanke återfinns i Erickson & Gustafsson (2020), där det varnas för att önskan att förtydliga kan medföra att ”*stoff förvandlas till stoft, som inte går att rekonstruera till den helhet som i slutänden ska bedömas på ett tillförlitligt och likvärdigt sätt*”.

Slutligen kan konstateras att hela denna artikel egentligen handlar om validitet, det vill säga det överordnade och inklusiva kvalitetsbegrepp, som har fokus på användning och konsekvenser av bedömningar av skilda slag och med en tydligt värderelaterad och etisk prägel. En utmaning för oss alla, enskilt och i samverkan, torde vara att konkretisera detta så att det blir ett brukbart och hållbart verktyg i den breda bedömningsverksamhet som kännetecknar skola och utbildning i skilda kontexter och på olika nivåer i olika systemet. Ett försök att göra detta finns i de Guidelines for Good Practice in Language Testing and Assessment, som utvecklats inom ramen för EALTA, men mycket arbete återstår att förtydliga, sprida och exemplifiera

både principer och goda exempel till gagn både för dem som lär sig och deras lärare samt för samhället i stort.

6. Referenser

- Bachman, L. F. (1990). *Fundamental Considerations in Language Testing*. Oxford University Press.
- Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80(2), 139-148.
<https://doi.org/10.1177/003172171009200119>
- Borger, L. (2018). *Investigating and validating spoken interactional competence: Rater perspectives on a Swedish national test of English*. PhD thesis. University of Gothenburg: Department of Education and Special Education.
- Brooks, L. (2009). Interacting in pairs in a test of oral proficiency: Co-constructing a better performance. *Language Testing*, 26(3), 341-366.
<https://doi.org/10.1177/0265532209104666>
- Bøhn, H. (2016). *What is to be assessed? Teachers' understanding of constructs in an oral English examination in Norway*. PhD thesis. University of Oslo: Department of Teacher Education and School Research.
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1, 1-47.
- Council of Europe (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge University Press.
<https://www.coe.int/en/web/common-european-framework-reference-languages/level-descriptions>
- Council of Europe (2020). *Common European Framework of Reference for Languages: Learning, teaching, assessment – Companion Volume*.
<https://www.coe.int/en/web/education/-/common-european-framework-of-reference-for-languages-learning-teaching-assessment-companion-volume>
- Davies, A. (2017). Ethics, professionalism, rights, and codes. I E. Shohamy, I. Or & S. May (red.), *Encyclopedia of language and education (3rd ed.)*, Volume: *Language testing and assessment* (s. 397-415). Springer.
- Erickson, G. (2009). *Nationella prov i engelska – en studie av bedömersamstämmighet*.
<https://www.gu.se/nationella-prov-frammande-sprak/rapporter-och-skrifter>
- Erickson, G. (2011). Handle with care...! Om referensramen och bedömning av språklig kompetens. I C. Söderberg (red.), *Språklärarens stora blå? En samling texter om Gemensam europeisk referensram för språk*, (s. 31-37). Fortbildningsavdelningen för skolans internationalisering, Uppsala universitet.
- Erickson, G. (2016). Bedömningens dubbla funktion – för lärande och likvärdighet. *Communicare – et fagdidaktisk tidskrift fra Fremmedspråksenteret*, 2016, 4-9.
<https://www.fremmedspraksenteret.no/communicare>
- Erickson, G. (2018). Bedömning av språklig kompetens. I Erickson, G. (red.), *Att bedöma språklig kompetens – Rapporter från projektet Nationella prov i främmande språk*, (s. 15-39). Göteborgs universitet, Rapporter från Institutionen för pedagogik och specialpedagogik, nr 16. <http://hdl.handle.net/2077/57683> .

- Erickson, G., & Gustafsson, J-E. (2005). *Some European Students' and Teachers' Views on Language Testing and Assessment*. European Association for Language Testing and Assessment. <https://www.ealta.eu.org/resources.htm>
- Erickson, G. & Gustafsson, J-E. (2020). Bedömningens dubbla funktion – för lärande och likvärdighet. Grundbok för lärare. I U. P. Lundgren, R. Säljö & C. Liberg (red.), *Lärande, skola, bildning* (5:e utg., s. 589–620). Natur & Kultur.
- Erickson, G., Borger, L., & Olsson, E. (2022). National assessment of foreign languages in Sweden: A multifaceted and collaborative venture. *Language Testing*, 39(3), 474-493. <https://doi.org/10.1177/02655322221075067>
- Foucault, M. (1977). *Discipline and punishment* (A. Sheridan, transl.). Allen Lane.
- Galaczi, E. D. (2014). Interactional Competence across Proficiency Levels: How do Learners Manage Interaction in Paired Speaking Tests? *Applied Linguistics*, 35(5), 553-574. <https://doi.org/10.1093/applin/amt017>
- Gardner, J. (red.) (2012). *Assessment and Learning* (Second edition). London: Sage Publications.
- Gipps, C. (1999). Socio-Cultural Aspects of Assessment. *Review of Research in Education* 24, 355-392. <http://rre.sagepub.com/content/24/1/355.short?rss=1&ssource=mfc>
- Harlen, W. (2012). On the Relationship between Assessment for Formative and Summative Purposes. I J. Gardner (red.), *Assessment and Learning*, Second edition (s. 87–102). Sage Publications.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Educational Research* 77(1), 81-112.
- Herlitz, K. (2022). Skolverkets syn på bedömning är ett hot mot bildningen. *Skolvärlden online* (4. maj 2022). <https://skolvärlden.se/bloggar/karin-herlitz/skolverkets-syn-pa-bedomning-ar-ett-hot-mot-bildningen>
- Hymes, D. H. (1972). On communicative competence. I J.B. Pride & J. Holmes (red.), *Sociolinguistics* (s. 269–293). Penguin Books.
- Isager, J. (2021). "Hvis du vakler, er der ridser i lakken". *Mundtlige danske studentereksaminer fra elevperspektiv*. Syddansk Universitet: Ph.d.-afhandling, Institut for Kulturvidenskaber.
- Jönsson, A. (2021). Analytisk och holistisk bedömning. I Å. Hirsch & C. Lundahl (red.), *Hållbar bedömning Bildning, välbefinnande och utveckling i skolans bedömningsarbete*, (s. 188-200). Natur & Kultur.
- Kane, M. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50(1), 1-73.
- Kornhall, P., Svensson, S., Allelin, M., & Karlsson, B. (red.). (2022). *När skolan blev en marknad. Trettio år med friskolor*. Natur & Kultur.
- Kunnan, A. (2004). Test fairness. I M. Milanovic & C. Weir. (red.), *European Language Testing in a Global Context. Proceedings of the ALTE Barcelona conference, July 2001*. Cambridge University Press.
- Lindblad, T. (1985). On the whys and hows of language testing. I V. Kohonen & A.J. Pitkänen (red.), *Language testing in school: AFinLA yearbook 1985*. Tammerfors universitet.
- Lindblad, T. (1992). Oral tests in Swedish schools: A five-year experiment. *System*, 20(3), 279-292. [https://doi.org/10.1016/0346-251X\(92\)90040-A](https://doi.org/10.1016/0346-251X(92)90040-A)
- Little, D. (2009). *The European Language Portfolio: where pedagogy and assessment meet*,

- 1-22. <https://www.coe.int/en/web/portfolio/elp-related-publications>
- Lortie, D. (1975). *Schoolteacher: A Sociological Study*. University of Chicago Press.
- May, L. (2011). Interactional Competence in a Paired Speaking Test: Features Salient to Raters. *Language Assessment Quarterly*, 8(2), 127-145.
<https://doi.org/10.1080/15434303.2011.565845>
- Messick, S. A. (1989). Meaning and Values in Test Validation: The Science and Ethics of Assessment, *Educational Researcher*, 18(2), 5–11. <https://doi.org/10.1002/j.2330-8516.1988.tb00303.x>
- Messick, S. A. (1996). Validity and washback in language testing. *Language Testing*, 13(3), 241–256. <https://doi.org/10.1177/026553229601300302>
- Nihlén, C. (2004). It's better to talk! I Skolverket *Att visa vad man kan: En samling artiklar om ämnesproven i år 5*. <https://www.skolverket.se/publikationer?id=1271>
- O'Sullivan, B. (2002). Learner acquaintanceship and oral proficiency test pair-task performance. *Language Testing*, 19(3), 277-295.
<https://doi.org/10.1191/0265532202lt205oa>
- O'Sullivan, B., & Weir, C.J. (2011). Test development and validation. I B. O'Sullivan (red.), *Language testing: Theories and practices* (s. 13-32). Basingstoke: Palgrave Macmillan.
- Ounis, M. (2017). A Comparison between Holistic and Analytic Assessment of Speaking. *Journal of Language Teaching and Research*, 8(4), 679-690.
<http://dx.doi.org/10.17507/jltr.0804.06>
- Sadler, R. (1983). Evaluation and the improvement of academic learning. *Journal of Higher Education*, 28(2), 60-79.
- Scriven, M. (1967). The Methodology of Evaluation. I R. W. Tyler, R. M. Gagné & M. Scriven (red.), *Perspectives of curriculum evaluation*. Rand McNally.
- Shohamy, E. (2001). *The Power of Tests. A critical perspective on the uses of language tests*. Pearson Education Limited.
- Skolverket (2002). *Att bedöma eller döma. Tio artiklar om bedömning och betygssättning*.
- Skolverket (2004). *Att visa vad man kan. En samling artiklar om ämnesproven i år 5*.
- Skolverket (2009). *Bedömaröverensstämmelse vid bedömning av nationella prov*. Dnr 2008:286.
- Skolverket (2018). *Analyser av familjebakgrundens betydelse för skolresultaten och skillnader mellan skolor. En kvantitativ studie av utvecklingen över tid i slutet av grundskolan*. Rapport 467. <https://www.skolverket.se/publikationer?id=3927>
- Takala, S., Erickson, G., Figueras, N., & Gustafsson, J-E. (2016). Future Prospects and Challenges in Language Assessments. I D. Tsagari & J. Banerjee (red.), *Contemporary Second Language Assessment, Contemporary Applied Linguistics* (s. 299-315). Bloomsbury Academic.
- Taras, M. (2005). Assessment – Summative and Formative – Some Theoretical Reflections. *British Journal of Educational Studies*, 53(4), 466–478.
- Torrance, H. (2007). Assessment as learning? How the use of explicit learning objectives, assessment criteria and feedback in post-secondary education and training can come to dominate learning. *Assessment in Education*, 14(3), 281-294.